STATISTICS

Economic Surveys and Data Analysis

CIRET Conference Proceedings, Paris 2000

OECD

CIRET

KOF

Economic Surveys and Data Analysis

CIRET Conference Proceedings, Paris 2000

Edited by
Günter Poser in co-operation with Daniel Bloesch

OECD

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Pursuant to Article 1 of the Convention signed in Paris on 14th December 1960, and which came into force on 30th September 1961, the Organisation for Economic Co-operation and Development (OECD) shall promote policies designed:

- to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;
- to contribute to sound economic expansion in Member as well as non-member countries in the process of economic development; and
- to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

The original Member countries of the OECD are Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The following countries became Members subsequently through accession at the dates indicated hereafter: Japan (28th April 1964), Finland (28th January 1969), Australia (7th June 1971), New Zealand (29th May 1973), Mexico (18th May 1994), the Czech Republic (21st December 1995), Hungary (7th May 1996), Poland (22nd November 1996), Korea (12th December 1996) and the Slovak Republic (14th December 2000). The Commission of the European Communities takes part in the work of the OECD (Article 13 of the OECD Convention).

Preface

The Centre for International Research on Economic Tendency Surveys (CIRET) held its 25th conference at the joint invitation of the Institut National de la Statistique et des Etudes Economiques (INSEE) and of the Organisation of Economic Co-operation and Development (OECD) in Paris from October 11 to October 14, 2000.

As key note speaker Professor Edmond Malinvaud addressed the conference participants with his opening speech on "The Role of Business Survey Data in Macroeconomic Research and Practice". This volume contains the opening speech as well as a selection of papers that were presented at the conference.

The papers cover a wide range of topics: (i) Macroeconomic analysis and forecasting, (ii) Leading indicators, (iii) Innovation and technology, (iv) Labour market analysis and consumer surveys, (v) Analysis of survey data with advanced econometric methods and (vi) Use of survey data for economic policy recommendations.

On behalf of CIRET I like to thank all persons who have contributed to the success of the conference:

Hélène Erkel-Rousse (INSEE) and Ronny Nilsson (OECD) and their staff were competent organizers and charming hosts. Mr. Seiichi Kondo, Deputy Secretary-General at the OECD, gave the closing speech. Petra Huth and Ruth Vognstrup at the Swiss Institute for Business Cycle Research (KOF) in Zurich supported the local organizers INSEE and OECD.

We thank the French Ministry of Economy, Finance and Industry, the French Economic Forecasting Department and the European Central Bank for sponsoring the 25th CIRET conference.

We thank Jean-Jacques Vanhaelen at the National Bank of Belgium for his support in the reconstitution of CIRET as an association with legal seat in Brussels. The first general assembly of the renewed CIRET, headed by Professor Bernd Schips, new President of CIRET, was held in Paris on October 10, 2000.

Finally we appreciate the endeavour of Ronny Nilsson (OECD) who made this publication possible and we thank Susanne Boehm for formating the papers.

Darmstadt, October 2002 Günter Poser

Table of Contents

III Innovation and Technology

IV Labour Market Analysis and Consumer Surveys

V Analysis of Survey Data with Advanced Econometric Methods

VI Use of Survey Data for Economic Policy Recommendations

List of Contributors

Elzbieta Adamowicz	SGH, Warsaw School of Economics, Warsaw, Poland
Marlene Amstad	Economic Research, Swiss National Bank, Zurich, Switzerland
Benoît Arnaud	OECD Secretariat, Paris, France
Spyros Arvanitis	KOF, Swiss Institute for Business Cycle Research, Swiss Federal Institute of Technology, Zurich, Switzerland
Anirvan Banerji	Co-Director of Research, ECRI, Economic Cycle Research Institute, New York, USA
Daniel Bloesch	Head of CIRET Information and Documentation Centre, KOF, Swiss Institute for Business Cycle Research, Swiss Federal Institute of Technology Zurich, Switzerland
Olivier Brunet	OECD Secretariat, Paris, France
John A. Carlson	Loeb Professor of Economics, Department of Economics, Purdue University, West Lafayette, USA
Paolo Carnazza	Head of Economic Research, Confartigianato, Rome, Italy; formerly ISAE, Rome, Italy
Cheng-Mount Cheng	Assistant Research Fellow, TIER, Taiwan Institute of Economic Research, Taipei, Taiwan
Marcella Corsi	Associate Professor, LUISS, Libera Universita Internazionale degli Studi Sociali, Rome, Italy
Richard Curtin	Senior Associate Research Scientist, Institute for Social Research, University of Michigan, Michigan, USA
Jan De Mulder	National Bank of Belgium, Brussels, Begium
Luc Dresse	National Bank of Belgium, Brussels, Begium
Richard Etter	Head of Survey Research, KOF, Swiss Institute for Business Cycle Research, Swiss Federal Institute of Technology, Zurich, Switzerland
Anna Gutkowska	SGH, Warsaw School of Economics, Warsaw, Poland
Heinz Hollenstein	KOF, Swiss Institute for Business Cycle Research, Swiss Federal Institute of Technology, Zurich, Switzerland

Ulrich Kaiser	Associate Professor, Department of Economics, University of Southern Denmark at Odense, Odense, Denmark
Pertti Kangassalo	Statistics Finland, Helsinki, Finland
Philip A. Klein	Professor, Senior Research Scholar, ECRI, Economic Cycle Research Institute, New York, USA
Jacek Kotlowski	SGH, Warsaw School of Economics, Warsaw, Poland
Edmond Malinvaud	Professor, Collège de France, Paris, France
Zbigniew Matkowski	Senior Lecturer, SGH, Warsaw School of Economics, Warsaw, Poland
Gernot Nerb	Head of Survey Research, ifo Institute for Economic Research, Munich, Germany
Ronny Nilsson	Administrator, OECD, Paris, France
Veijo Notkola	Statistics Finland, Helsinki, Finland
Giuseppe Parigi	Research Department, Bank of Italy, Rome, Italy
Günter Poser	Professor, Darmstadt University of Technology, Germany
Stanley Presser	Professor, University of Maryland, Maryland, USA
Eleanor Singer	Senior Research Scientist, University of Michigan, Michigan, USA
Alexandra Spitz	ICT Research Group, ZEW, Centre for European Economic Research, Mannheim, Germany,
Markus Taube	Professor, University of Duisburg, Germany
Neven T. Valev	Assistant Professor, Department of Economics, Georgia State University, Atlanta, USA
Jean-Jacques Vanhaelen	Head of Economic Research, National Bank of Belgium, Brussels, Begium
Konrad Walczyk	SGH, Warsaw School of Economics, Warsaw, Poland
Anita Wölfl	Young Professional, OECD, Paris, France; formerly IWH, Institute for Economic Research Halle, Halle, Germany

Opening Address

The Role of Business Survey Data in Macroeconomic Research and Practice

Edmond Malinvaud

This twenty-fifth conference of the Center for International Research on Economic Surveys bears witness to the longevity and vitality of a research program that was launched in the nineteen fifties. My opening address is, first and foremost, the declaration of someone who early on worked close to those who launched the program. This was then a natural outgrowth of the setting up of regular national systems of business surveys, in the spirit of the initiators, particularly of the Munich Ifo Institute. For us young macroeconomists and statisticians of the early fifties, the program was part of the drive for enlightened economic management, which we wanted to promote more generally. Even though during the following decades we had to realize that parts of our motivations were definitely naïve, the business surveys survived pretty much as they had been conceived and they kept performing the functions envisaged for them from the beginning.

Rather than elaborating on this initial remark I may be more useful to you today by choosing a different subject, to which however I must also give the historical tone expected from an old man like me. The subject will be related to my scientific interests in macroeconomics. It will be: the role of business survey data in macroeconomic research and practice. The main point I want to make is precisely that the role in question should be better embodied in research on business cycle theory and in its applications.

Concentrating on this subject means that I shall be silent on other subjects which will also take a large part in your discussions and are very interesting on their own right. I must at least mention in this connection all papers dealing with business survey techniques. I must also recognize that business surveys bring essential informations about the structures of our economies, or about such very relevant topics as the importance and nature of innovations. Please accept my apologies and this recognition that my address would be very incomplete if it was meant to cover the whole range of the subjects to be investigated during this conference.

In order to deal with the role of business survey data in macroeconomics the first half of this talk will speak of the demands addressed by research and practice to those in charge of business surveys. The concerns of academic research will be the topic of my first part, those of practitioners in government or business services will be the topic of my second part. The two last parts will discuss more specifically to what extent business surveys contribute to knowledge, respectively in two areas which are both challenging for research and practice, namely the characterization of market disequilibria in the first place, the continuous observation of fluctuating expectations and intentions of market participants in the second place.

1 The Needs as Perceived through the Evolution of Business Cycle Research over the last Fifty Years

Clearly, business cycle research needs to precisely refer to empirical data. That may not have been obvious during the first part of this century, when the theoretical writings surveyed in the famous book of Godfried Haberler, first published in 1936, presented only qualitative descriptions and explanations of business cycles. But already in those times, it had become common to state that the business cycle did not cope with the fact that business fluctuations much varied, more even than could be explained by random shocks recurrently hitting the economies. Some economists had realized that a good understanding of this diversity required a close examination of data, probably also a more detailed analysis than was offered by general theories.

Undoubtedly many small research teams initiated work on empirical research projects about business fluctuations. But only two of those early projects could be pursued with sufficient energy and perseverance to bring significant contributions to business cycle theory. I like to quote first the project launched in the 1930s at the National Bureau of Economic Research in New York, a project still active today. It is not the place here to discuss the significance of its contributions, which appears to have been definitely lower than its initiators had hoped. But the quantitative analysis of business fluctuations had begun and would later grow.

Before I mention the second major empirical project on business fluctuations initiated about at the same time in Europe, let me very briefly sketch the evolution of business cycle theory since the first edition of Haberler's book. This is simple. The advent of Keynesian theory and of aggregate demand analysis provided a unification of prevailing explanations of business fluctuations. The little book published in 1950 by John Hicks, adding occasional supply constraints to the multiplier-accelerator model, provided a widely accepted overall reference for at least two decades. Actually, it had no competitor for that function until publication, in the early 1980s, of the first articles arguing in favor of a real business cycle theory, which would extend to fluctuations the neoclassical theory of economic growth.

Few economists endorsed the new theory as such, but many found in its methodology a way to build alternative models suiting their own views. We might argue that the set of models so built now provides the overall reference, which replaced Hicks' model. There is, however, one problem with this last assertion, namely that it hides the importance of the mounting flow of detailed research on such diverse factors as inventory accumulation, the scrapping of obsolete productive capacities, credit rationing, and so on. The largest part of this flow makes no explicit reference to an overall embracing theory and, in most cases, would fit into Hicks' approach broadly understood.

The second major empirical project to appear in the 1930s aimed at building structural macroeconometric models. Initiated by Jan Tinbergen, it was later taken over by Lawrence Klein who promoted it throughout the world in the 1950s and 1960s. In those years at least, the structural models had the same Keynesian inspiration as Hicks' model. These models were sometimes used for research on business cycle phenomena. But they were applied mainly for providing macroeconomic forecasts and policy analyses. This is why I shall discuss their data needs in the second part of my talk only.

Looking at the empirical part of the rest of business cycle research, I am led to stress that it uses, almost exclusively, the body of data made available by official statistical offices, mainly the time series of economic and social indicators, as well as series of national account aggregates. More and more often now, academic research in economics also uses cross-sections or panels of microeconomic data. This applies in particular to the flow of research projects on detailed aspects of business fluctuations, a flow to which I drew your attention a moment ago.

Since your concern in this conference is data coming from business surveys, I must say that direct use of the results coming from these surveys is fairly rare, at least if we have a narrow conception of business surveys, namely surveys similar to those early made at the Ifo Institute. I know that a broader conception prevails in your program, but the narrow definition, which was common when CIRET was founded, is more relevant for my talk, I believe. However, there is also indirect use when business survey results have served statisticians for the estimation of indicators or aggregates, such as productive capacity or even investment.

Although fairly rare now direct use of business survey results is valuable for business cycle research, so valuable that it should be more frequent. This is so because of the importance of the two aspects which I shall more precisely discuss in the third and fourth part of this address, namely the nature of market disequilibria and the formation of expectations or intentions of market participants. As you well know, questioning business firms on their expectations or intentions, and on market slacks or tensions they are experiencing, is indeed the main object of business survey narrowly conceived.

Clearly, the dynamics of business fluctuations much depends on the nature of market disequilibria, which appear along these fluctuations and which constrain opportunities, behaviors and adjustments. Clearly, this dynamics also depends on what are expectations and intentions, hence on how they are formed. Econometricians taking part in business cycle research know these facts; hence, their models often contain variables about the nature of disequilibria or about private expectations. But often also econometricians deal with such variables as if their values were not directly observed: the variables in question are then latent variables, whose values are estimated simultaneously with the parameters of the model.

Actually, business surveys give information about those variables. Neglecting the information leads to a loss of efficiency. Perhaps, business survey results do not give quite the values of the variables entering the model estimated by the econometrician, and this has then to be taken into account. But, as long as the results bring relevant information, they should be used.

Moreover, there is no reason in principle, which would forbid improving the business surveys, so that their results come closer to what research needs. Indeed, this is an opportunity for me to advocate participation of academic users in the conception or revision of business surveys.

2 The Needs of Practitioners Aware of the Fact that Market Tensions and Business Anticipations Matter

We do not need a long demonstration in order to argue that economic forecasts and policy analyses often have to focus on market tensions and business expectations. The spontaneous evolution of the economy in the months and the years to come will much depend on what are present tensions and anticipations, on how they are likely to evolve. The effects to be expected from alternative policies at the short-term horizon of one year or two will similarly depend on the diagnosis that can now be made about the business climate and about the market constraints firms are facing. Even the medium-term effects, namely the results after five years or so, will be similarly affected since, for instance, the degrees of capacity utilization at such a horizon will depend on investments made in the near future and since the way in which investments will react to policy decisions will depend on the business climate and constraints when decisions will be made.

Macroeconomists working on forecasts and policy analyses, whether in government or in private consulting firms, are well aware of these facts, much better aware, it seems, than some of those working in academia. By the way, this is why Keynesian modes of thinking remained commonly used by forecasters and policy analysts, even at the time when it was fashionable in academic circles to discredit these modes of thinking.

My comment at this point should not be misunderstood. I do not want to disparage all the research work which was done in academia during the last two decades by new-classical macroeconomists. Some of it is quite valuable and brings irreversible improvements in our modes of thinking. In particularly, the way in which we think about the shocks hitting the economy has much improved. There is also improvement in the fact that, in policy analyses, we now immediately think about the credibilities of alternative policies and about the possible feedbacks from a change in the policy regime to private patterns of behavior.

Many such improvements are due to the work of new-classical macroeconomists. This must be recognized, even when we stand firm is saying that market disequilibria exist and are often important, or that we cannot have full confidence in the way in which most rational expectation models describe the formation of private expectations.

Actually, even though Keynesian modes of thinking remain valid in these respects, it does not mean that the tools used in the older days for computing forecasts or policy assessments were perfectly suited. For instance, the systematic use of adaptive expectations may also create problems. In the last part of my address I shall wonder about alternative treatments of expectations in structural econometric models.

For the time being I shall rather draw your attention towards another imperfection of those models, namely that, according to me, they took too little advantage of business survey data; they still now take too little advantage of them. This is fundamentally the same remark as the one I made at the end of the first part of my talk about the needs of research. But it is now formulated with respect to structural econometric models, the main tool at the disposal of macroeconomic forecasters and policy analysts.

The teaching about these models seems to completely overlook the fact that business survey data contain information that would be useful for practitioners. The data base of structural models is assumed to contain only time series of the values taken by traditional statistical indicators and aggregates. Thus students may even ignore that other kinds of data, provided by business surveys, also exist.

Actually, real practice is not as deficient in this respect as is the teaching of practice. In the first place the specification of some structural models involves variables whose values is meant to come from business surveys. I am not in a position to know how many such macroeconometric models now exist or existed in the past. But I remember to have seen such models already long ago. In the second place, when using their models in order to make forecasts, practitioners often bring adjustments either to the results derived from the model, or even to the model itself before it is used for the calculation of a projection. Such adjustments are often due to the information coming from business surveys. For instance, the projection obtained from the model predicts a level of investment which does not agree with what is announced by an investment business survey and the economists in charge of the forecast change the value to be given to future investment so that it comes closer to what the survey announced. Or even, these economists revise their model because they know that the situation when the forecast is being made is special in some respects. For instance, this is the time of a credit crunch and the investment equation gives little weight to the level of firms indebtedness; revising the equation so as to increase this weight then appears to be wise.

3 Business Surveys for the Characterization of Market Disequilibria

Let us now look more precisely into the phenomena for the study of which information coming from business surveys is particularly useful. I consider first market disequilibria. You know what we mean by a market disequilibrium, namely the fact that the degree of tension or lack between supply or demand varies. Concerning the markets for goods or services, this means that buyers are more or less rationed because they have more or less difficulty in finding what they want to buy. On the other side, sellers experience changes in their outlets; producing firms then have to adapt the levels of their output or their inventories, usually both. Concerning the market for labor, more or less numerous workers are unemployed, many or few employers experience difficulty in finding the people they would like to recruit. Concerning the credit market, banks have more or less tight standards in evaluating the solvency of clients asking for loans.

The concept of market disequilibrium extends beyond variations in the quantities desired for exchange respectively by suppliers and demanders. There are also simultaneously price disequilibria, because many prices are somewhat rigid. The typical case of a price disequilibrium is when the profitability of production is abnormally high or low. Market forces then stimulate a correction, down or up, but that will come more or less quickly depending on how evolve other factors acting on prices, wages and interest rates.

Asking market participants in order to know which disequilibria they were experiencing in the markets was the dominant aim of the initiators of business surveys, defined in the narrow sense I selected for this address. The information so collected was and is still now useful for both business analysts and macroeconomists. Broadcasting the information is therefore useful. But what should users do with it? In order to discuss the answer to this question I shall limit attention to macroeconomists: what should macroeconomists do with the information on market disequilibria which comes from business surveys?

The first and obvious answer is that they should take account of it. This is so obvious that no one would think of giving this answer if there were not quite a few macroeconomists who take no account of the available information on the existence of market disequilibria, as I hinted earlier when speaking of new-classical macroeconomics. I could even quote textbooks, meant to teach macroeconomics, in which the concept of a market disequilibrium does not appear (the fact does not speak in favor of the academic profession). After mentioning the point, I shall restrict attention to macroeconomists who are not blind to the existence of market disequilibria.

The second answer to the question I posed is to say that macroeconomists should use the information coming from business surveys in order to distinguish between various combinations of market disequilibria. For instance, one combination would be: excess supply on both the markets for goods and for labor, equilibrium on the credit market, normal profitability. Actually, this combination was much studied by the Keynesian literature. In order to designate it by a simple expression, we can speak of Keynesian unemployment. Another combination would be: excess demand on the market for goods, excess supply on the market for labor, equilibrium on the credit market, abnormally low profitability; we can then speak of classical unemployment. Giving this second answer to the question, macroeconomists would then take the habit of using a typology of combinations of market disequilibria. Research would study each one of these combinations. In applied macroeconomics, practitioners would use the typology for their diagnoses, forecasts and policy analysis.

There is a, still better, third answer to the question, namely to study the factors explaining the nature and intensity of disequilibria. The result of this study would be, on the one hand, a theory of market disequilibria and, on the other hand, a new family of structural macroeconometric models. Let me make right away two comments about structural models of this new family. First, they will have to be dynamic, as may be suggested already by the word disequilibrium. Second, structural models now in use recognize the existence of market disequilibria, but in a rather loose way. Hence, they belong to the family, which ought, however, to be more explicitly characterized.

Research has undertaken the study I just described. But, as I hinted when using the conditional mood in my description, the study did not reach the stage where we could say that a full theory of market disequilibria has been established and a new family of immediately usable structural models has been delivered to practitioners. I may even add that, faced with serious difficulties, the study marked a pause. Undoubtedly it will have to be resumed, but probably from an approach which will have to dodge part of the difficulties in order to cope with the main remaining problems.

However, it is clear to me that the information gathered by business surveys about market tensions or slacks will play a significant part in whatever approach will be taken in the future by the resumed research effort. This information was already embodied in some relevant macroeconomic work where market disequilibria were identified. For instance such a work concerned European unemployment. We should draw the lessons of this experience, more fully than they have been up to now. Moreover, in order to assist in the evolving methodology of structural models, we should welcome a research program on how business survey results on market disequilibria are being already used in forecasts and policy analyses, and how they should best be used in the future.

4 Business Expectations and Intentions as Captured through Surveys

Beginning now the last part of my talk, I must first briefly comment on the role of expectations in the macroeconomy and on the exact power of the rational expectations hypothesis in macroeconomics. The hypothesis is so often used in the modern theoretical literature that we could be misled into thinking that it is perfectly appropriate and very powerful.

Expectations are undoubtedly important. I can quote here Robert Solow writing that all economists "would agree that the response of the macroeconomy to disturbances will depend on the beliefs, perceptions and expectations of participants". But he adds: "I would not know what set of statements about these things deserves to be included in the core of usable macroeconomics. I feel acutely uncomfortable with this fudge factor that is capable of having drastic effects but is so conjectural that it can be used to explain just about anything".[1] The two last sentences are provocative and certainly too extreme. But they are welcome in order to challenge the dogmatism of some writers and to cool down the subsequent naïve enthusiasm of some of their readers.

There are indeed a number of reasons for judging that the scope of validity of the rational expectations hypothesis is much more narrow than the scope of its presence in modern theoretical literature. I have no time to dwell on that. But I shall, first, say that, embodied within a model, the hypothesis means that actual expectations have to be consistent also with all other hypotheses of the model. I shall, second, note that the hypothesis was not found to perform so well empirically when econometricians tried to test it independently of the models in which it could be embodied: for the purpose econometricians referred only to the set of public information available to the agents whose expectations were analyzed. Econometricians then tested whether forecasts made by individual agents optimally used this available information.

However, it is noteworthy that econometricians found also that the adaptive expectations hypothesis, the only existing competitor with the rational expectations hypothesis, did not either perform so well empirically. Faced with the dilemma resulting from these econometric findings, modern builders of structural models, who know how to technically deal with rational expectations, usually limit to a small number the variables about which rational expectations appear in their models. Some such builders also provide two versions of their model, one with rational expectations, the other with adaptive

---

[1] In the symposium "Is there a core of practical macroeconomics that we should all believe?", American Economic Review, May 1997.

expectations only. Users are then in a position to know when and how the choice of the hypothesis about the formation of expectations affects the forecasts or policy analyses made.

Aware of this context and of some others of its features, we may wonder whether present business surveys give, about expectations and intentions, the data which would best fit the needs of researchers and practitioners. I am not a good judge of this question because my acquaintance with these surveys is partial or dated. What I am going to say should therefore be understood as queries rather than as leading to recommendations.

I remember to have read studies showing that intentions collected by surveys of households and firms were fairly imperfect predictors. For instance, intentions about the mount of investment to be made a year later turned out to differ substantially from investment which were really made by the respondents. There is a natural explanation of the act, namely that the intentions were conditional and that conditions changed in the meantime. Now that we have a large experience with data on intentions, can we say that this explanation olds? Or is it rather that answers about intentions are made without care and are so subject to a large amount of insignificant noise?

I also remember to have been told that answers about expectations could reflect as such the current prevailing "business climate" as real forecasts about the future. This should be disturbing for someone who, like me, is advocating the use of business survey data in macroeconomic research and practice. In our investigations and applications should we often introduce a variable meant to represent the perceived business climate? How should then the variable be exactly defined and measured? Which hypotheses should we test about the effect of this variable on decisions taken by agents? Would the principle of rational economic behavior suffice in order to suggest the proper hypotheses? Or should we also turn to psychology?

Common sense suggests that this, up to now loose, variable called here "perceived business climate" matters mainly for answers to questions about the economy in general rather than for expectations about what specifically concerns the respondent: his or her earnings next year, the prices next year for the goods or services sold by the firm, and so on. But can we rule out off hand the idea that the mood of respondents might be sensitive to the business climate and that this might bias his or her answers even to specific questions on the future values of variables about which he or she should be well informed?

The various queries I just listed suggest that the introduction, in macroeconomic research and practice, of business survey results about intentions and expectations may be less direct and simple than the wording of the survey questions would itself imply. A collaboration between survey specialists and macroeconomists may be rewarding in clarifying the issues. Within such a collaboration it would be good to also wonder whether business surveys could not more often be used for collecting data about anticipations up to longer horizons than is now common, i.e. up to medium-term horizons which are so challenging for us in research and practice.

Chapter I

Macroeconomic Analysis and Forecasting

A Comparative Study of Forecasting Performance: European Long Range Gauge vs. European Business Surveys

Anirvan Banerji and Philip A. Klein

Introduction

For many years there has been debate about the relative forecasting properties of indexes based on quantitative data and those based on qualitative data. While at the end of the day there is no doubt much to be said for utilizing both kinds of data, for analytical purposes we need to know what each kind of indicator contributes to such a forecasting tool. In the current paper we ask this question with new data. Because of the attention paid to the emergence of the Eurozone, many students of economic instability have asked what will be the impact of this new global configuration on important economic variables, and our ability to forecast economic instability.

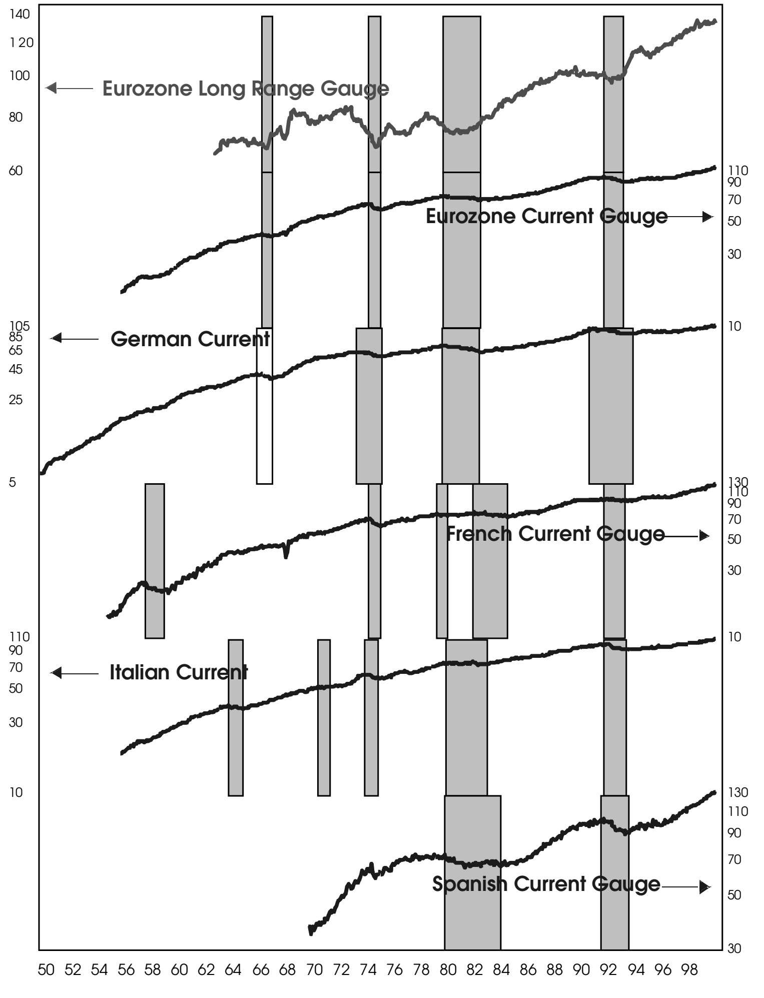
Constructing Eurozone Composite Indexes

At the Economic Cycle Research Institute (ECRI) we regularly monitor cyclical developments for Germany, France Italy, and Spain. These four countries account for approximately 82% of Eurozone GDP. It seems reasonable, therefore, to combine the information available for these four countries and utilize the resultant composite indexes to ask how well we can use them to forecast economic activity for the entire Eurozone.

Inasmuch as this means that we shall be using these four countries to represent the Eurozone, we need to consider first how well Eurozone composite indexes reflect cyclical developments in the individual countries. (It would be possible to have an index based on data representing 82% of Europe but still not doing a good job of reflecting the cycles in the individual countries included in the composite index).

Our procedure for choosing business cycle chronologies has always involved, in addition to looking at the turning points in the composite indexes, looking at the coincident indicators used in the composite index for each country, and examining the clusters in turning points among these individual series at both peaks and troughs. This enables us to evaluate the degree to which the composite indexes for each country are reflected in the turns in the component series. (In our work in individual countries we have calculated the medians of the timing of individual series.) In somewhat similar fashion, we began here by asking how good a job the four-country composite index did of reflecting the cycles in the four component countries.

Chart 1. Tracking Cyclical Movements in the Eurozone and Major Eurozone Economies



Note: Shaded areas covering the top two panels represent cyclical downturns in the Eurozone Current Gauge.
Shaded areas in the third panel represent German business cycle recessions.
Shaded areas in the fourth panel represent French business cycle recessions.
Shaded areas in the fifth panel represent Italian business cycle recessions.
Shaded areas in the sixth panel represent Spanish business cycle recessions.

Chart 1 suggests the answer. It shows at the top the Eurozone Long Range Gauge (EZLRG) and the Eurozone Current Gauge (EZCG) for the period 1972-2000. There are three clear recessions during this relatively long period – 1974-75, 1980-82, and 1992-93. How well are we capturing the cycles in the component countries? The EZLRG shows one extra, albeit mild, cycle in 1976-77. It is visible in the current gauges but only as slightly reduced rates of growth. In none of the gauges shown in Chart 1 does the Bry-Boschan program (Bry and Boschan, 1971), an algorithmic formulation of the classical NBER procedure for choosing turning points, regard these turning points around 1976-77 as significant.

In fact, the chart shows the German Current Gauge and clearly the same three cycles are exhibited, although the third cycle is plainly longer in Germany than for the Eurozone. In France the same three cycles are visible but this time the middle cycle in the early 1980s is slightly later and is preceded by a short recession in the late 1970s visible primarily in pervasive weakening in the components rather than the amplitude of the composite gauge. In Italy there are three cycles, but the middle recession is somewhat longer. In Spain the first cycle is visible but is too mild for the Bry-Boschan algorithm to include it in the recessions it registers. Again the middle recession is longer – this time it is the longest of any of the four recessions we examine.

In sum, Chart 1 suggests that the EZCG does a good job of reflecting the majority of the significant cyclical changes in the four component countries. Because they represent such a large part of the Eurozone, we may assume that the chronology based on the four largest countries is a reasonable reflection of Eurozone cyclical activity. The EZCG is a good proxy for aggregate economic activity in the European Monetary Union and the findings are unlikely to be very different if the remaining EMU countries were to be included.

Cyclical Timing of the Eurozone Long Range Gauge

If we focus now on the turning points in the EZLRG and the EZCG the lead of the EZLRG at business cycle turns is very impressive – it averages 19 months at the peaks and 12 months at the troughs (Table 1).

We turn now to a comparison of the timing of the EZLRG and the timing of a Business-survey-based composite index based on available business survey data for the same four countries. This approach will enable us to discern which of the two leading indexes has the longest and/or the most reliable leads at peaks, at troughs, and overall. The differences will be reassessed after factoring in the differences in the timing of data availability. This has often been the major advantage which survey-based indicators have in forecasting.

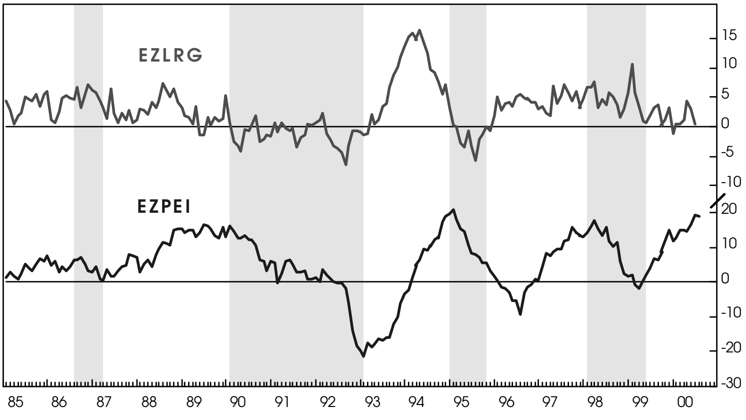
Chart 2 shows the Eurozone Long Range Gauge and a Production Expectations Index based on survey data. We do not have any qualitative indicators available for all four countries for as long a time span as the quantitative data are available. The Eurozone Production Expectations Index (EZPEI) shown in Chart 2 is a GDP-weighted average of the production expectations for the four major countries we are utilizing, all of which are published by the European Commission. This series can be pushed back only to 1985, therefore covering only the last of the three cycles. We call attention to the fact that long range gauges for each country are "level" rather than "change" series. Therefore the Eurozone Long Range Gauge is a "level" series. In order to make a proper timing comparison, therefore, the series dealing with Production Expectations (which is a "change" series inasmuch as it reflects first differences or changes in production expectations from month to month) need to be cumulated before comparing the turns in the Long Range Gauge to turning points in the Production Index. When this is done the production expectations actually lag at turning points.

Table 1. Lead/Lag, Eurozone Long Range and Current Gauges

EZCG Turning Points		EZLRG Turning Points		Lead (-) / Lag (+) in Months		
Troughs	Peaks	Troughs	Peaks	Troughs		Peaks
	7/1966		2/1965			-17
5/1967		12/1966		-5		
			12/1968			extra
		3/1970		extra		
	7/1974		11/1972			-20
6/1975		1/1975		-5		
			1/1976			extra
		5/1977		extra		
	2/1980		12/1978			-14
11/1982		10/1980		-25		
	2/1992		12/1989			-26
7/1993		8/1992		-11		
				Troughs		Peaks
					Overall	
			Average	-12		-19
					-15	
			Median	-8.0		-18.5
					-15.5	
			Percent Lead	100		100
					100	

In the case at hand, the survey results are actually available about one month earlier than the long range gauge. Therefore the differences in leads or lags need to be adjusted by one month to reflect the differences in data availability. In the only business cycle for which this comparison can be made (the latest recession) the production expectations actually lag by six months at the trough and three months at the peak, whereas the EZLRG leads by 11 months at the trough and 26 months at the peak. Therefore the EZLRG leads the Composite of the Production Expectations (cumulated) by 29 months at the peak and 17 months at the trough (Table 2). It is clear, therefore, that even allowing for the one-month gain in data availability, that the EZLRG leads the Cumulated Production Expectations Index by a considerable margin. In that sense, the EZLRG is therefore a more useful forecasting tool for anticipating Eurozone business cycle turning points than is the survey-based index.

Chart 2. Eurozone Long Range Gauge, Growth Rate (%) and Production Expectations Index



Shaded areas represent cyclical downturns in the Eurozone Current Gauge.

Table 2. Lead/Lag, Eurozone Long Range Gauge and Cumulated Eurozone Production Expectations

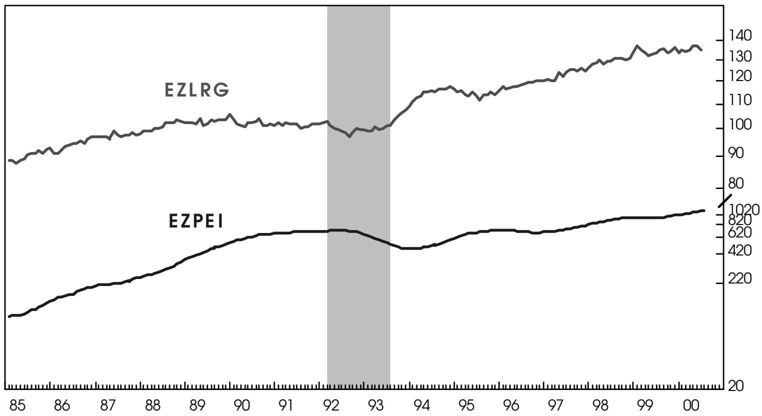
EZPEI (cum.)		EZLRG			
Turning Points		Turning Points		Lead (-) / Lag (+) in Months	
Troughs	Peaks	Troughs	Peaks	Troughs	Peaks
	5/1992		12/1989		-29
1/1994		8/1992		-17	
				Troughs	Peaks
				Overall	
			Average	-17	-29
				-23	
			Median	-17	-29
				-23	
			Percent Lead	100	100
				100	

Table 3. Lead/Lag, Eurozone Long Range and Current Gauges, Growth Rates

EZCG, Growth Rate Turning Points		EZLRG, Growth Rate Turning Points		Lead (-) / Lag (+) in Months		
Troughs	Peaks	Troughs	Peaks	Troughs		Peaks
1/1975		10/1974		-3		
	4/1976		1/1976			-3
9/1977		9/1976		-12		
	5/1979		1/1978			-16
12/1980		4/1980		-8		
	4/1982					miss
9/1982				miss		
			1/1983			extra
		9/1983		extra		
			8/1984			extra
		3/1985		extra		
	7/1986		11/1986			4
3/1987		7/1987		4		
	1/1990		7/1988			-18
1/1993		8/1992		-5		
	12/1994		2/1994			-10
10/1995		7/1995		-3		
	1/1998		1/1999			12
5/1999				miss		
				Troughs		Peaks
					Overall	
			Average	-5		-5
					-5	
			Median	-4.0		-6.5
					-4.0	
			Percent Lead	83		67
					75	

We can do somewhat better in comparing the two types of series at growth rate cycle turns. The series are longer, and therefore, there are more cycles to be compared. In Chart 3 we compare the EZLRG growth rate with the Eurozone Production Expectations Index (EZPEI) in a comparable form. During the period 1985-1999, during which both the EZLRG and the EZPEI are available, the EZLRG growth rate leads the Eurozone growth rate cycle by three months on average at peaks, and by one month on average at troughs, though these leads increase to five months at both peaks and troughs when the data are extended back to 1974 (Table 3). The EZPEI leads the Eurozone growth rate cycle by four months at

Chart 3. Eurozone Long Range Gauge and Cumulated Production Expectations



Shaded areas represent cyclical downturns in the Eurozone Current Gauge.

Table 4. Lead/Lag, Eurozone Long Range Gauge and Production Expectations Index

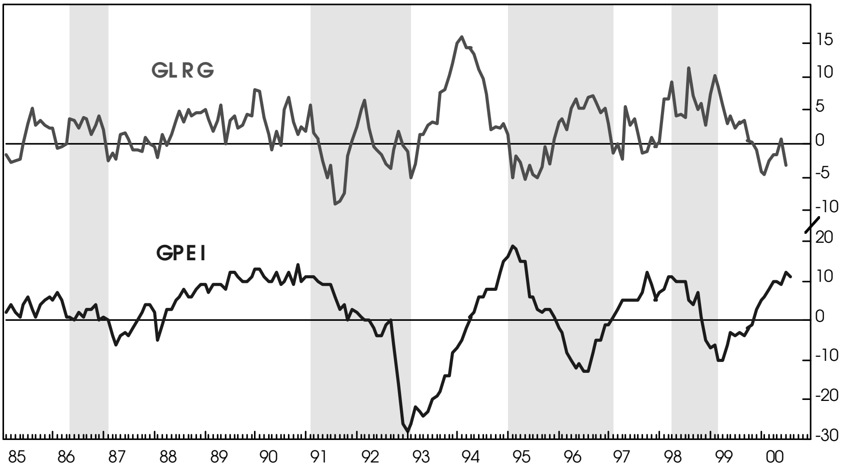
EZPEI, Level Turning Points		EZLRG, Growth Rate Turning Points		Lead (-) / Lag (+) in Months	
Troughs	Peaks	Troughs	Peaks	Troughs	Peaks
	12/1985		11/1986		11
3/1987		7/1987		4	
	6/1989		7/1988		-11
1/1993		8/1992		-5	
	1/1995		2/1994		-11
7/1996		7/1995		-12	
	3/1998		1/1999		10
3/1999				miss	
				Troughs	Peaks
				Overall	
			Average	-4	0
				-2	
			Median	-5.0	0.0
				-5.0	
			Percent Lead	67	50
				57	

troughs and lags three months at peaks, on average. The EZLRG leads the survey-based Production Expectations Index by four months at troughs, coincides at peaks, and leads by two months overall (Table 4). Allowing for the difference in publication dates of the data, the quantitative series adds about a month to the average forecasting ability of the qualitative index.

Predicting German Cycles

Germany is the largest economy in the European Monetary Union, and there is one more test we can perform, this time involving only German data. Just as the Eurozone Production Expectations Index is roughly coincident with Eurozone growth rate cycles, especially in recent years (Chart 3, bottom line), the Production Expectations series for Germany is roughly coincident with the German growth rate cycle (Chart 4). But the German Long Range Gauge (GLRG) growth rate, although exhibiting considerable noise in the earlier years, leads German growth rate cycles, on average, by eight months at peaks and 11 months at troughs (Table 5, Chart 4). This suggests strongly that the GLRG leads by a longer period of time than the production expectations do. These results are duplicated in general for the other three countries we have been considering.

Chart 4. German Long Range Gauge Growth Rate and Production Expectations Index



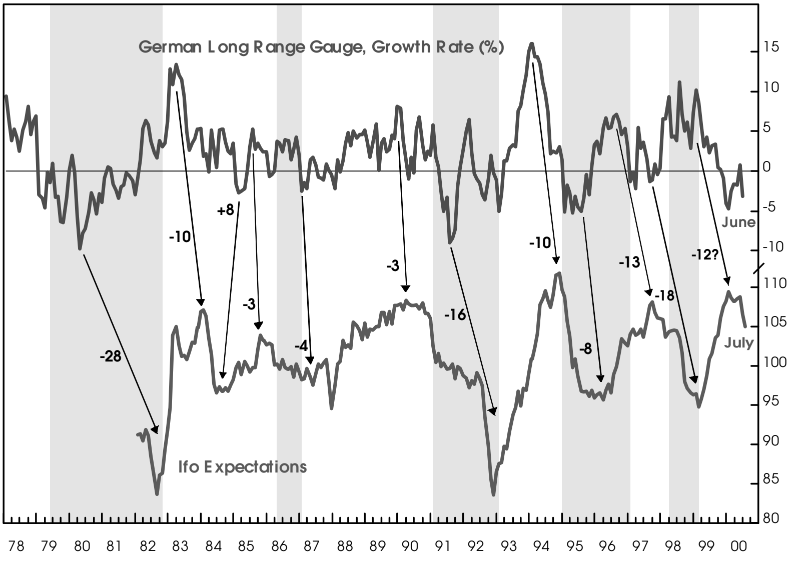
Shaded areas represent German growth rate cycle downturns.

Table 5. Lead/Lag, German Long Range Gauges Growth Rate and Growth Rate Cycle

German Growth Rate Cycle Turning Points		GLRG Growth Rate Turning Points		Lead (-) / Lag (+) in Months	
Troughs	Peaks	Troughs	Peaks	Troughs	Peaks
3/1967		7/1965		-20	
			4/1967		extra
		2/1968		extra	
	1/1969		10/1968		-3
9/1971		2/1970		-19	
	1/1973		2/1972		-11
12/1974		12/1973		-12	
	4/1976		11/1975		-5
7/1977		12/1976		-7	
	5/1979		1/1978		-16
10/1982		4/1980		-30	
			3/1983		extra
		2/1985		extra	
	4/1986		7/1985		-9
1/1987		1/1988		12	
	1/1991		12/1989		-13
1/1993		7/1991		-18	
	12/1994		1/1994		-11
		4/1995		extra	
			8/1996		extra
1/1997		8/1997		7	
	3/1998		7/1998		4
				Troughs	Peaks
				Overall	
			Average	-11	-8
				-9	
			Median	-15.0	-10.0
				-11.0	
			Percent Lead	75	88
				81	

In the case of Germany we can examine the relationship between our Long Range Gauge and expectations in another way, not available for other countries. The Ifo Institute in Munich has long produced a well-known Expectations Index. This is not a production expectations index alone but is rather a measure of general business optimism about the immediate future of the economy as seen by entrepreneurs. As Chart 5 suggests, this index, unlike the Productions Index, rather than coinciding, even roughly, with the growth cycle chronology, leads it instead (Table 6). But Chart 5 makes it clear that the GLRG growth rate is a leading indicator of German business expectations as reflected in the Ifo Expectations Index. The Ifo index in turn, though leading at a number of turning points, has missed several in recent years. The chart shows that GLRG growth has led the Ifo expectations at every turning point except the 1984 trough for the entire period where comparisons are possible.

Chart 5. German Long Range Gauge and Ifo Expectations Index



Shaded areas represent German growth rate cycle downturns.

Table 6. Lead/Lag, German Production Expectations and Growth Rate Cycle

German Growth Rate Cycle Turning Points		GLRG Growth Rate Turning Points		Lead (-) / Lag (+) in Months		
Troughs	Peaks	Troughs	Peaks	Troughs		Peaks
10/1982		8/1982		-2		
			1/1984			extra
		6/1984		extra		
	4/1986		10/1985			-6
1/987		5/1987		4		
	1/1991		3/1990			-10
1/1993		11/1992		-2		
	12/1994		11/1994			-1
1/1997		3/1996		-10		
	3/1998		9/1997			-6
2/1999		2/1999		0		
				Troughs		Peaks
					Overall	
			Average	-2		-6
					-4	
			Median	-2.0		-6.0
					-2.0	
			Percent Lead	70		100
					83	

The Current Outlook

We have thus far concentrated on telling a story in historical terms. It is appropriate to say something about the most recent developments. Chart 4 suggests German production expectations are up quite unambiguously, while GLRG growth is in a downtrend. This suggests that optimism about the immediate future may be less warranted than the production index suggests. The ambiguity is reflected in Chart 5.

The GLRG growth rate is in a clear downtrend while the Ifo Expectations have been in a strong uptrend. It is only very recently that the Ifo Expectations have started declining. The latest data on German industrial production (June saw the biggest monthly drop in five years) is beginning to reinforce the picture suggested earlier by the German Long Range Gauge – and in fact, the latest Ifo Expectations index numbers are now suggesting the possibility of a cyclical downturn in German growth, in line with earlier indications of weakness from the German Long Range Gauge. If so, the Long Range Gauge would once again have provided early warning of directional changes in the survey data, and even earlier warning of a turning point in the growth rate of the actual economy.

Postscript

With the benefit of hindsight, two years after the paper was presented at the Paris conference, it is clear that the GLRG behaved precisely as expected. Specifically, in the sense of out-of-sample performance, it provided "early warning of directional changes in the survey data, and even earlier warning of a turning point in the growth rate of the German economy" in line with the hope expressed in the concluding paragraph of the paper. At the time of the conference, the general expectation was that the German economy would remain strong in 2001, but the GLRG strongly disagreed. In fact, the German economy entered a business cycle recession two months later, in January 2001.

Reference

Bry, G. and C. Boschan, 1971. *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, Technical paper 20, National Bureau of Economic Research, New York.

Towards the Elaboration of Business Climate Indicators for the Italian Economy

Paolo Carnazza and Giuseppe Parigi

Abstract

The paper concerns the cyclical evolution of the Italian economy, with special regard for the role of confidence indicators. In the first part of the paper various confidence indicators, computed according to the European Commission methodology, are investigated. In particular, the analysis concentrates on the dynamics of some business climate indicators referred to the "supply side" of the economy (i.e. manufacturing, retail and construction industries).

In the second part we build up new confidence and business climate indicators exploring the wide informative set characterising the ISAE business surveys. The ability of these indicators in predicting the short-term evolution of GDP is compared with that of the previous European Commission confidence indicators. Finally, in the third part, we present some estimates of the relationship between the business confidence in Italy and some driving variables.

1 Introduction

Confidence indicators, based on the aggregation of qualitative series derived from consumer and business surveys, have significantly grown in importance in the recent past. The aggregation of various series, generally chosen on the basis of the National Bureau Economic Research (NBER) methodology, into a single, synthetic indicator can be justified either on theoretical or on empirical grounds.

Since most economic decisions are influenced by agents' opinions and expectations business surveys are particularly relevant as emerges from the analyses of the European Commission (EC, hereafter): "….. *in the business surveys firms indicate what their expectations are. In the very short term firms' expectations are realised. For example, when the head of a firm states that he is going to increase output, he does so on the basis of an objective assessment of his order books. Furthermore, he has also generally taken the technical measures necessary to ensure that output does in fact rise. If he has assessed the general economic situation wrongly and if he subsequently notes a full-off in demand, he will be compelled to revise his production plans progressively although the increased output indicated in the business survey will have been achieved initially*" (EC, 1991, page 5). In other terms, by assessing the importance of the psychological "mood" and of expectations in influencing the human behaviour, the combination of different qualitative series (opinions and expectations) can be mainly justified on the basis that, very often, economic agents' opinions cannot be considered isolated by expectations and *vice versa*.

The aggregation of different series into a synthetic indicator may partially reduce the problems of "false signals" of turning points and of the volatility of the series due to seasonal and accidental factors. The superiority of the synthetic indicator - with respect to that of its single components - in forecasting economic activity has been shown by many empirical analyses (for Italy, see Schlitzer, 1993; and Carnazza and Oneto, 1996). The EC - which is in charge of the harmonised business and consumers surveys for the countries in the European Union - regularly computes a set of confidence indexes for the single countries and some aggregate areas, such as the Euro area and the whole EU, whose predictive ability has been recently assessed (EC, 1997).

In this paper we concentrate exclusively on business confidence indicators for the Italian economy computed according to different procedures and using three different surveys. In the first section the characteristics of the confidence indicators proposed by the EC methodology are analysed. Then the three confidence indexes related to the manufacturing, the retail trade and the construction industries are aggregated into what we call a business climate (BC hereafter) indicator. The indicators ability to predict the short-term evolution of GDP is then examined according to different statistical approaches: correlation analysis; Granger causality tests; some statistical indicators derived from the regression of GDP growth on its past values and on the past values of each confidence indicator in turn; and, finally, the cyclical behaviour of the chosen synthetic indicator around the turning points of the reference series.[1]

The EC approach has, however, a drawback insofar as it is based only the aggregation of on a few series. To overcome this limit, in the third section, we explore the wide informative set of the ISAE business surveys and try to build up new confidence and BC indicators. By following the same statistical procedures as before, the predictive ability of several indicators is compared with that of the previous EC confidence indexes, and with that of the leading indicator for the Italian economy jointly elaborated by ISAE and the Research Department of the Bank of Italy (see Altissimo et al., 1999).

In the next section given the good results on the predictive ability of our proposed BC indicator we investigate whether it may be interpreted only as a synthesis of macroeconomic forces which can be traditionally measured or its informative content is wider, with a significant role for psychological factors. To do this we perform a regression analysis of the BC indicator on a set of potentially driving factors.

The last section draws some conclusions and discusses some suggestions for further research.

---

1 It should be observed however that to anticipate a turning point is conceptually different from the minimisation of the mean squared forecasting error. On this regard a leading indicator may prove to be useful for one purpose but not for the other.

2 The EC Confidence Indicators: An Application for the Italian Economy

At the beginning of nineties, the EC elaborated some confidence indicators for the whole EURO-11 area and for each European economy (EC, 1991). These confidence indicators are quite simple since they are based on a few harmonised series. More specifically, they concern the following items:

- the manufacturing confidence indicator, defined as the arithmetic mean of the answers to the questions on production expectations, opinions on order books and on the level of the stocks of finished products (the latter with the negative sign);
- the retail trade confidence indicator, defined as the arithmetic mean of the answers to the questions on the present and the future business situation and the opinions of the level of stocks (with negative sign);[2]
- the construction confidence indicator expressed as the arithmetic mean of the answers to the questions on the level of order books and on the expected evolution of employment.

The choice of the components of each confidence indicator is influenced by the restrictive information set used by the EC and is carried out on empirical grounds.

The EC quantifies the qualitative answers to the questions by using the balance between positive and negative replies on a seasonally adjusted basis.[3] Alongside these indicators the EC provides also the consumer confidence indicator[4] and the so-called "Economic Sentiment indicator", given by the aggregation of the manufacturing and the consumer indicators (with equal weight), the construction indicator and the share-price index (with half the weight of the former two). According to the EC approach, the Economic Sentiment indicator "…. s*hould combine judgements and attitudes of the principal actors in the economic process, the producers, consumers and investors* [….] *It was hence thought appropriate to add the attitudes of financial investors, as they are reflected in the share price index*" (EC, 1997, pag.26).

The combination of the consumers and business confidence indexes along with the share price index does not seem to be soundly based. Moreover, its usefulness is not adequately supported by the empirical evidence. In this paper we have preferred to confine

---

2 This confidence indicator was implemented by the EC since May 1997. A detailed analysis of its statistical properties is contained in EC (1997).

3 The EC uses a specific method for seasonal adjustment called "Dainties", originally developed by Eurostat, synthetically described in EC (1997). It is worth noting that the EC does not provide the raw series so that it is not possible to check the results of the seasonal adjustment procedures.

4 This indicator is defined as the arithmetic mean of the (seasonally adjusted) replies to five questions: the assessment of the households' financial situation in the last 12 months and the prospects for the next 12 months; the assessment of the general economic situation in the last 12 months and the prospects for the next 12 months; the advisability of making major purchases of durable consumer goods.

ourselves to the supply side of the economy by building an aggregate BC indicator for the Italian economy based on the manufacturing, the retail trade and the construction confidence indicators. We have therefore disregarded the consumers, or the demand side, index for the time being. The relationship between the BC indicator and the consumers confidence index is however analysed in section four.[5]

Our procedure is the same as the EC one apart from the fact that by exploiting the whole ISAE data set we start from the raw data (suitably transformed into indices, 1995=100) and seasonally adjust them by adopting the Tramo-Seats technique.[6]

In order to relate the various confidence indicators to the evolution of the Italian GDP, the monthly qualitative series should be transformed into quarterly series. To do this we have followed two approaches: the more common approach, generally adopted by the EC and other institutions, based on the arithmetic mean of the three months in the quarter; an alternative one based on the three-term moving average centred on the last month of the quarter. This last choice stems from the assumption that the agents' assessments of the economic evolution (their own and of the whole economy) is essentially based on the most recent past; in other terms, economic agents are assumed to exploit the largest informative set in taking their decisions. This is particularly true - as we shall see later on - for the questions in the retail and the construction surveys.

The final step corresponds to the computation of some aggregate business climate indicators, obtained as the average of the three single confidence indicators, weighted by the their shares of the value added at factor costs in 1995. More specifically, we built three different business climate indicators: one given by the usual monthly to quarterly series aggregation (arithmetic average); for the second one the aggregation is based on the three-term moving average centred on the last month of the quarter; the third one is given by a mixture of the two aggregation procedures: the arithmetic average for the manufacturing confidence indicator and the three-term moving average, centred on the last month of the quarter, for the retail trade and the construction indicators.

The indicators' ability to predict the GDP growth in the short period is assessed through some statistical approaches:[7] the correlation analysis; Granger causality tests; some goodness of fit indicators from a regression of GDP growth on its past values and on the past values of each confidence indicator in turn.

Over the period from the first quarter of 1986 to the fourth quarter of 1999 the manufacturing and the retail trade indicators show a strong correlation with GDP growth (0.69 and 0.74, respectively, at one quarter ahead, Tab.1), while for the construction confidence indicators the results are disappointing. On the other hand, the BC indicators

---

5 On the poor results of the combination of the two indexes see Strighel (1988), p. 15.

6 The Tramo-Seats procedure did not detect any seasonality in the series from the retail trade and the construction surveys. On the reverse, the series from the manufacturing survey have shown a marked seasonality, with the exception of the short-term expectations on the evolution of the economy.

7 Santero and Westerlund (1996) and EC (1997) have found some evidence for the empirical relationship between the confidence indicators and the output components for Italy and other European countries.

seem to be characterised by fairly high values of contemporaneous and one quarter ahead correlation with GDP (0.85 and almost 0.80, respectively, for the three versions of the BC indicators).

Table 1. Correlation between EC confidence indicators and GDP (*) (Period: 1986.1 – 1999.4)

	GDP_{t-3}	GDP_{t-2}	GDP_{t-1}	GDP_{t}	GDP_{t+1}	GDP_{t+2}	GDP_{t+3}
Manufacturing Conf. Indicator (A)							
(1)	0.44	0.63	0.77	0.77	0.69	0.52	0.31
(2)	0.51	0.69	0.78	0.76	0.63	0.44	0.24
Retail trade Conf. Indicator (B)							
(1)	0.35	0.53	0.68	0.76	0.74	0.67	0.52
(2)	0.41	0.58	0.72	0.76	0.70	0.63	0.47
Construction Conf. Indicator (C)							
(1)	-0.03	0.16	0.35	0.52	0.58	0.62	0.56
(2)	-0.02	0.22	0.41	0.54	0.60	0.61	0.52
(A+B+C)							
(3)	0.37	0.59	0.78	0.85	0.82	0.71	0.52
(A+B+C)							
(4)	0.44	0.65	0.81	0.85	0.78	0.64	0.44
(A+B+C)							
(5)	0.40	0.62	0.80	0.85	0.80	0.68	0.48

Notes. (*) GDP is expressed as quarterly to quarterly growth.

Legend. (1) The transformation from monthly to quarterly series is expressed as average mean. (2) The transformation from monthly to quarterly series is expressed as three-term moving average centred on the last month of the quarter. (3) Average mean (weights are given by added values at 1995 factor prices). (4) The transformation from monthly to quarterly series is expressed as three-term moving average centred on the last month of the quarter. (5) The transformation of series A from monthly to quarterly is expressed as average mean; the transformation of series B and C from monthly to quarterly series is expressed as three-term moving average centred on the last month of the quarter.

According to the Granger causality test only the manufacturing confidence indicator seems to Granger-cause GDP, while the other two confidence indicators reveal somewhat mixed results (Tab. 2). In particular, no Granger-causality could be detected for the construction confidence indicators. As to the three BC indicators, Granger-causality was found only for the first and the third version.

Table 2. GDP growth and confidence indicators

Indicators	Granger causality tests: Null hypothesis: Indicator does not Granger cause GDP (% p-value)	Null hypothesis: Indicator Granger causes GDP (% p-value)	Regression analysis: R^2	Adjusted R^2	Standard error of the regression
		Manufacturing Confidence Indicator (A)			
(1)	0.226	53.781	0.360	0.238	0.00526
(2)	1.787	27.590	0.285	0.149	0.00555
		Retail trade Confidence Indicator (B)			
(1)	4.501	4.704	0.248	0.105	0.00570
(2)	1.322	0.317	0.297	0.163	0.00551
		Construction Confidence Indicator (C)			
(1)	62.398	45.448	0.112	-0.0569	0.00619
(2)	15.489	57.491	0.193	0.0387	0.00590
		A+B+C			
(3)	0.814	35.391	0.315	0.184	0.00544
		A+B+C			
(4)	0.182	3.160	0.367	0.246	0.00523
		A+B+C			
(5)	0.110	12.417	0.383	0.266	0.00516

Notes. (*) GDP is expressed as quarterly to quarterly growth.

Legend. (1) The transformation from monthly to quarterly series is expressed as average mean. (2) The transformation from monthly to quarterly series is expressed as three-term moving average centred on the last month of the quarter. (3) Average mean (weights are given by added values at 1995 factor prices). (4) The transformation from monthly to quarterly series is expressed as three-term moving average centred on the last month of the quarter. (5) The transformation of series A from monthly to quarterly is expressed as average mean; the transformation of series B and C from monthly to quarterly series is expressed as three-term moving average centred on the last month of the quarter.

The forecasting ability of the indicators is gauged through the $\overline{R}^2$ from a regression of the GDP on its four lags and four lags of the confidence indicators over the period 1987.2-1999.4 (Tab. 2). Given the previous results it comes to no surprise that the best performing indicator is that of the manufacturing sector. Among the BC indicators the best predictor of GDP is the third version, depicting the highest value of $\overline{R}^2$. In this context the application of ordinary least squares has been found to yield fairly satisfactory results, especially with respect to more complicated non linear procedures (see Hamilton and Peres-Quiroz, 1996; Diebold and Rudebusch, 2000). In our analysis we have used he first difference filter for the confidence indicators. This may appear unjustified, as these series should be dominated by short-term cyclical factors. Some preliminary analyses have however shown that the evolution of all indicators aver the estimation period is clearly characterised by a sort of trend, thus justifying the use of first differences. This is not an unusual result for the

Italian economy; Parigi and Locarno (1997) have found a similar pattern for the consumers' confidence index as well.

Synthetically, the chosen confidence variable – based on the EC approach – is built as a combination of three specific indicators characterised by different aggregation procedures. There is no *a priori* theoretical or empirical reason, which can lead to the choice of a specific aggregation method. Our results show that for the manufacturing survey the more appropriate procedure seems to be the average mean of the three months in the quarter, while for the other two surveys the three terms moving average centred on the last month of the quarter is to be preferred.

3 The ISAE Confidence Indicators

The aim of this section is to build up new confidence and BC indicators for the Italian economy considering the wide informative set of the ISAE business surveys and the results of some recent studies carried out at ISAE (see the set of leading indicators proposed for private consumption and equipment investment by Carnazza and Oneto, 1996; Carnazza, 1998a).

Accordingly, for the manufacturing sector, we have computed six different confidence indicators by combining the (seasonally adjusted) replies to questions on the level of order books, production, the general trend of the economy, selling prices expectations and the stocks of finished products (with inverted sign). We have also considered a specific measure of the degree of uncertainty[8], but its performance proved to be unsatisfactory (for a more detailed description of the confidence indicators, see the Appendix).

For the retail trade sector we employed the leading indicator proposed by Martelli (1997) and different combination of the replies to some questions of the ISAE survey. Finally, for the constructions sector a wide range of questions have been used, but the results have always been totally disappointing.

The general BC indicators were then computed by aggregating the different confidence indicators by using the weights computed according to the share of their value added at factor costs in 1995. As before, the monthly series have been seasonally adjusted with the Tramo-Seats procedure.

The Granger-causality tests show that all manufacturing confidence indicators Granger-cause GDP (Tab. 3). It is interesting to observe that Granger-causality does not hold for the two confidence indicators obtained by considering the proxy for degree of uncertainty. Similar results hold for the retail trade confidence indicators while for the constructions sector no sign of Granger-causality have been found.

The components of the BC indicators were chosen among those indicators showing the highest forecasting power of GDP growth (Tab. 3). Three indicators were built – see

8 This indicator – expressed as the inverse of the share of firms, which have indicated stable and favourable expectations on the short-term evolution of production – has been successfully used by Carnazza (1998b) as a leading indicator for equipment investment in Italy.

Appendix – due to the different method of temporal aggregation of the components. The best in terms of predictive ability proved to be the one given by aggregating the type "F" manufacturing confidence indicator (the mean of the replies given by the industrial firms to three questions: on order books opinions and expectations and on the stocks of finished products; the transformation from monthly to quarterly series is obtained as the average of the three months in the quarter) and the retail trade confidence indicator type "I" (the mean of the replies to three questions: on employment and order-books expectations and on the stocks of finished products; the transformation from monthly to quarterly series is obtained as the three-term moving average centred on the last month of the quarter). The constructions confidence indicator was not used given its poor statistical results.

Table 3. GDP growth and ISAE confidence indicators

	Granger causality tests		Regression analysis		
Indicators	Null hypothesis: Indicator does not Granger cause GDP (% p-value)	Null hypothesis: Indicator Granger causes GDP (% p-value)	R^2	Adjusted R^2	Standard error of the regression
Manufacturing Confidence Indicator (A)					
(1)	3.109	23.914	0.263	0.123	0.00564
(2)	7.452	42.342	0.226	0.079	0.00578
Manufacturing Confidence Indicator (B)					
(1)	1.516	54.944	0.291	0.156	0.00553
(2)	2.693	66.082	0.269	0.13	0.00562
Manufacturing Confidence Indicator (C)					
(1)	1.181	35.613	0.301	0.168	0.00549
(2)	3.145	46.844	0.263	0.122	0.00564
Manufacturing Confidence Indicator (D)					
(1)	21.155	45.826	0.176	0.020	0.00596
(2)	33.609	12.205	0.151	-0.011	0.00605
Manufacturing Confidence Indicator (E)					
(1)	20.049	29.854	0.179	0.023	0.00595
(2)	53.554	4.011	0.122	-0.045	0.00615
Manufacturing Confidence Indicator (F)					
(1)	0.143	25.237	0.375	0.255	0.00520
(2)	2.664	29.679	0.269	0.13	0.00562

Legend. (1) Quarterly series as average of monthly series. (2) Quarterly series as three-term moving average centred on the last month of the quarter.

Table 3. continued

	Granger causality tests		Regression analysis		
Indicators	Null hypothesis: Indicator does not Granger cause GDP (% p-value)	Null hypothesis: Indicator Granger causes GDP (% p-value)	R^2	Adjusted R^2	Standard error of the regression
		Retail trade Confidence Indicator (G)			
(1)	30.446	74.948	0.157	-0.004	0.00603
(2)	30.967	96.944	0.156	-0.005	0.00604
		Retail trade Confidence Indicator (H)			
(1)	3.107	36.408	0.263	0.123	0.00564
(2)	0.614	3.233	0.325	0.196	0.0054
		Retail trade Confidence Indicator (I)			
(1)	1.980	59.632	0.281	0.144	0.00557
(2)	0.209	9.143	0.362	0.241	0.00525
		Retail trade Confidence Indicator (L)			
(1)	6.943	18.271	0.229	0.0821	0.00577
(2)	1.711	1.934	0.287	0.0.151	0.00555
		Construction Confidence Indicator (M)			
(1)	20.831	21.370	0.178	0.021	0.00596
(2)	10.261	14.606	0.212	0.062	0.00583
		Construction Confidence Indicator (N)			
(1)	13.625	13.194	0.199	0.047	0.00588
(2)	11.223	4.026	0.208	0.057	0.00585
		Construction Confidence Indicator (O)			
(1)	62.398	45.448	0.112	-0.057	0.00619
(2)	15.489	57.491	0.193	0.039	0.00590
		Business Climate Indicators			
F+I+O (3)	0.157	21.742	0.372	0.252	0.00521
F+I+O (4)	0.036	37.565	0.417	0.306	0.00502
F+I (5)	0.026	61.187	0.427	0.318	0.00497

Legend. (1) Quarterly series as average of monthly series. (2) Quarterly series as three-term moving average centred on the last month of the quarter. (3) Quarterly series F and O as average of the monthly series. Quarterly series I as three-term moving average centred on the last month of the quarter. (4) Quarterly series F as average of the monthly series. Quarterly series I and O as three-term moving average centred on the last month of the quarter.(5) Quarterly series F as average of the monthly series. Quarterly series I as three-term moving average centred on the last month of the quarter.

Although the forecasting performance of the composite BC indicator appears to be slightly better (Tab. 4), there is a great similarity with the indicator built on the basis of the EC methodology: the cyclical evolution of the two variables is practically the same with respect to GDP growth (Fig. 1, 2).

Figure 1. EC Business Climate Indicator

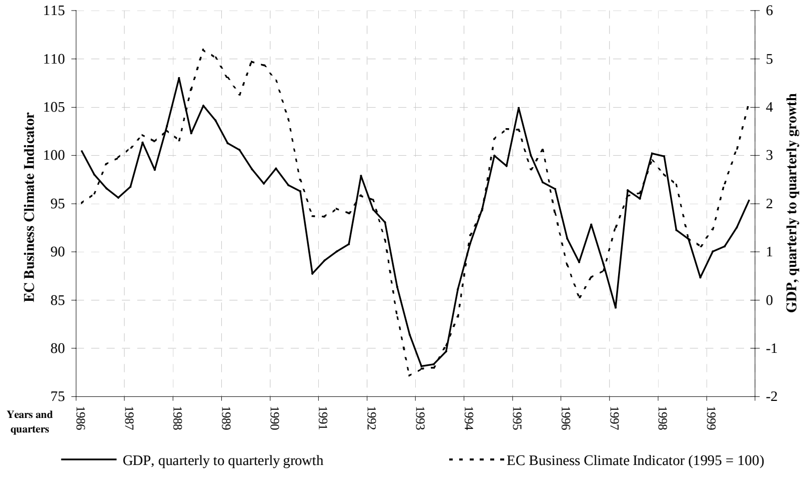


Figure 2. ISAE Business Climate Indicator

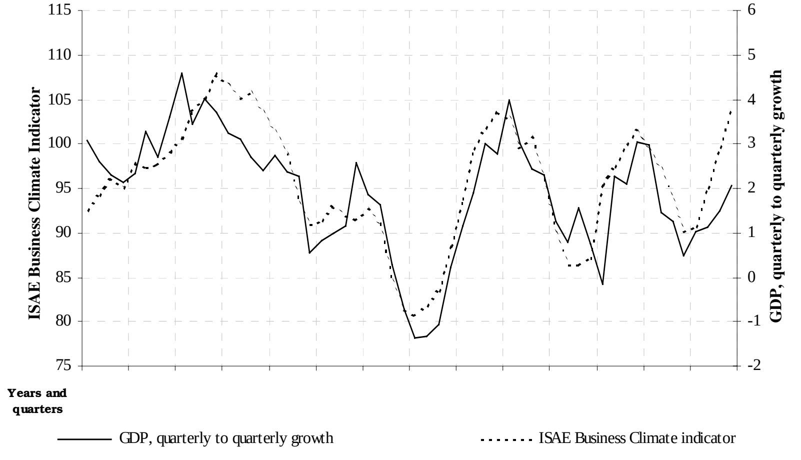


Table 4. Comparison between the best EC and the ISAE confidence indicators

Indicators	R^2		Adjusted R^2		Standard error of the regression	
	EC	ISAE	EC	ISAE	EC	ISAE
Manufacturing Confidence Indicator	0.360	0.375	0.238	0.255	0.00526	0.00520
Retail trade Confidence Indicator (1)	0.297	0.362	0.163	0.241	0.00551	0.00525
Construction Confidence Indicator	0.193	0.212	0.0387	0.062	0.00590	0.05830
Business Climate Indicator (2)	0.383	0.427	0.266	0.318	0.00516	0.00497

Notes. (1) The EC retail trade confidence indicator doesn't Granger explain GDP, while the ISAE version does.
(2) The ISAE indicator does not include the construction confidence indicator.

The conformity of the cyclical components of the two indicators is also confirmed by the analysis based on the Bry-Boschan (1971) methodology (Tab. 5).[9] Both series are characterised by the same number of cycles (from trough to trough) and of turning points; the same average length of cycles; a fairly high correlation with GDP (0.85 for EC Business Climate indicator at one quarter ahead, 0.79 for ISAE Business Climate Indicator two quarters ahead); a satisfactory ability in anticipating the turning points of GDP in the nineties (Tab. 5).

The results of the analysis for the BC indicators are even reinforced when compared to the evolution of a recently elaborated leading indicator for the Italian economy (see Altissimo et al., 1999). Although based on a higher number of series[10] and on a more elaborated methodology the leading indicator performance appears to be very similar to that of the BC indicators (Tab. 5).

---

9 This analysis is based on the concept of "growth cycle" instead of the "classical cycle" (for a more general discussion on the difference between the two concepts see Schlitzer, 1993). In order to de-trend the series we have estimated a polynomial trend following the procedure implemented in Altissimo et al. (1999).

10 The leading indicator includes nine series; production expectations, level of domestic orders of consumption goods; stocks of finished goods; the households confidence indicator; the hours of wage supplementation fund in manufacturing; bank deposits deflated by the index of consumption prices; the spread between interest rates on private and public debt; total merchandise imports at constant prices; the German industrial production index.

Table 5. Cyclical features of GDP, the EC and the ISAE business climate indicators and the Leading indicator (Growth cycles; 1986.1-1999.4)

	GDP	EC business climate indicator	ISAE business climate indicator	Leading indicator
Number of turning points	9	7	7	9
Number of cycles (from trough to trough)	4	3	3	4
Average length of cycles	37	33	33	35
Average length of an expansion	20	16	16	18
Average length of a recession	17	17	17	17
Correlation with GDP (*)	1	0.85 (1)	0.79 (2)	0.79 (2)
Turning points				
Trough	87.1			87.2
Peak	89.3			89.3
Trough	90.4	90.4	90.4	90.4
Peak	92.1	91.4	91.4	91.4
Trough	93.3	93.2	93.1	93.2
Peak	95.2	95.1	94.4	94.4
Trough	96.4	96.3	96.3	96.3
Peak	97.4	97.4	97.4	97.4
Trough	99.2	99.1	99.1	99.1

Notes. (*) The highest values of correlation coefficients are reported at the correspondent lead (-) a lag (+).

4 The Main Factors Affecting the ISAE Business Climate Indicators

In this section we concentrate on the ISAE business climate indicator which has been preferred to its EC counterpart on the basis of its better forecasting performance. Given the results of the previous sections - in particular the forecasting ability of GDP growth – a further interesting point is to understand whether the BC indicator may be interpreted only as a good synthesis of traditional cyclical forces or its information content may be considered wider, covering phenomena only indirectly related to short term dynamics. For this kind of analysis we refer to the literature dating back to the early sixties, where confidence indicators, especially consumers confidence indicators, were regressed on a set of macroeconomic quantitative variables. Locarno and Parigi (1997) provide a detailed survey of this literature, as well as an application to the determinants of the Italian consumers' confidence index. Their main result is that the index cannot be uniquely explained by macroeconomic variables, as suggested by the high significance of some "social" variables included in the regression. In other terms, "*it is necessary to rely on the people themselves to provide information on how precipitating circumstances and* *information received have influenced their attitudes and therefore their responses*" (Katona, 1977, p. 101). In the case of the BC indicator we have taken into account three groups of variables related to cyclical factors, general climate conditions and economic policy decisions. In the first group we have included the industrial output index (PROD); the total number of hours in the wage supplementation fund (CIG); the total number of overtime hours in the manufacturing sector (OHOURS). We have also considered a measure of competitiveness (COMPET).[11] In the second group we have the manufacturing confidence index of the countries in the euro area (excluding Italy; EUBC); the ISAE households confidence index (CLIMA). The third group includes the discount rate (TASC)[12] and the public debt to GDP ratio (DEBT).

The estimation technique has followed the general-to-specific approach. Some dynamic terms of the dependent variable have been considered in order to take account of possible inertia effects in the agents' behaviour or mood. As shown in table 6 all estimated coefficients have the *a priori* expected sing and are highly significant. The fit is satisfactory as evidenced by the low value of the standard error of the regression. The specification does not present particular problems and is fairly stable, as the Chow test and the recursive residuals (not reported) suggest.

As a general comment it appears that the cyclical variables are not sufficient by themselves to track the evolution of the BC indicator. They can be fruitfully integrated with general climate and economic policy variables. In the first case the positive and contemporaneous link between the BC indicator and the consumer confidence index suggests that the two series have followed the same pattern over the estimation period. In some sense this result supports our decision of not combining the two indexes. More recently some evidence at the euro area level presents an opposite picture, with a negative correlation between the two indexes.[13] According to some preliminary estimates a weakening of the link seems to have occurred in the last three years, but it is too early to say a more definitive word on this problem.

The positive influence of EUBC may be interpreted as a sign of the growing integration among the cyclical evolution of the countries in the euro area. The higher interdependence of our economy may also help explaining the negative link between BC and the degree of competitiveness (COMPE). It is interesting to observe that the estimated coefficient in this case is higher starting from 1996, when the lira rejoined the EMS agreements and the exchange rate with the European partners became virtually fixed.

---

[11] The degree of competitiveness coincides with the real effective exchange rate computed on the basis of manufacturing output prices of main Italian trade partners. Positive values of the index mean competitive losses.

[12] TASC is given by the official discount rate until December 1998; the European Central Bank policy reference rate from January 1999.

[13] See the monthly Bulletin of ECB (page 22-23).

Table 6. Confidence and Macroeconomic factors
Dependent Variable: ISAE BC
(Estimation period: 1986.2 – 1998.4)

Variables	Coefficients	White t-statistics
CONSTANT	123.92	4.782
BC_{t-1}	0.471	4.374
$MAVE(DEBT,3)_{t-3}$	-0.387	-4.415
$TASC_{t-2}$	-0.969	-4.997
$MAVE(\Delta(PROD), 9)_{t-6}$	0.270	4.032
CIG_t	-0.054	-2.447
$MAVE(\Delta(OHOURS), 4)_{t-1}$	0.108	3.135
$\Delta(EUBC)_t$	0.537	10.010
$\Delta(CLIMA)_t$	0.163	4.045
$COMPET_t$	-0.224	-4.315
$COMPET_t$*DU961	-0.036	-4.263

$\overline{R}^2$ = 0.97

S.E.(%) = 1.24

Misspecification tests
(p-value in parentheses)

Autocorrelation			Unit root test on residuals			Heteroscedasticity		
DW	2.34		ADF	-5.46		$ARCH_{1-4}$	0.74	(0.95)
LM_{1-4}	0.86	(0.50)						
LB_4	3.23	(0.52)						
			General specification			Predictive power		
			RESET	0.42	(0.66)	CHOW	0.72	(0.58)

Legend: S.E., regression standard error; DW, Durbin-Watson test; LM_{1-4}, Lagrange multiplier test for residual autocorrelation of order 1 to 4, F(4,36); LB_4, Lijung-Box test for residual autocorrelation of order 1 to 4, $\chi^2(4)$; CHOW, Chow test of predictive power over the period 1999.1-1999.4, F(4,44); RESET, functional form test F(2,38); $ARCH_{1-4}$, autoregressive conditional heteroscedasticity test up to the 4th lag, $\chi^2(4)$; ADF, augmented Dickey-Fuller test (10% critical value: -5.0); MAVE(*x,y*) is the (uncentered) *y* terms-moving average of *x*; DU961 is a *dummy* variable with 1 from 1996.1 on. All variables, except for TASC and DU961, are (seasonally adjusted) index numbers, 1995=100.

The relationship of the BC indicator with the variable related to policy decisions on one hand confirms the results obtained by Locarno and Parigi (1997) for the consumers confidence index; on the other presents some new and interesting features. In the first case, the negative sign of the estimated coefficient of TASC – whose changes may be interpreted as the announcement of the monetary policy stance - may represent how such policy decisions affect entrepreneurs expectations. In our case, a restriction has a negative effect on expectations. In the case of the debt to GDP ratio the negative and highly significant coefficients suggests that Italian entrepreneurs may have been influenced by the poor quality of public finances. In other terms, a high public debt might have generated expectations of higher interest rates and of a more restrictive fiscal stance. On this regard this result may also indicate that the recent reduction of interest rates and of the public debt to GDP ratio should support business confidence. However a more accurate analysis is needed to assess the effective role of DEBT in this context, for instance by considering income and the expenditure components separately.

5 Concluding Remarks

The paper compares – for the Italian economy – the predictive power of confidence indicators based on the EC methodology and new confidence and BC indicators built up by exploring the wide informative set of the ISAE business surveys. This comparison has been developed trough different approaches: correlation analysis; Granger causality tests and some other statistical indicators.

The main results point out that the ISAE business climate indicator – although based on a wider informative set and on some recent contributions – seems to be characterised by a slightly better predictive ability with respect to the EC business climate indicator. Moreover, the short-term forecasting performance of these two indicators has been found to be fairly similar to that of a more sophisticated leading indicator for the Italian economy. This last result is important since it means that the BC indicators can effectively be used as a preliminary estimate of the leading indicator. Their main advantages lay in the fact that they can be easily computed each month after the release of the results of a new survey; in addition, they are not affected by the problem of revisions, which normally characterises the elaboration of most other composite leading indicators. Consequently, such business climate indicators can be very useful to anticipate the short-term evolution of the reference series (here represented by GDP) on the basis of the more recent information.

A natural extension of this paper is the computation of similar indicators for the European Union as a whole and, in particular, for the EURO area. To this end, the variables provided by the harmonised EC surveys could be safely used, if the results for Italy should be confirmed at the euro area level.

Then, the paper – trough a regression analysis – has tried to explore the "main determinants" of the evolution of the ISAE business climate indicator. It has been shown that, apart from a strong correlation with some cyclical variables, our indicator appears to be affected by some policy variables such as the short-term interest rate and the public debt to GDP ratio. From this analysis some interesting policy implications can be drawn: more specifically, policy makers - by reducing the public debt and/or the level of interest rates – may contribute to improve significantly the "climate" of firms with positive effects on their investment programs. Anyway, this topic should be further investigated. Another result of the regression analysis that deserves further investigation is the relationship between the BC and the household confidence index. A deeper analysis could be developed aimed at understanding the cyclical evolution and the main factors affecting the two composite indicators.

References

Altissimo, F., D. J. Marchetti and G. P. Oneto (1999), The Italian Business Cycle: New Coincident and Leading indicators and Some Stylised Facts, Documento di lavoro, ISAE n. 8.

Bry G. and C. Boschan (1971), Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, NBER Technical Paper, n.20.

Carnazza P. and G. P. Oneto (1996), Searching for a Leading Indicator of Household Consumption in the Italian Economy, in Oppenländer K.H. and G. Poser (eds.) Business Cycle Surveys: Forecasting Issues and Methodological Aspects. Selected Papers, presented at the 22nd CIRET Conference, Singapore, 1995. Avebury, Aldershot.

Carnazza, P. (1998a) Il Ruolo degli Indicatori Sintetici nell'Analisi Congiunturale, Rassegna dei lavori dell'ISCO, XV, n. 3, p. 145-170.

Carnazza P. (1998b), A leading indicator of equipment investment demand in the Italian economy, in Oppenländer K.H. and G. Poser (eds.), Business Cycle Surveys: Forecasting Issues and Methodological Aspects. Selected Papers, presented at the 23rd CIRET Conference, Helsinki, 1997. Avebury, Aldershot.

Diebold, F. X. and G. Rudebusch (2000), Business Cycles, Princeton University Press, Princeton, New Jersey.

European Central Bank (1999), Monthly Bulletin, October.

European Commission (1991), The general principles underlying the business surveys, Supplement B, Special edition, July.

European Commission (1997), The Joint harmonised EU Programme of Business and Consumer Surveys, Supplement B, n.6.

Hamilton, J. D. and G. Perez-Quiros (1996), What Do the Leading Indicators Lead?, Journal of Business, 69, p. 27-49.

Katona, G. (1977), Psychological Economics, Elsevier, New York.

Locarno A. and G. Parigi (1997), Clima di fiducia e consumi delle famiglie: movente economico o psicologico?, Ricerche quantitative per la politica economica, Vol. II.

Martelli B.M. (1997), An Attempt in Assessing a Leading Indicator for the Retail Trade in Italy on the Basis of ISCO Survey Data, Paper presented at the 23rd CIRET Conference, Helsinki, 1997.

Santero T. and N. Westerlund (1996), Confidence indicators and their relationship to changes in economic activity, Working Papers, OECD Economic Department n. 70.

Strighel, W. H. (1988), Il Clima Economico come Indicatore Anticipatore, Rassegna della letteratura sui cicli economici, ISCO, XXX, n. 1, p.1-39.

Schlitzer G. (1993), Nuovi strumenti per la valutazione e la previsione del ciclo economico in Italia, Temi di discussione, Banca d'Italia, n. 200.

Appendix

Confidence Indicators Definitions

Manufacturing confidence indicators

All variables have been adjusted for seasonal factors with the Tramo-Seats procedure and have been transformed into indices 1995=100. The quarterly aggregation procedures are the simple average of the monthly observations in the quarter and the three terms moving average centred on the last month of the quarter.

A = arithmetic mean of the replies to the questions on order books, production, general economy, selling prices (expectations); stocks of finished products.

B = arithmetic mean of order books, production, selling prices (expectations); stocks of finished products.

C = arithmetic mean of order books and production (expectations); stocks of finished products.

D = arithmetic mean of order books, production, and selling prices (expectations); stocks of finished products; degree of uncertainty. The series have been standardised.

E = arithmetic mean of order books expectations; stocks of finished products; degree of uncertainty.

F = arithmetic mean of order books opinions and expectations; stocks of finished products.

Retail trade confidence indicators

G = arithmetic mean of employment expectations; order books opinions (expressed by manufacturing firms); degree of uncertainty.

H = arithmetic mean of employment, order books and business situation (expectations); stocks of finished products.

I = arithmetic mean of employment and order books (expectations); stocks of finished products

L = arithmetic mean of employment and business situation (expectations); stocks of finished products.

Construction confidence indicators

M = arithmetic mean of order books opinions; order books, employment and activity (expectations).

N = arithmetic mean of order books and activity expectations.

O = arithmetic mean of order books opinions and employment expectations.

Business climate indicators

(1) F + I + O The quarterly version of I is obtained as the three-term moving average centred on the last month of the quarter.

(2) F + I + O The quarterly versions of I and O are obtained as the three-term moving average centred on the last month of the quarter.

(3) F + I The quarterly version of I is obtained as the three-term moving average centred on the last month of the quarter.

Compiling and Comparison of Taiwan's Business Indicators

Cheng-Mount Cheng

Introduction

Good business indicators are like mirrors such that they can truthfully reflect real macro economic situations. They are important references for government to make policy decision and for private firms to determine production and investment. In short, business indicators are key to economic decisions and therefore should be treated carefully in the process of sampling and compiling.

With the development of newest telecommunication system, we can now collect, analyze, and publish economic information in a much faster way. And thanks to various statistic tools, we can classify data according to their characteristics and form different functions of business indicators. In general, there are three types of business indicators. Business cycle indicators are to reflect the direction and fluctuations of economic activities. Monitoring indicators are for government to make policy decisions and business climate indicators are surveys of subjective viewpoints of the economic situations.

In Taiwan, there are two main institutions that publish monthly business indicators. The Council for Economic Planning and Development (CEPD), a government agency, publishes Business Cycle Indicators (including leading and coincident indicators), Monitoring Indicators, and Survey of Firm Operation every month. On the other hand, Taiwan Institute of Economic Research (TIER), an independent not-profit organization, releases monthly Business Climate Index (BCI). The TIER's BCI is based on survey results that TIER conducts on manufacturing firms with a methodology resembling the Ifo's BCI. In this paper we will first describe the method of compiling these indicators and then compare them with one another.

1 The Method of Compiling Business Indicators in Taiwan

1.1 The Monitoring Indicator

Since the monitoring indicator is for government to make policy decisions, the endogenous variables it contains are highly related with the synchronous policy. The CEPD's goals are to maintain domestic economic stability and growth as well as balance of international accounts. Under those objectives the CEPD chooses 9 variables in the content of Monitoring Indicator, which include 4 financial variables and 5 variables from the real side. Financial variables included are variations in money supply M1b , loans & discounts, bank clearings, and stock prices. Five real side variables are variations in manufacturing new orders, exports, industrial production, manufacturing inventory ratio, and non-agricultural employment. In addition, variations in consumer price index and wholesale price index, leading and coincidence indicators, and GDP growth rates are also used as reference. The calculation of Monitoring Indicators are through the following steps:

Step 1: Adjust the 9 variables by seasonal factors, then calculate the variations from the same period of last year. Among other variables manufacturing inventory ratio is equal to inventory divided by total sales.

Step 2: Compare the variations with the critical value and signal them with different lights. Each light represent different scores from 1 to 5.

Step 3: Add the scores together to get an overall score, i.e., the monitoring indicator.

Step 4: Compare the overall score with the criteria and determine lights for the month.

Table 1. Critical Values of the Monitoring Indicator

		Red (5)		Yellow red (4)		Green (3)		Yellow blue (2)		Blue (1)
Variations in financial variables	Money Supply M1b(*)	←	15	—	12	—	7	—	5	→
	Loans & Discounts	←	23	—	20	—	14	—	11	→
	Bank Clearings	←	25	—	17	—	6	—	0	→
	Stock Prices	←	36	—	21	—	0	—	0	→
Variations in real side variables (%)	MFG's New Orders	←	12	—	9	—	3	—	0	→
	Exports	←	15	—	10	—	3	—	0	→
	Industrial Production	←	12	—	9	—	3	—	0	→
	MFG's Inventory Ratio	←	52	—	56	—	67	—	72	→
	Non-agricultural Employment	←	4.0	—	3.2	—	1.6	—	0.8	→
Total Scores	Lights	Red		Yellow red		Green		Yellow blue		Blue
	Policy Suggestion	Tight		Cautious		Safe		Cautious		Expan-sionary
	Critical Values	45-38		37-32		31-23		22-17		16-9

Note:

(*) By definition M1b include currency in circulation, checking deposits, demand deposits, and saving deposits.

1. Figures inside brackets are the score each light represents.

2. The critical values are based on the variations of the 9 variables from 1982 to 1994, taking consideration of the variations between 1991 and 1994 as well as the long-and-medium term targets.

Sources: CEPD, Business Indicators, monthly.

Even though the main purpose of the monitoring indicator is for policy reference, private businesses can also use it to understand the current economic situation as well as possible policy measures of the government in the future and thus readjust their investment plans or business strategies. According to the definition of CEPD, the meaning of different lights and government's corresponding policy are listed in the following table.

Table 2. Lights of Monitoring Indicator and Government's Corresponding policy

Light	Economy	Policy Suggestion
Red	Overheating	Government should take tightening policy to cool down the economy.
Yellow Red	Booming	Take no action
Green	Stable	Take neutral fiscal and financial policy to stabilize and encourage continuous economic growth.
Yellow Blue	Recessed	Watch cautiously but take no action
Blue	Depressed	Take expansionary measures to stimulate the economy.

Sources: CEPD, Business Indicators, monthly.

1.2 Business Composite Index

The CEPD also compiles business cycle indicators in the form of the business composite index, which includes both leading and coincidence indicators. There are 7 variables including in the leading indicator whereas 6 in the coincidence indicator. The compilation of the business composite index follows the method of those adopted by the commerce department of the U.S. and has several steps.

Step 1: Adjust variables by seasonal factors.

Step 2: Calculate the variation ($B_{i,t}$) from the previous month for each variable ($A_{i,t}$). There are two variations. The first one is called symmetrical variation, which is equal to

$$B_{i,t} = \frac{200\times(A_{i,t} - A_{i,t-1})}{(A_{i,t} + A_{i,t-1})}$$ if all $A_{i,t}$'s are greater than zero.

The second variation is simple variation, which is equal to

$B_{i,t} = A_{i,t} - A_{i,t-1}$ if $A_{i,t}$ appears zero or negative.

Step 3: Standardize the variations. The denominator is the moving average of the 60 $B_{i,t}$'s absolute values.

Step 4: After standardization, calculate the weighted average according to their weighs and we get two new sequence of data for leading and coincidence indicators.

Step 5: Standardize the 2 new series, the denominator is the average of the absolute values of each series during some fixed period. The standardized series is called $Z_{j,t}$.

Step 6: Add $Z_{j,t}$ together and get the initial complex index (CI')

$$CI'_{j,t} = \frac{200 + Z_{j,t}}{200 - Z_{j,t}} \times CI'_{j,t-1}, CI'_{j,1} = 100$$

Step 7: Use 1996 as the base year and transform the initial complex index into a new series.

Step 8: Calculate the average K_1 of the first CI' cycle and K_2 of the latest CI' cycle and then calculate the average monthly variation from K_1 to K_2.

Step 9: Add $Z_{j,t}$ of the leading indicator with the average monthly variations of the coincidence indicator.

Step 10: Accumulate the results from step 9 and get a new complex index.

Step 11: Adjust the base year and get the final version of the complex index.

Table 3. Variables in the leading and coincidence indicators and their weights

Leading indicator	Weights	Coincidence indicator	Weights
Variation in MFG's New Orders	57	Variation in industrial production	77
Average monthly working hours in the MFG	54	Variation in MFG production	75
Variation in exports	67	MFG's sale	64
Variation in money supply M1b	81	Average monthly pays from the MFG	61
Variation in WPI	69	Variation in bank clearings	77
Variation in stock prices	77	Domestic shipment	71
Applying areas for construction	61		

According to the statistic's economic importance, statistic sufficiency, cycle correspondence, time consistency, smoothness of the curve, and data's in time and accuracy, the range of the weights are from 1 to 100, the more the variable weighs, the more they can represent the business cycle.

1.3 Survey of Firm Operation and Business Climate Index

The CEPD's Survey of Firm Operation is similar to IFO's Business Test of Germany and Economic Survey of Selected Enterprises of Japan and is conducted in the way of monthly questionnaire asking firm's current running situation and future prospect. The questionnaire contains both qualitative and quantitative questions.

Table 4. Questionnaire of the CEPD

	Qualitative	Quantitative
Question	Ask firms if their running situations are better, same as or worse than the previous month. Variations in production, sales, prices and stocks. Prediction on the prospect in the future 3 months.	Ask firms about the figures of sales and orders.
	Add together the proportion of firms who are better, same or worse in the questionnaire and calculate the Diffusion Index (DI), which is equal to DI=1*(the proportion of better firms)+0.5*(the proportion of firms who are the same)	Add the figure by their industry classification, then calculate the variation from last year, and finally transform the series into an index by compare it with the base year

To understand and to predict the business climate, the Taiwan Institute of Economic Research (TIER) also adopted IFO's method and has conducted monthly survey since 1979. The TIER chooses 1,200 manufacturing firms as the survey sample and asks only questions of quality. The questionnaire contains two parts: 1) ask firms whether their production, inventory, price, and cost are rising, sustaining, or falling during the month (and in the future three months); 2) ask firms whether their businesses are turning better, the same, or worse compared to the previous month and their viewpoints about the future six months.

One difference between the CEPD's Survey on Firm Operation and the TIER's Business Climate Survey is that the CEPD asks firm's prospect in the future three months while the TIER ask firm's viewpoints about the future six months. In addition, the CEPD only calculates the Diffusion Index with the quality data while the TIER combines firm's viewpoints about the business climate during the month and in the future six months to compile the business climate index. The business climate index is the only one index that has been conducted and published regularly by private non-profit organizations in Taiwan. It is compiled through the following steps:

Step 1: Take weighted average of firm's reply. The weight is calculated by dividing the industry's (where the firm belongs) net production value by the total manufacturing production in the latest year.

Step 2: Calculate the net increase of manufacturing firms that express optimistic views.

(1) Net increase of manufacturing firms whose business are better in the month (GA_t) = the proportion of firms whose business are better – the proportion of firms whose business are worse.

(2) Net increase of manufacturing firms that express optimistic views about the future six months (GB_t) = the proportion of optimistic firms – the proportion of pessimistic firms.

Step 3: Take geometric average of GA_t and GB_t .

$$S_{1,t} = \sqrt{(GA_t + 200) \cdot (GB_t + 200)}$$

Step 4: Adjust by seasonal factor.

$$S_{2,t} = \frac{(S_{1,t})}{season\ factor_t}$$

Step 5: take 3 month moving average.

$$S_{3,t} = \frac{S_{2,t-2} + S_{2,t-1} + S_{2,t}}{3}$$

Step 6: Adjust the base year (1991=100).

1991price = $(\sum S_{3,m})/12, m = 1991$ January, February, … , December

$S_{4,t} =$ the business climate index = $(S_{3,t}/1991\,price) \times 100$

2 Comparison of Various Indicators in Taiwan

As mentioned before, a good indicator should be capable of reflecting true economic situation and predicting the future trend at the same time. In the following we use cross-correlation and causality tests to compare the explaining and forecasting power of the leading indicator, the coincidence indicator, the monitoring indicator and the business climate index to the economic growth.

2.1 Cross-Correlation

As table 5 shows, all indicators are positively related with the GDP growth rate. The monitoring indicator has the best performance since it has the highest correlation coefficient with the GDP growth rate. Even with the data led or lagged by one period, the monitoring indicator still has high correlation coefficients. Similarly, the business climate index performs relatively well. The leading indicator ranked third in the cross-correlation test. On the other hand, the coincidence indicator performs very poor. It even shows negative relation with the GDP growth rate when the coincidence indicator led by 1 period.

Table 5. Cross-Correlations

	Leading Indicator		Coincidence Indicator		Monitoring Indicator		Business Climate Index	
	Level	First-degree diff.	Level	First-degree diff.	Level	First-degree diff.	Level	First-degree diff.
K=-4	-0.062	-0.119	-0.321***	-0.311***	0.167*	-0.291***	-0.321***	-0.382***
K=-3	-0.026	-0.209***	-0.304***	-0.266***	0.319***	-0.396***	-0.098	-0.481***
K=-2	0.043	-0.123	-0.287***	-0.210***	0.522***	-0.290***	0.185	-0.536***
K=-1	0.083	0.027	-0.275***	-0.022	0.673***	-0.083	0.509***	-0.355***
K=0	0.069	0.280***	-0.279***	0.358***	0.719***	0.229***	0.723***	-0.029
k=1	-0.033	0.428***	-0.312***	0.412***	0.604***	0.425***	0.731***	0.217*
k=2	-0.186***	0.473***	-0.348***	0.456***	0.387***	0.493***	0.600***	0.404***
k=3	-0.347***	0.443***	-0.380***	0.359***	0.128	0.485***	0.360***	0.529***
k=4	-0.514***	0.224***	-0.419***	0.136	-0.121	0.296***	0.046	0.425***

Note: 1. The above figures are correlation coefficients for GDP_t and $Indicator_{t-k}$, where k=4 means the correlation between the GDP at current time and the indicator 4 periods before.

2. * means significant at 10%;
** means significant at 5%;
*** means significant at 1%.

Sometimes the level value of the indicators may not be able to reflect true economic peak or trough while the variation of indicators in two periods could catch the fluctuation of the business climate. To test this idea we first take the first degree of difference with the variables and then run the same cross-correlation tests. The result shows a little surprise. The coincidence indicator, the poorest performer with the level value, jumps to the top, followed by the monitoring indicator, the business climate index and the leading indicator. All indicators show leading relation with the GDP growth rate. From this we confirm that the first degree of difference of all indicators can really catch the fluctuations of the GDP growth to some extent.

In theory, the leading indicator, coincidence indicator, and the monitoring indicator belong to the so-called hard data. Their functions are mainly to explain and forecast the direction and the fluctuation range of the GDP growth. On the other hand, the business climate index belongs to soft data and can only reflect the direction of the GDP growth. But through the cross-correlation tests we find out that the business climate index also has good explaining and forecasting power of the GDP growth.

The CEPD only calculates the diffusion index from their survey results. To compare the explaining power of the surveys conducted by the CEPD and the TIER, we also calculate the diffusion index from TIER's survey result by using the same formula. Then we run the cross-correlation tests of both surveys to the GDP growth rate and to the industrial production. But the results show that both survey's diffusion indexes do not have a significant relation with either.

2.2 Causality

From the cross-correlation tests we see that all indicators are either roughly leading or lagging behind the GDP growth rate. If there do exist a causality relation, then we can use one variable to predict another variable. In other words, it would be more precise if we could use the past information of the business cycle in the forecast of the future. In the following we run Granger causality tests between all indicators and the GDP growth rate. The two variables in the Granger model are:

$$Y_t = \alpha_1 + \sum_{i=1}^{k_1} \alpha_{1t} \cdot Y_{t-i} + \sum_{i=1}^{m_1} \beta_{1j} \cdot X_{t-j} + \varepsilon_{1t} \qquad \text{(E1)}$$

$$X_t = \alpha_2 + \sum_{i=1}^{m_2} \alpha_{2t} \cdot X_{t-i} + \sum_{i=1}^{k_2} \beta_{2j} \cdot Y_{t-j} + \varepsilon_{2t} \qquad \text{(E2)}$$

where X_{t-i}, Y_{t-j} are leading i period's ; $X, Y; k_1, k_2, m_1, m_2$ are the periods of leading set up in the model. By F-test we could examine the two null hypotheses:

$$H_0 : \beta_{1j} = 0 \quad (j = 1,2,...m_1) \ldots \qquad \text{(E3)}$$

$$H_0 : \beta_{2j} = 0 \quad (j = 1,2,...k_2) \qquad \text{(E4)}$$

If the null hypothesis in (E3) can not be rejected, that means the time series X can not be regarded as the cause of Y, i.e., the past information of X is not helpful in the forecast of Y. Similarly, if the null hypothesis in (E4) can not be rejected, then the past information of Y is not helpful in the forecast of X. On the contrary, if one of the two null hypotheses is rejected, then there exists a unilateral causality relation between X and Y. If both null hypotheses are rejected, the X and Y are affecting each other, i.e., there exists a feedback relation between them. Because Granger causality tests require the time-series data be stationary, we use the Augmented Dickey-Fuller unit root tests to test for the stationarity.

The results show that except for the GDP growth rate all indicators have unit roots. After taking the first-degree difference on all indicators they become stationary. Then we can run the causality tests between the GDP growth rate and the first-degree difference of all indicators. As to the determination of the lagged periods, we follow the principle developed by Akaike and Schwarz and come out with 4 for coincidence and leading indicators and 5 for the monitoring indicator and the business climate index.

Table 6 shows that all indicators are significantly the causes of the GDP growth rate. However, the GDP growth rate is the cause of the business climate index only. This shows that only the business climate index has the feedback relation with the GDP growth rate. The adoption of the first-degree difference on all indicators in the causality tests also once again shows that using the difference between two periods has better explaining or forecasting power than using the level values.

Table 6. Granger Causality Tests

Null Hypothesis:	Coincidence Indicator	Leading Indicator	Monitoring Indicator	Business Climate Index
1. Indicator does not Granger Cause GDP Growth	7.858***	12.915***	7.140***	6.374*
2. GDP Growth does not Granger Cause Indicator	1.775	3.384	4.971	8.434*

Note: * means rejected the null hypothesis at the 10% significance level;
** means rejected the null hypothesis at the 5% significance level;
*** means rejected the null hypothesis at the 1% significance level.

3 Conclusion

In this paper we first introduce how major business indicators are compiled. Then we evaluate the explaining and forecasting power of all indicators by running cross-correlation and Granger causality tests on them. The result shows that the monitoring indicator and the business climate index perform better than leading and coincidence indicators. The first-degree difference of indicators also performs better than the level value. Indeed, the purposes of these indicators are different and not just specific on the forecast of the GDP growth rate. For example, the priority of the monitoring indicator is to reflect government's policy measures; the purpose of leading and coincidence indicators are to catch the turning points of the economic cycle. Nonetheless, from this paper we are able to show that the monitoring indicator and the business climate index are useful in the explanation and forecast of the GDP growth rate and therefore are useful references in the understanding of the business dynamics.

General Indicators of Business Activity for Poland Based on Survey Data

Zbigniew Matkowski

1 Introduction

This paper refers to a broader research on composite indicators of economic activity for Poland, carried out since 1994 at the Research Institute of Economic Development (RIED), Warsaw School of Economics, by a team headed by the author under three successive research projects instituted by the Committee of Scientific Research and the Committee on European Integration. The aim was to develop a system of composite indicators, based on quantitative and qualitative data, which might be used for analysing changes in the aggregate economic activity as well as for monitoring purposes. The full account of results is contained in three volumes of RIED's "Papers & Proceedings" devoted to the subject (Z. Matkowski, ed., 1997, 1998, 1999).

Even in times of the so-called centrally planned system, Poland's economy revealed well pronounced fluctuations quite similar to those observed in the developed market economies, elsewhere known as business vs. growth cycles. Growth fluctuations were mostly attributed to investment cycles and to the five-year span of medium term planning, as well as to exogenous factors, such as fluctuations in export markets or changing weather conditions. Nowadays, with the progress of transformation towards an open market system, Poland becomes increasingly exposed to internal and external fluctuations and shocks typical of a market economy. This justifies our concern about the development of a monitoring system that could help us to evaluate the state and prospects of national economy. For this purpose, we have developed three types of composite indicators.

The first one is a composite indicator of aggregate economic activity, denoted GCI (general coincident indicator). It is a weighted average of indices reflecting production or sales volume in five major sectors of economy: industry, construction, agriculture, transport, and trade, weighted by their yearly shares in GDP (till 1989, in GMP). The index has been compiled on a monthly basis for the period starting in January 1975; after the last updating the time series has been brought till April 2000. This index gives a good indication of month-by-month changes in total domestic output (quarterly GDP data are only available for the last few years). The cyclical component of GCI was used in our analysis of growth cycles in Poland over the last 25 years. It is also used as reference cycle in our work on composite leading indicators. Autoregressive 12-month projections of the indicator enabled us to make very exact forecasts of GDP growth rates for the last few years, much more precise than any other sources did.

The second one is the composite leading indicator (CLI) based on quantitative and qualitative data, compiled according to the OECD methodological standards. As a matter of fact, we have developed and tested many variants of CLI designed for monitoring and forecasting purposes, with different component variables and different leads. Composite leading indicators combines the selected set of economic variables that precede cyclical changes in aggregate economic activity.

The third one is the economic sentiment indicator (ESI) which reflects the opinion of economic agents (entrepreneurs and consumers) on the tendency of business and current economic conditions. Likewise, we are testing several variants of ESI, with different formulas and different sources of survey data. The indicator is based exclusively (or almost exclusively) on survey data. It is compiled as a weighted average of appraisals obtained from individual sectors of the economy.

The main results of our research have been summarised in papers prepared for the successive CIRET conferences. The paper presented at the 23rd CIRET Conference in Helsinki (Z. Matkowski, 1997) brought the analysis of growth cycles seen in the development of Poland's economy, as reflected by our reference index of aggregate economic activity. The paper prepared for the 24th CIRET Conference in Wellington (Z. Matkowski, 1999) was mainly devoted to the analysis of our composite leading indicators, and it discussed the use of survey data in their construction. This paper focuses on the application of survey data in economic sentiment indicators. The results of our research on the composite indicators of business activity for Poland have also been reported in the paper presented at the First International Meeting on Economic Cycles in Ourense, in June 2000 (Z. Matkowski, 2000).

Apart from section 1 – the introduction, this paper includes the following. Section 2 elucidates the concept of economic sentiment indicator and specifies alternative formulas of the indicator. Section 3 presents empirical results. Section 4 brings some conclusions.

2 Alternative ESI Concepts

The usefulness of survey data in monitoring systems designed to detect changes in economic activity has been evidenced by many theoretical and empirical studies. It has been proved that certain qualitative indicators, reflecting the opinion of economic agents (entrepreneurs, consumers, etc.) on their own situation and on the tendency of business in general, may be a valuable source of information on the actual course of business. On a theoretical plane, the concept of rational expectations explains the link between real economic developments and their reflection in people's attitudes and judgements. On an empirical plane, the analysis of leads and lags between micro- and macroeconomic variables provides strong proof in favour of monitoring and forecasting properties of such qualitative indicators as investment plans, order-books and capacity utilization rates of enterprises, or major purchase intentions of the consumers. Therefore, survey data are widely used in macroeconomic assessments and forecasts.

In our search for macroeconomic barometers that could be used to estimate current condition and prospects of Poland's economy we have developed and tested several variants of the economic sentiment indicator (ESI) based on qualitative data from business and consumer surveys. The basic idea behind this attempt is to combine composite indicators of business activity in individual sectors of economy, compiled from survey data, into a single indicator reflecting the general condition of national economy. This kind of macroeconomic index resembles the EU concept of economic sentiment indicator (ESI), though some of the tested formulas have different coverage and weights.

Several institutions in Poland are continuously testing public opinion about the course of business in individual sectors of the economy. The most important sources include the surveys made by the Research Institute of Economic Development at the Warsaw School of Economics (RIED) and by the Central Statistical Office (CSO). The RIED surveys now cover five sectors: industry, agriculture, construction, trade (retail & wholesale), and households. The CSO surveys cover three sectors: manufacturing industry, construction, and retail trade. These surveys are supplemented by the consumer survey made by Demoskop, a private polling company (recently, CSO has started its own consumer survey).

The first attempts to construct a general indicator of business activity for Poland based on survey data were made in 1993 independently by two researchers using the RIED survey data (K. Stanek, 1993; Z. Matkowski, 1993). Both of them calculated a general indicator inspired by the EU ESI concept, but different in coverage. Since 1994, K. Stanek continues to compile two versions of his synthetic indicator covering industry, agriculture, households, construction, and trade (since 1997, also transport), using his own system of constant weights and two different concepts of sectoral composite indicators. The first is called "RIED formula", and the second one is denoted "EU formula", though the latter significantly differs from the harmonized ESI concept both in coverage and weights.

In 1998, this author developed several alternative formulas of a general qualitative indicator of economic activity for Poland and filled them with RIED and CSO survey data, supplemented by consumer survey data from Demoskop. A concise analysis, covering the period 1994-1997, was included in the paper presented at the 24th CIRET Conference in Wellington (Z. Matkowski, 1999). Here we wish to present an updated and deepened analysis, covering the period from 1994 till 1999.

Tables 1 and 2 show the list of component indicators entering the composite index based on RIED and CSO survey data. Most of them are composed of 2-5 single time series representing business tendency in the given sector. Component variables have been taken from the source or calculated by the author according to his own (ZM) or harmonized (EU) formulas. Supplementary series describing consumer attitudes and Warsaw Stock Exchange Index were also included in some variants.

Table 1. Component variables entering economic sentiment indicators based on the RIED survey data

Code	Indicator / Formula	Compiled as:
ZH01	Business Indicator Industry (RIED)	Avg. of production appraisal & forecast
ZH03A	Industrial Activity (ZM)	Avg. of production appraisal & forecast
ZH05A	Industry Confidence (EU)	Future production, change in stocks, order book
ZH11A	Construction Activity (ZM)	Avg. of production appraisal & forecast
ZH09	Construction Confidence (EU)	Order book & expected change in employment
ZH17	Business Indicator Trade (RIED)	Future sales, anticipated change in supplies, commodity stocks
ZH20A	Trade Sales (ZM)	Avg. of sale appraisal & forecast
ZH25	Business Indicator Agriculture (RIED)	Current & future revenue, adjusted for anticipated change in economic condition
ZH28	Consumer Confidence (RIED)	Change in income and savings, modified by anticipation factor; since 1995 - EU formula
ZW01	Warsaw Stock Exchange Index	Undetrended
ZWR01	Warsaw Stock Exchange Index R/T	Detrended

Table 2. Component variables entering economic sentiment indicators based on the CSO survey data

Code	Indicator / Formula	Compiled as:
ZG01	Industrial Climate (CSO)	Avg. of enterprise situation & prospect
ZG03A	Industrial Activity (ZM)	Avg. of production appraisal & forecast
ZG05A	Industry Confidence (EU)	Future production, change in stocks, order book
ZG08	Construction Climate (CSO)	Avg. of production appraisal & forecast
ZG08A	Construction Activity (ZM)	Avg. of production appraisal & forecast
ZG10A	Construction Confidence (EU)	Order book & expected change in employment
ZG13	Retail Trade Climate (CSO)	Avg. of enterprise situation & prospect
ZG15A	Retail Sales (ZM)	Avg. of sale appraisal & forecast
ZD05	Consumer Confidence (DEMO)	EU formula
ZW01	Warsaw Stock Exchange Index	Undetrended
ZWR01	Warsaw Stock Exchange Index R/T	Detrended

Alternative formulas of the composite indicator based on RIED data have been denoted ZHG1 to ZHG5 while similar formulas filled with CSO data have been recorded as ZGG1 to ZGG5.

Variant 1 is the closest implementation of the harmonized EU ESI concept. Both with the RIED and CSO data, it was calculated according to the formula:

$$\{ZHG1 \equiv ZGG1\} = {}^{1}/_{3}\ \textit{Industry} + {}^{1}/_{3}\ \textit{Households} + {}^{1}/_{6}\ \textit{Construction} + {}^{1}/_{6}\ \textit{Share Prices}.$$

It includes the following components: industrial confidence indicator, consumer sentiment indicator, construction confidence indicator (all compiled according to EU standards), and the detrended share price index of Warsaw Stock Exchange. The latter is supposed to reflect confidence and expectations in the capital market.

Variant 2 does not significantly differ from variant 1 except that share price index was included in undetrended form, which facilitates the calculation. The formula is the same as in variant 1.

Variant 3 excludes the share price index that, in case of Warsaw Stock Exchange, does not yet correctly reflect real business developments. At the same time, it includes confidence indicator for trade and, in case of RIED data, also agriculture. Industry and households are given weights twice as high as each of the remaining sectors. The resulting formulas are as follows:

ZHG3 = $^{2}/_{7}$ Industry + $^{2}/_{7}$ Households + $^{1}/_{7}$ Construction + $^{1}/_{7}$ Trade + $^{1}/_{7}$ Agriculture,

ZGG3 = $^{1}/_{3}$ Industry + $^{1}/_{3}$ Households + $^{1}/_{6}$ Construction + $^{1}/_{6}$ Trade.

This indicator retains the logic of the original EU ESI concept and applies similar system of constant weights, but it differs in coverage. Therefore, it will be also referred to as "modified EU formula".

Variants 4 and 5 differ significantly from the EU ESI concept as to the coverage and weights. They cover productive sectors: industry, construction, and trade (in case of RIED, also agriculture), which directly contribute to the creation of GDP. Households and stock exchange are not included. Business tendency in each sector is determined according to EU concepts except for agriculture where it is expressed according to RIED’s own formula. Unlike in the preceding variants, component indicators reflecting the situation in individual sectors are combined using no arbitrary constant weights, but weights that represent their actual share in GDP (more precisely, in gross value added). The weights are changed each year, according to the changing structure of the economy. For the period covered by this analysis, the average shares were: industry - 30.8%, construction - 7.6%, agriculture - 6.2%, trade - 19.6%. Since the above sectors do not cover the whole economy (amounting to slightly less than 2/3 of GDP), sum total of weighted indices must be divided by the sum of weights. The resulting algorithms are:

$$\text{ZHG4} = \frac{a_1 \times Industry + a_2 \times Construction + a_3 \times Trade + a_4 \times Agriculture}{a_1 + a_2 + a_3 + a_4},$$

$$\text{ZGG4} = \frac{a_1 \times Industry + a_2 \times Construction + a_3 \times Trade}{a_1 + a_2 + a_3},$$

where weights a_1, a_2, a_3, a_4 reflect the contribution of individual sectors to GDP in the given year.

Variant 5 has the same coverage and the same weights as variant 4, but it is filled with homogeneous component indicators reflecting business tendency in individual sectors. Instead of different concepts of business tendency for individual sectors (a practice adopted by EU harmonized standards), this variant employs a uniform concept of business tendency for each sector, namely the average of output (sales) appraisal and forecast (except of agriculture where the available survey data do not contain the relevant information). The algorithms here used are the same as in variant 4.

3 Empirical Results

Economic sentiment indicators calculated according to the above formulas have been filled with survey data of the period from Nov. 1993 till Jan. 1999, available at the time of analysis. Quarterly data were transformed into monthly series by interpolation. All component series entering the combined indicator have been smoothed with 3-month moving average and standardised according to the following formula (OECD, 1987):

$$\left[(x-\bar{x}):\frac{\sum\left|x-\bar{x}\right|}{n}\right]+100 ,$$

where x denotes numerical values of the given variable, $\bar{x}$ - arithmetic mean of the time series values, n - number of observations (months). Standardised component series and composite indicators derived from them usually take values ranging between 95 and 105.

It should be noted that the resulting indicators express relative deviations of the values observed from their long run average (= 100), so they should not be mistaken for simple dynamic indices. Anyway, numerical values of indicator higher than 100 mean that current business activity is assessed positively (above average) while values lower than 100 indicate that current condition of the economy is perceived as rather poor (below the long-run trend). The increase of the indicator suggests a rising tendency of economic activity whereas its decrease is tantamount to a slack.

Alternative variants of economic sentiment indicators calculated from RIED survey data are shown in Figure 1, and similar variants based on CSO survey data (supplemented by Demoskop) are presented in Figure 2. For clarity, variant 2 (almost identical with variant 1) has been omitted in this presentation.

All the indicators display quite a regular seasonal pattern. In order to trace the tendency, they must be deseasonalized. This has been done using X11-ARIMA procedure. Seasonally adjusted and smoothed time series of ESI are shown in Figures 3 and 4. The charts now show the tendencies (trend + cycle) of indicators in the form of Henderson curve (9- or 13-term moving average).

Figure 1. Economic sentiment indicators based on the RIED survey data

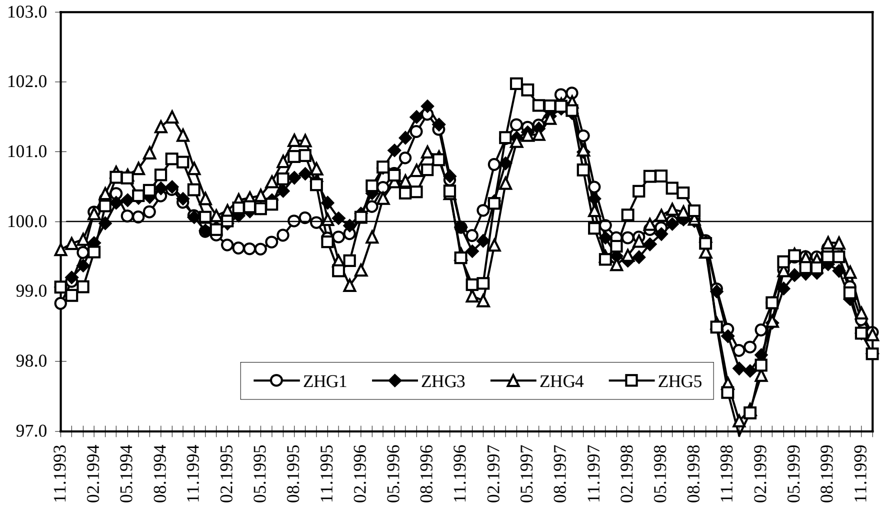


Figure 2. Economic sentiment indicators based on the CSO survey data

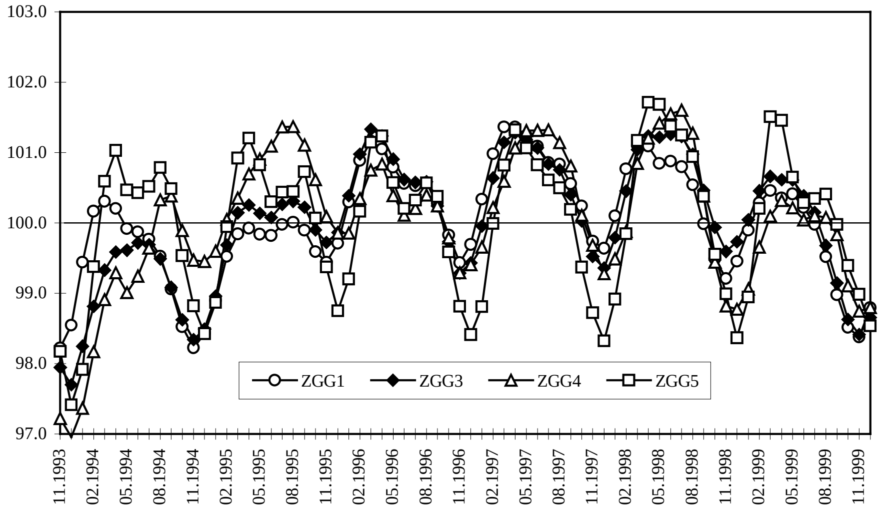


Figure 3. General business tendency according to RIED-based ESI

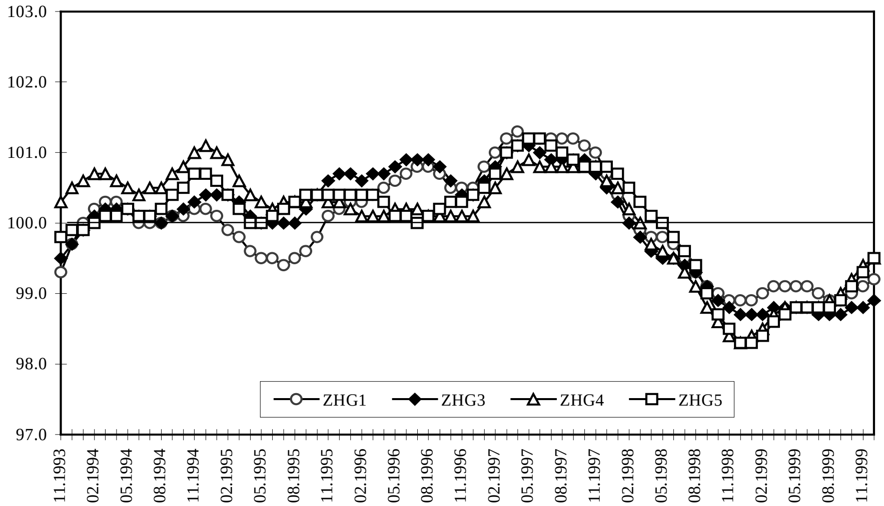
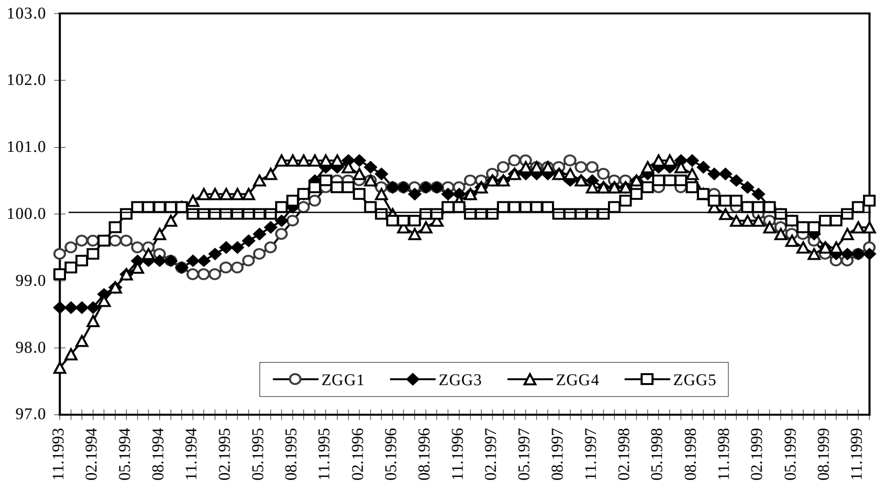


Figure 4. General business tendency according to CSO-based ESI



Looking at the figures, we can make two important observations. First, strikingly enough, it turns out that different ESI formulas give quite similar tendency of general business if they are filled with survey data from the same source. The alternative variants of the indicator give different empirical values, but after intensive smoothing applied to remove seasonal and irregular changes, we get similar tendencies reflecting trend and cycle, irrespective of the particular formula used. Secondly, the indicators compiled from RIED data depict more pessimistic tendencies of general business as compared to those obtained from CSO data. The RIED-based indicators tend to rise less sharply when economy is booming (as in he first half of the analysed period), and they tend to fall more deeply when economic conditions deteriorate even slightly (as in the second half of the period). We can say that the RIED-based ESIs are less sensitive to favourable trends and more sensitive to unfavourable trends in the economy.

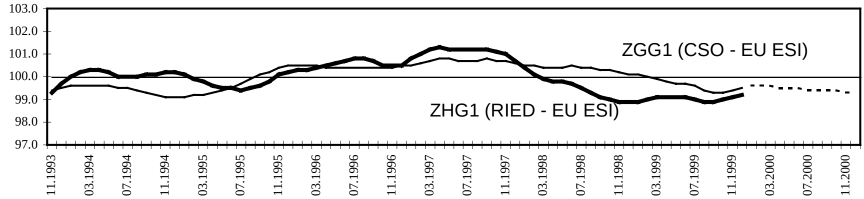
The difference between the RIED- and CSO-founded indicators can be explained partly by the fact that the first include agriculture, the sector most hit by economic reforms and foreign competition. But the difference also exists in variant 1, which does not include that sector. If the major aim of ESI would be to signal the worsening of economic climate (rather than its improvement), then RIED-based ESIs could have an advantage over the other source. On the other hand, CSO-based indicators seem to be more consistent with the actual development of the economy, marked by the acceleration of economic growth in the first half of the period and its deceleration during the second half (but no recession).

In order to check how different source of data (and slightly different coverage) may affect indicators calculated according to alternative formulas, we should look at Figures 5. Each chart shows two versions of ESI belonging to the same variant but filled with survey data from the two sources: RIED (thick line) and CSO (thin line). The solid lines reflect business tendency (trend + cycle). The dotted line indicates the ARIMA projection of seasonally adjusted values for the next 12 months (whenever available). Figures (a) and (b) show the RIED-based and CSO-based indicators compiled according to the original or modified EU formula (variants 1 and 3) while Figures (c) and (d) show the indicators developed according to author's own concepts (variants 4 and 5). The RIED-based indicators seem to display, at least in the second half of the period, some leads in comparison with the respective CSO-founded indicators.

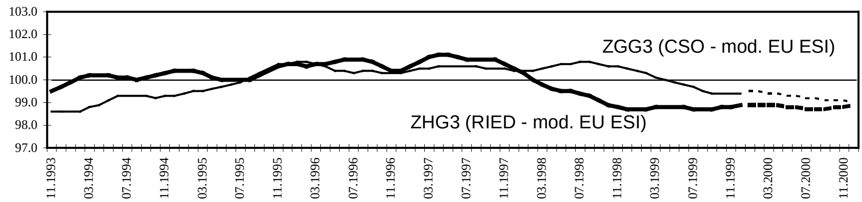
Table 3 shows important statistical properties of the ESI time series. All the calculated ESI variants have very good performance characteristics (QCS < 0.6, MCD • 3) that makes them suitable for monitoring purposes. The amount of irregular movements, after preliminary smoothing of component indicators by 3-months moving average, has been brought to minimum. The amount of seasonal changes is quite high (in some variants extremely high), but relatively stable seasonal patterns enable us to remove seasonality from the time series without any problems. With two exceptions, all the variants allow us to generate 12-month ARIMA forecasts. Correlation with the reference indicator GCI is rather low, but it heavily depends on the detrending method used to isolate cyclical components. Correlation between the time series detrended by linear regression is rather insignificant. However, with the reference indicator detrended by the OECD PAT procedure, we get a closer correlation and longer leads, as shown in the table. Most ESI indicators then reveal quite long leads against the reference cycle (between 3/4 and 1 1/2 year) which would make them useful in macroeconomic forecasts.

Figure 5. General business tendency according to the alternative ESI variants

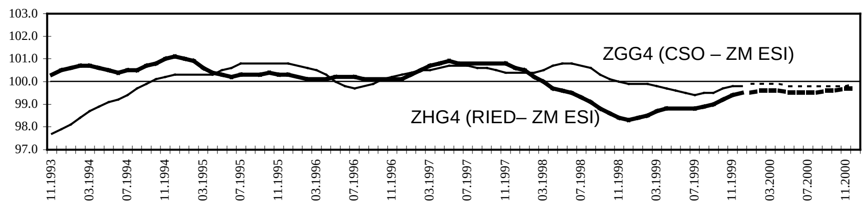
(a) EU ESI (ZHG1&ZGG1)



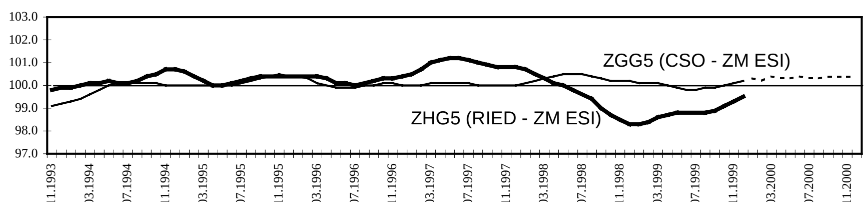
(b) Modified EU ESI (ZHG3&ZGG3)



(c) ZM ESI (ZHG4&ZGG4)



(d) ZM ESI (ZHG5&ZGG5)



Detrended and smoothed time series of the alternative ESI variants have also been correlated against each other in order to check how closely they move together and whether they display any significant leads or lags which might be essential in choosing the most useful formula. For the lack of space, we shall not present here the correlation matrix, but the results of cross-correlation may be summarised as follows. As it could be expected, formulas of similar coverage and weights, such as 1 and 2, or 4 and 5, give similar empirical patterns, which are closely correlated. Formulas 4 and 5 give different distribution of empirical values as compared with formulas 1 and 2. The source of data affects the distribution of empirical values of the indicator, but it does not significantly alter the total variance as long as its formula and coverage remain similar or the same. Over the whole period, all the indicators that are mutually significantly correlated do not reveal distinct leads or lags. However, RIED-based indicators, as already noted, seem to display some lead over the CSO-based equivalents in the second half of the period.

Table 3. Statistical properties of ESI variants

Code	Coverage	QCS	MCD	Relative contribution to stationary variance			ARIMA forecast	Cross-correlation against GCI	
				I	S	TC		lead (–)	R
ZHG1	I + C + H + S	0.59	1	1.3	28.2	69.3	no	–12	0.690
ZHG2	I + C + H + S	0.55	1	1.2	28.0	69.2	yes	–11	0.721
ZHG3	I + C + T + H	0.41	1	0.6	35.6	60.3	yes	–12	0.740
ZHG4	I + C + T + A	0.46	1	1.4	61.0	36.2	yes	–9	0.609
ZHG5	I + C + T + A	0.50	1	1.7	49.9	44.8	no	–15	0.644
ZGG1	I + C + H + S	0.27	1	1.0	48.6	41.8	yes	–13	0.614
ZGG2	I + C + H + S	0.25	1	0.8	49.6	42.4	yes	–14	0.631
ZGG3	I + C + T + H	0.20	1	0.7	48.0	44.9	yes	–19	0.587
ZGG4	I + C + T	0.42	1	1.4	38.2	48.7	yes	.	ns
ZGG5	I + C + T	0.30	3	1.7	84.9	5.90	yes	.	ns

I - industry, C - construction, H - households, S - stock price index, T - trade, A - agriculture.

In the light of the calculated ESI indicators, the general tendency of Poland's economy during the last few years is somewhat uncertain. All indicators reveal a clear-cut downward tendency since 1997 or 1998, much more pronounced in the indicators based on RIED data. The indicators compiled from CSO survey data give a more optimistic picture. Nevertheless, they also show the worsening of economic climate towards the end of period. The early indications of a slack have been ultimately supported by statistical data that showed a marked slowdown in economic growth in two indicated years. Our composite indicators based on survey data signalled the recent slowdown about 1.5 year in advance. On the other hand, most indicators announce some improvement of economic climate in the second half of 1999, and their projections for 2000 suggest something like stabilization.

The examination of economic merits and statistical properties of various ESI variants leads us to conclude that variants 1 (or 3) and 4 seem to be most promising and deserving further testing. Variant 1 is fully comparable with the harmonized EU ESI formula whereas the modified variant 3 seems to be more suitable for the transition economy (with relatively less developed capital market and relatively well developed trade). There is no significant difference between variants 1 and 2 if the share-price index does not show a distinct trend, but in the long run detrending of the index may be necessary to make it stationary. Formulas 4 or 5 are a real alternative to the EU ESI concept; both represent a consistent ESI concept coming close to the real category of aggregate economic activity and more comparable with real GDP data. Out of the two, variant 4 has slightly better statistical properties.

The most adequate formula of the general indicator of business activity derived from survey data will be chosen by comparing the evolution of alternative indicators over a longer period with the actual development of the economy, as reflected by GDP growth rates and other objective statistical measures, including our monthly reference index.

4 Conclusion

1. For analytical and monitoring purposes, the author has developed the following composite indicators of general economic activity for the Polish economy: (a) synthetic index of aggregate economic activity based on monthly output data from five major sectors of the economy (reference cycle); (b) economic barometers (composite leading indicators) developed according to OECD standards; (c) composite indicators of business activity based on survey data (economic sentiment indicators), compiled according to EU standards and supplemented by author's own concepts.
2. Several alternative formulas of a general indicator of economic activity, filled with survey data from different sources, have been developed and tested in order to find the most adequate variant.
3. Our composite indicators of business activity for Poland require further testing, but some of them can already be used for operational purpose, i.e. to evaluate current economic situation and to make macroeconomic forecasts.
4. Autoregressive projections of our reference indicator GCI enabled us to make very exact forecasts of GDP growth rates for 1998 and 1999. Current slowdown of economic growth was signalled by our CLIs one year in advance, and some of ESI variants announced it even earlier. The reference indicator has revealed a peak in 1998, indicating a slowdown in economic growth.
5. The author believes that the experience gained in our work on economic barometers for Poland may also be of interest for the analysts engaged in developing similar monitoring systems for other economies in transition.

Bibliography

Matkowski Z. (1993), In Search for an Economic Sentiment Indicator for Poland: The RIED Approach and Preliminary Results, Workshop on Business Surveys in Transition Countries, Poznan 1993.

Matkowski Z. (1997), Business Cycles in Poland, 23rd CIRET Conference, Helsinki 1997, in: Social and Structural Change - Consequences for Business Cycle Surveys (Oppenlaender K.H., Poser G., eds.), Ashgate, Aldershot - Brookfield - Singapore - Sydney 1998.

Matkowski Z. (ed.) (1997), Z prac nad syntetycznymi wskaznikami koniunktury dla gospodarki polskiej (Synthetic Indicators of Business Activity for the Polish Economy), „Prace i Materialy IRG", vol. 51, SGH, Warszawa 1997.

Matkowski Z. (ed.) (1998), Zlozone wskazniki koniunktury dla gospodarki polskiej oparte na standardach UE i OECD (Composite Indicators of Business Activity for Poland based on EU and OECD Standards), „Prace i Materialy IRG", vol. 61, SGH, Warszawa 1998.

Matkowski Z. (ed.) (1999), Barometry koniunktury dla gospodarki polskiej (Composite Indicators of Business Activity for Poland), „Prace i Materialy IRG", vol. 64, SGH, Warszawa 1999.

Matkowski Z. (1999), The Use of Survey Data in Economic Barometers for Poland, paper presented at the 24th CIRET Conference, Wellington 1999.

Matkowski Z. (2000), Composite Indicators of Business Activity for Poland, in: First International Meeting on Economic Cycles, Ourense, 21-23rd June 2000, UNED, Madrid 2000.

OECD Leading Indicators and Business Cycles in Member Countries 1960-1985, „Main Economic Indicators. Sources and Methods", no. 39, OECD, January 1987.

Stanek K. (1993), Koncepcja barometru gospodarczego koniunktury (The Concept of an Economic Barometer), in: Badania koniunktury gospodarki Polski, SGH, Warszawa 1993.

Confidence Indicators and Composite Indicators

Ronny Nilsson

1 Introduction

Most institutes conducting business tendency surveys select a set of survey series and combine them into a single cyclical composite or confidence indicator. This is done in order to reduce the risk of false signals, and to provide a cyclical indicator with better forecasting and tracking qualities than any of its individual components.

The reason why a group of indicators combined into a composite indicator should be more reliable over a period of time than any of its individual components is related to the nature and causes of business cycles. Each cycle has its unique characteristics as well as features in common with other cycles. But no single cause explains the cyclical fluctuation over a period of time in overall activity. The performance of individual indicators will then depend on the causes behind a specific cycle. Some indicators will perform better in one cycle and others in a different cycle. It is therefore necessary to have signals for the many possible causes of cyclical changes, and to use all potential indicators as a group.

This paper looks into the construction of the economic sentiment indicators calculated by the European Commission (EC) and the composite leading indicators calculated by the OECD. The aim of the paper is to evaluate the current rather coincident performance of the EC Economic Sentiment Indicators for a number of countries and to investigate possibilities of constructing sentiment indicators which would forecast cycles in economic activity with longer lead times. In particular, the following issues are discussed:

- *Selection of component series:* standard set of components for individual countries or aggregates or best performing series for individual countries;
- *Weighting of component series:* different or equal weights for individual components or aggregates;
- *Normalisation of cyclical amplitudes* and *smoothing* of component series;
- *Prompt availability* of component series and *revisions* to components and composite indicators.

The study is restricted to the EC Economic Sentiment Indicators (ESI) and OECD Composite Leading Indicators (CLI) calculated for France, Germany, Italy, the United Kingdom and the European Union as a whole. In Section 2, the performances of the EC ESIs and component series are evaluated over the period 1970-1999. The performances are evaluated both at turning points and over the whole cycle (cross-correlation) against total industrial production as a proxy for economic activity. These results are compared with the performance of the OECD CLIs for the same countries.

A standard set of components is used in the EC system. This set is mainly based on qualitative data from business or consumer tendency surveys. The EC ESI combines the following component series:

- industrial confidence indicator;
- construction confidence indicator;
- consumer confidence indicator;
- share price index.

Section 3 looks into the cyclical performance of the individual component series

in order to identify the components with the best leading characteristics. The results of this evaluation are then used as a basis for selection of country specific component series and the construction of alternative ESIs. A set of alternative ESIs for each country is constructed and evaluated in Section 4.

The effect of weighting of component series is also studied in Section 4. Different weights are used in the current EC system. The performances of the current EC ESIs are evaluated against the performance of the same ESIs but with equal weighting for individual component series.

A standard set of components across countries is used in the EC system while the OECD system includes series with best leading performance at turning points for individual countries. In Section 5, the component series used for the construction of the OECD CLIs are investigated and compared with the components used in the EC system.

The basic steps for the calculation of a composite index or confidence index are outlined in Section 6 and differences in methods applied by the OECD and the European Commission are discussed. In particular, the method of normalisation used in the current EC system is discussed. This method standardise each component series so that their average month-to-month changes are equal i.e. by dividing the month-to-month changes with the average month-to-month change. This method however gives little weight to the more irregular series in the cyclical movements of the composite index, unless some prior ad-hoc smoothing is performed. In the current EC system no smoothing is applied to component series and the effect of smoothing of components with a moving average equal to the month for cyclical dominance (MCD) measure, which is used in the OECD system, is studied.

Timeliness and absence of excessive revisions are obvious requirements of good cyclical indicators and these characteristics are also discussed Section 6 in connection with the aggregation of component series for the construction of composite indicators.

Finally, a summary of the main results from each section is presented and some general conclusions about the possibilities of constructing alternative ESIs with better leading performance are discussed.

2 Performance of the EC Economic Sentiment Indicators

The historical cyclical performance of the EC ESIs for France, Italy, the United Kingdom and the European Union as a whole are evaluated over the period 1970-1999. The performances are evaluated both at turning points and over the whole cycle (cross-correlation) against total industrial production as a proxy for economic activity and reference series for dating of cyclical turning points. The performances of the EC ESIs are compared with the performance of the OECD CLIs for the same countries and the EU as a whole. The evaluation results are presented in Table 1 and the cyclical developments of the EC ESIs and the OECD CLIs against the reference series are illustrated in Charts 1-10.

The results of the evaluations of the ESIs show rather coincident performance for all investigated countries and the EU as a whole according to the turning point analysis. The median lead of the ESIs at all turning points against the reference series is in the range of 1 to 3 months for all countries and the EU zone. However, the ESIs for Germany and the EU zone show longer leads at peaks, while the ESI for the United Kingdoms shows a better performance at troughs. The correlation results indicate better performances for Germany, the United Kingdom and the EU zone with leads in the range of 6 to 8 months. The cyclical profiles between the ESIs and the reference series are, however, not very close for any of the countries or the EU zone. In particular, the ESIs for Italy and France show very low correlation coefficients of 0.38 and 0.52 respectively.

In contrast to above results, the OECD CLIs for the same countries and the EU as a whole show in general much better results with longer leads at turning points and a closer correspondence with the reference series over the whole cycle. The median lead of the OECD CLIs at all turning points against the reference series is between 5-10 months for the EU zone and all countries except Germany. The lead is, in addition, rather good at both peaks and troughs for the EU zone and all countries except Germany which shows shorter leads at troughs and the United Kingdom, which shows shorter, leads at peaks. The cyclical profiles between the OECD CLIs and the reference series are very good with correlation coefficients in the range of 0.67 to 0.83 for all countries and the EU zone.

The construction of composite cyclical or leading indicators such as the EC ESI and the OECD CLI is the main objective of a cyclical indicator system, but the type of component series used in international indicator systems may be quite different. A standard set of indicators across countries may be used or an individual set of indicators per country may be used. An indicator system may be almost totally dependent on qualitative business and consumer survey series such as the EC system or based solely on quantitative statistical series which is the rule in the United States. The OECD system makes use of a blend of both qualitative business survey series and quantitative statistical series.

The use of a standard set of indicators across countries is a good approach for obtaining international comparability. However, cyclical indicators, which perform well in one country may not work well in another because of differences in economic structure and statistical system.

The EC ESI is calculated on a standard set of indicators while the OECD CLI is based on individually selected leading indicators for each country. The above results point in favour of the OECD CLI in comparison to the EC ESI and in the following section we will look into alternative combinations of component series for the construction of the EC ESI in order to improve the forecasting capacity for individual countries.

Table 1. Historical Performance of EC Economic Sentiment Indicators and OECD Composite Leading Indicators over the Period 1970-1999

Performance 1970-1999
Against ratio to trend of industrial production

Zone/Country/ Composite Indicator	Extra (x) or missing (m) cycles	Turning point analysis: Median lag (+) in months at Peak	Trough	All turning points	Mean absolute deviation around median	Cross-correlation: Months Lag (+)	Peak value
France							
OECD Composite Leading Indicator		-8	-6	-6	6.1	-8	0.78
EC Economic Sentiment Indicator		-3	0	-1	12.0	-3	0.52
Germany							
OECD Composite Leading Indicator	1x	-6	-1.5	-2	4.6	-4	0.72
EC Economic Sentiment Indicator	1x	-7	-0.5	-2	5.8	-6	0.62
Italy							
OECD Composite Leading Indicator		-10	-9	-9.5	4.4	-9	0.79
EC Economic Sentiment Indicator		-2	0	-1	7.1	-3	0.38
United Kingdom							
OECD Composite Leading Indicator	2m	-2	-7	-5	6.8	-11	0.67
EC Economic Sentiment Indicator	2m	-1	-5	-3	4.9	-8	0.68
European Union							
OECD Composite Leading Indicator		-7	-6	-7	4.1	-8	0.83
EC Economic Sentiment Indicator		-5	-2	-2	4.8	-6	0.61

Chart 1. France: EC ESI, OECD CLI and industrial production Ratio to trend

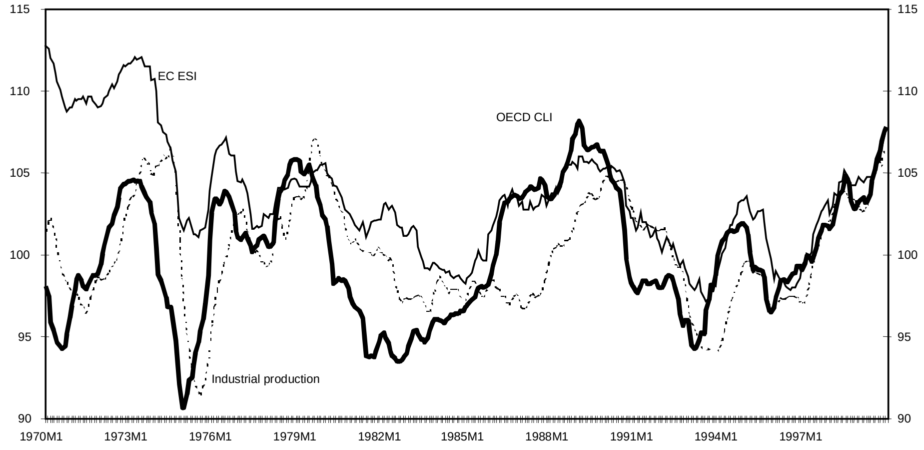


Chart 2. Germany: EC ESI, OECD CLI and industrial production Ratio to trend

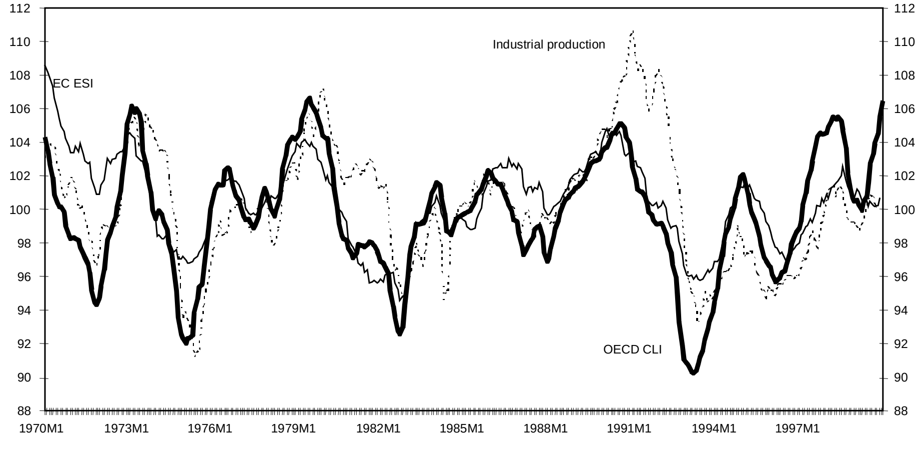


Chart 3. Italy: EC ESI, OECD CLI and industrial production Ratio to trend

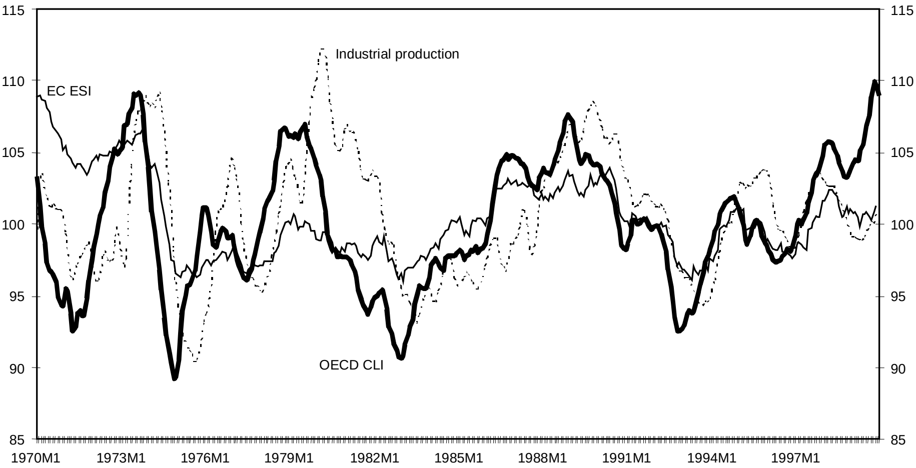


Chart 4. United Kingdom: EC ESI, OECD CLI and industrial production Ratio to trend

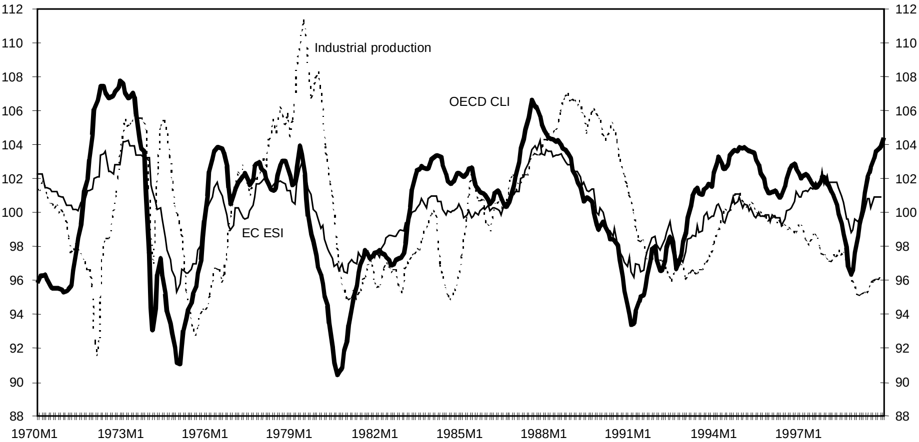
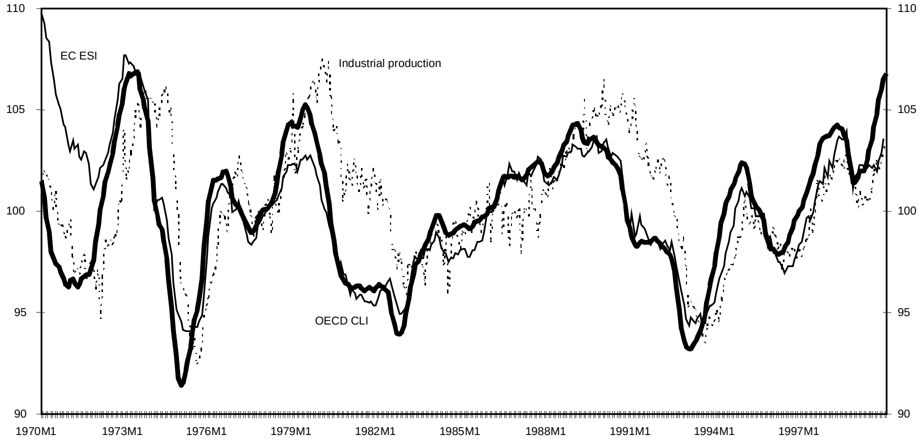


Chart 5. European Union: EC ESI, OECD CLI and industrial production Ratio to trend



3 Performance of Component Series Included in the EC Economic Sentiment Indicators

A standard set of components is used in the EC system. This set is mainly based on qualitative data from business or consumer tendency surveys. The EC ESI combines the following component series:

- industrial confidence indicator (ICI);
- construction confidence indicator (CCI);
- consumer confidence indicator (CSCI);
- share price index (SPI).

The industrial confidence indicator is the arithmetic average of the answers (balances) to the question on production expectations, order books and stocks of finished goods (inverted).

The construction confidence indicator is the arithmetic average of the answers (balances) to the questions on order books and employment expectations. The consumer confidence indicator is the arithmetic average of the answers (balances) to the four questions: on the financial situation of households; on the past and future general economic situation; and on the advisability of making major purchases (of consumer durable).

Different weights are used in the current EC system. The components are divided into two groups with equal weights to components in each group. The first group contains the industrial confidence indicator and the consumer confidence indicator, and the second group includes the construction confidence indicator and the share price index. The components in the second group are given half the weight of the components in the first group.

The results of the evaluation of the ESIs in section two showed rather coincident performances for all investigated countries. In this section, the cyclical performances of the individual component series are investigated in order to identify the components in each country with the best leading characteristics. The results of this evaluation are then used as a basis for selection of country specific component series and the construction of alternative economic sentiment indicators.

The historical performances of the ESIs and its component over the period 1970-1999 for France, Germany, Italy and the United Kingdom are set out in Table 2 and the cyclical profiles of the confidence indicator components are presented in Charts 6 - 9. The cyclical performance of the ESIs and its components are evaluated against industrial production as reference series.

The results for *France* show a rather coincident relationship between the reference series and two of the components, namely the industrial confidence indicator and the construction confidence indicator. Both series show a lead of only 1-2 months but a good correspondence with the reference series with correlation coefficients around 0.70. The consumer confidence indicator and the share price index, however, show good leads of 6-8 months according to the median lag, but show rather weak correspondence with the reference series with correlation coefficients of about 0.50.

The evaluation measures for *Germany* indicate different performances for two of the components. According to the median lag, the construction confidence indicator shows a long lead of 10 months against the reference series but only a 3 months lead as measured by the correlation lag. On the other hand, the consumer confidence indicator shows a zero lag according to the median but a long lead of 16 months as measured by the correlation lag. The industrial confidence indicator shows a coincident relationship with the reference series on both measures while the share price index shows a lead of 6 and 11 months on the two measures but the correlation is rather weak with a correlation coefficient of only 0.35.

In the case of *Italy*, the share price index shows no correspondence at turning points with the reference series and the correlation results are not significant. The construction confidence indicator shows coincident behaviour with a short lead of 2 months according to the median but a lag of 1 month as measured by the correlation lag. The industrial confidence indicator shows a lead of 5 months according to median and a good correspondence with the reference series with a correlation coefficient of 0.70. A lead of 10 months is registered for the consumer confidence indicator but the relationship with the reference series is rather weak with a correlation coefficient of 0.41.

Three of the four component series for the *United Kingdom* show rather long leads in the range of 6-l2 months as measured with both the median lag and the peak-correlation lag. Only the construction confidence indicator shows a shorter lead of 1 month as measured by the median but register a 6 months lead according to the peak-correlation lag. All components, with exception of the consumer confidence indicator, show a rather good correspondence with the reference series with correlation coefficients in the range of 0.60 - 0.77.

Table 2. Historical Performance of EC Economic Sentiment Indicator and Components

Performance 1970-1999 Against ratio to trend of industrial production							
		Turning point analysis				Cross-correlation	
	Compo-nent	Median lag (+) in months at			Mean absolute deviation around median	Months Lag (+)	Peak value
Zone/Country/ Composite Indicator		Peak	Trough	All turning points			
France							
EC Economic Sentiment Indicator		-3	0	-1	12.0	-3	0.52
	ICI	-3	1.5	0	7.4	-2	0.67
	CCI	-1	3.5	-1	13.9	-1	0.71
	CSCI	-8.5	-6	-6	10.9	-3	0.48
	SPI	-7	-8	-8	5.9	-7	0.51
Germany							
EC Economic Sentiment Indicator		-7	-0.5	-2	5.8	-6	0.62
	ICI	-1	-1	-1	4.5	-2	0.67
	CCI	-8	-12.5	-10	13.1	-3	0.61
	CSCI	-7	3	-2	10.9	-16	0.58
	SPI	-9	-4.5	-6	6.1	-11	0.35
Italy							
EC Economic Sentiment Indicator		-2	0	-1	7.1	-3	0.38
	ICI	-7	-4.5	-5	3.9	-3	0.7
	CCI	-0.5	-5.5	-1.5	7.6	1	0.53
	CSCI	-14.5	-7	-7	17.1	-10	0.41
	SPI	nc	nc	nc	nc	nc	ns
United Kingdom							
EC Economic Sentiment Indicator		-1	-5	-3	4.9	-8	0.68
	ICI	-1	-5	-3	4.9	-8	0.68
	CCI	-4.5	-6	-6	7.8	-8	0.61
	CSCI	0	-2	-1	5.9	-6	0.77
	SPI	-6	-4.5	-6	8.8	-12	0.44

ICI = Industrial confidence indicator
CCI = Construction confidence indicator
nc = no correspondence
CSCI = Consumer confidence indicator
SPI = Share price index
ns = not significant

Chart 6. France: EC Confidence Indicators Balance

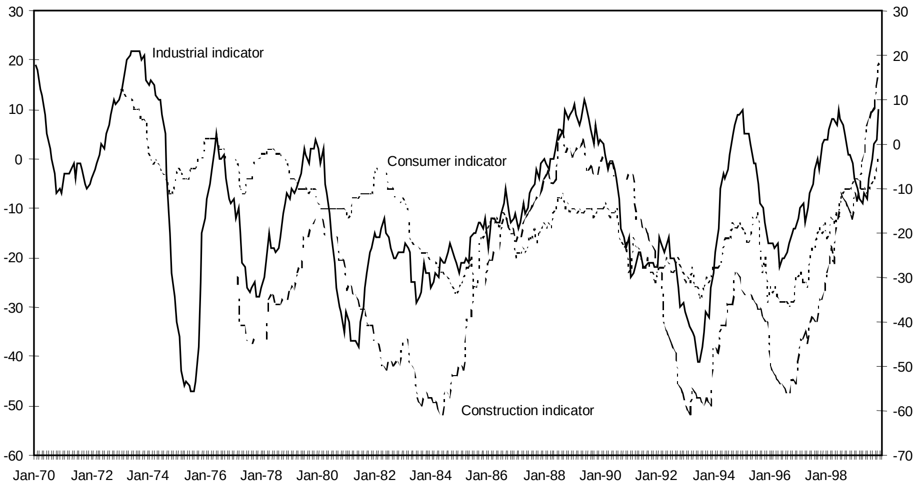


Chart 7. Germany: EC Confidence Indicators Balance

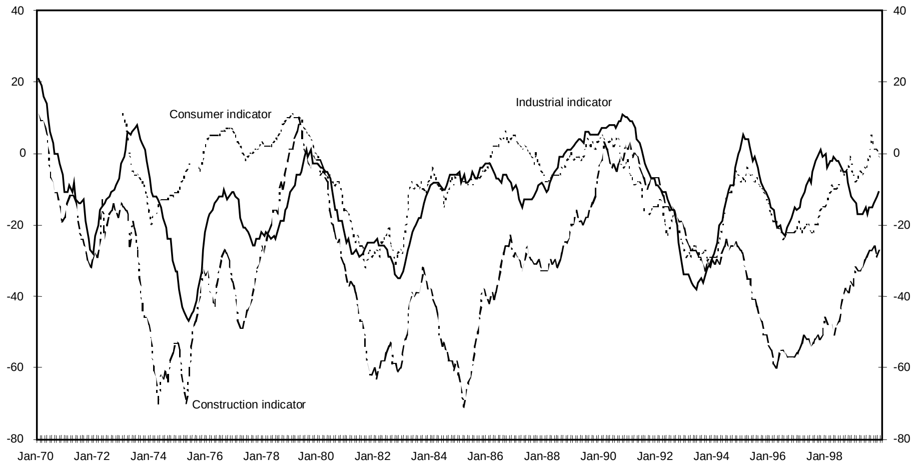


Chart 8. Italy: EC Confidence Indicators Balance

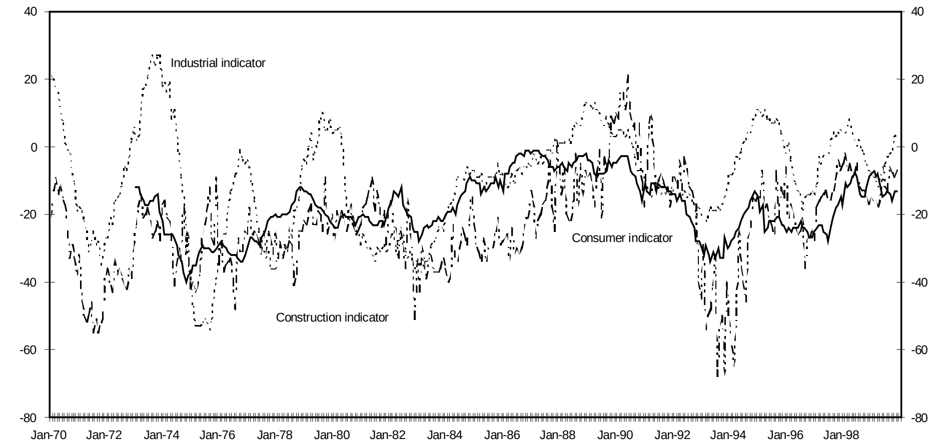
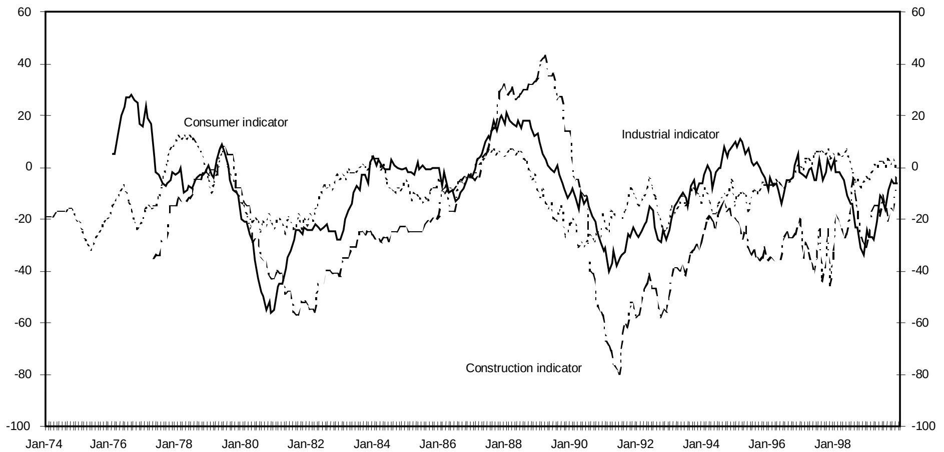


Chart 9. United Kingdom: EC Confidence Indicators Balance



4 Construction and Performance of Alternative Economic Sentiment Indicators

The results obtained in section three indicate that it may be possible to improve the cyclical performance by selecting and combining only the best components for each country. The historical performances of different sets of alternative ESIs are presented in Table 3.

The existing EC ESI use different weights for the aggregation of component series as noted above. However, if only the best components are selected and combined it would be reasonable to give equal weights to the components and the alternative ESIs evaluated are constructed with equal weights for the different components.

France: Alternative Economic Sentiment Indicators

The four alternative ESIs evaluated for France include the following components:

ESI1	Industrial confidence indicator (ICI) Construction confidence indicator (CCI) Consumer confidence indicator (CSCI) Share prise index (SPI)
ESI2	ICI, CSCI, SPI
ESI3	ICI, SPI
ESI4	CSCI, SPI

ESI1 includes the same four components as the existing EC ESI only the weighting system is different between the two indicators. The performance of the ESI1 over the period 1970-99 shows no major differences compared to the EC ESI. Both indicators show coincident relationships with the reference series at all turning points and a short lead of 3-4 months at peaks. However, the ESI1 shows a lag of 3 months at troughs while the EC ESI shows zero lag.

On the other hand, the ESI1 registers much higher correlation against the reference series with a correlation coefficient of 0.68 compared to 0.52 for the EC ESI.

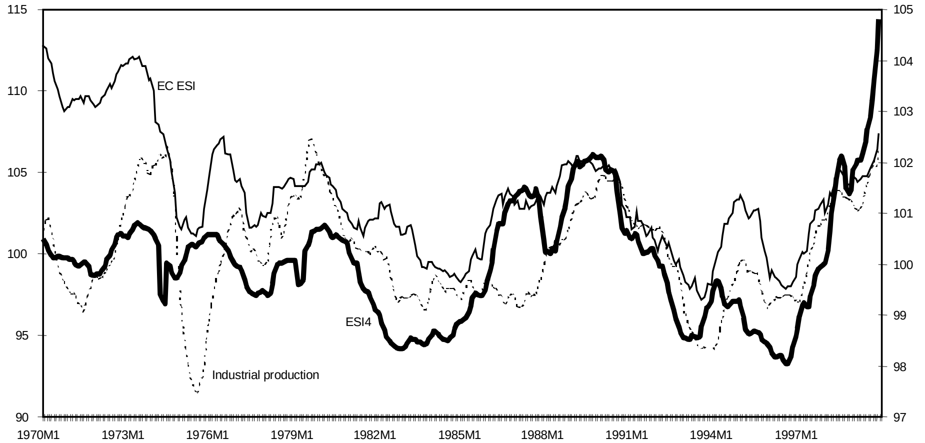
ESI2 includes three of the components in ESI1 with the longest leads against the reference series and excludes the shorter leading construction confidence indicator. However, this indicator shows no improvement in terms of lead times compared to the EC ESI but the correspondence with the reference series is clearly better with a correlation coefficient of 0.72.

Only two components are included in the ESI3, the industrial confidence indicator and the share prise index. The performance of this indicator shows more or less the same results as the ones registered for ESI2.

ESI4 includes the consumer confidence indicator and the share price index, the two components with the longest leads against the reference series. The consumer confidence indicator is only available from 1977 and the evaluation of the ESI4 is performed for the period 1977-99 for which both components are available and the period 1970-99. The results show a clear improvement in terms of lead times in comparison to the all other sentiment indicators. The lead at all turning points is 4 months and at the lead at peaks as long as 10 months. In comparison, the EC ESI for the period 1977-99 shows zero lag at both all turning points and at peaks and a lag of 3 months at troughs.

These results indicate, that the use of equal weights for components is not improving the performance of ESI1 compared the EC ESI, which use different weights for groups of components. However, an alternative sentiment indicator for France (ESI4) with better leading performance than the existing EC ESI could be constructed by selecting only the two component series with the longest lead included in this indicator (Chart 10).

Chart 10. France: EC ESI, ESI4 and industrial production
Ratio to trend



Germany: Alternative Economic Sentiment Indicators

The five alternative ESIs evaluated for Germany include the following components:

ESI1	Industrial confidence indicator (ICI) Construction confidence indicator (CCI) Consumer confidence indicator (CSCI) Share price index (SPI)
ESI2	ICI, CCI, CSCI
ESI3	ICI, CCI, SPI
ESI4	ICI, CCI
ESI5	CCI, SPI

ESI1 includes the same four components as the existing EC ESI only the weighting system is different between the two indicators. The performance of the ESI1 over the period 1970-99 shows no major differences compared to the EC ESI. Both indicators show coincident relationships with the reference series at all turning points and troughs, but a lead of 7 months at peaks.

All other alternative ESIs with different combinations of component series included in the EC ESI show about the same performance as the ESI1. Not even the ESI5, which only includes the two components with the longest leads against the reference series namely, the construction confidence indicator and the share price index, show an improved leading behaviour in comparison with any of the other ESIs.

These results indicate that it is difficult to construct an economic sentiment indicator for Germany with better performance than the existing EC ESI only by selecting and combining the component series included in this indicator.

Italy: Alternative Economic Sentiment Indicators

The two alternative ESIs evaluated for Italy include the following components:

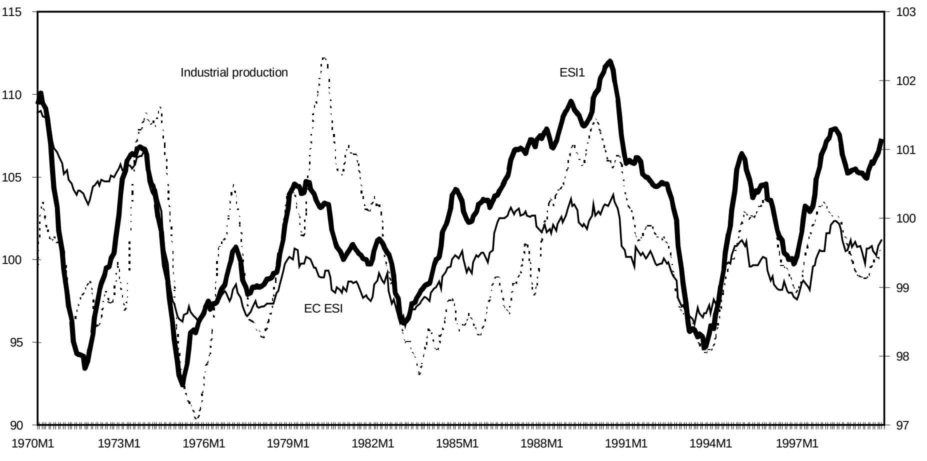
ESI1	Industrial confidence indicator (ICI)
	Construction confidence indicator (CCI)
	Consumer confidence indicator (CSCI)
ESI2	ICI, CSCI

The ESI1 includes three of the four components used in the EC ESI, it excludes the share price index, which shows no correspondence with the cyclical development of the reference series. This alternative sentiment indicator shows an average lead of 5 months at all turning points and peaks and 4 months at troughs over the period 1970-99. This performance is much better than the one registered for the EC ESI, which shows a coincident relationship with the reference series. In addition, the correspondence between the ESI1 and the reference series is rather good with a correlation coefficient of 0.65 compared to 0.38 for the EC ESI.

An even better cyclical performance is obtained if only the two components of the ESI1 with longest leads against the reference series are combined. These components are the industrial confidence indicator and the consumer confidence indicator which are included in the ESI2. This indicator shows an average lead of about 7 months at all turning points and at troughs and a lead of over 10 months at peaks. The correspondence with the reference series is also good with a correlation of 0.62.

These results show that alternative potential sentiment indicators for Italy with better leading performance than the existing EC ESI could be constructed. A first such indicator (ESI1) includes three of the four components used in EC ESI, but excludes the share price index, which shows no correspondence with the reference series. A second potential sentiment indicator includes only the two components with the longest leads against the reference series (Chart 11).

Chart 11. Italy: EC ESI, ESI1 and industrial production
Ratio to trend



United Kingdom: Alternative Economic Sentiment Indicators

The six alternative ESIs evaluated for the United Kingdom include the following components:

ESI1	Industrial confidence indicator (ICI)
	Construction confidence indicator (CCI)
	Consumer confidence indicator (CSCI)
	Share price index (SPI)
ESI2	ICI, CSCI, SPI
ESI3	ICI, CCI, SPI
ESI4	ICI, CCI
ESI5	ICI, SPI
ESI6	ICI, CSCI

The evaluations of the first four alternative sentiment indicators are performed both over the period 1970-99 and 1976-99. This is done because only data for the share price index is available back to 1970 while data for other components start between 1974 and 1977. The results for the period 1976-99 are more representative for the performance of the alternative sentiment indicators and are commented on in the following.

ESI1 includes the same four components as the existing EC ESI only the weighting system is different between the two indicators. The cyclical performance of this potential sentiment indicator over the period 1977-99 shows an average lead of 3 months at all turning points, an average lead of 5 months at troughs and 1 month at peaks. This performance is slightly better than the one registered for the EC ESI, which shows a more coincident behaviour with an average lead of 1 months at all turning points (Chart 12).

The two alternative sentiment indicators (ESI2 and ESI3) with different combinations of three of the components included in the EC ESI show about the same performance as ESI1.

An even better performance is obtained if only two of the best components with the longest leads against the reference series are combined. Different combinations of these components are included in sentiment indicators ESI4, ESI5 and ESI6. The best performance is registered for ESI5, which includes the industrial confidence indicator and the share price index. This indicator shows an average lead at all turning points of close to 10 months, an average lead of 11 months at troughs and over 7 months at peaks.

These results show that alternative ESIs for the United Kingdom with better leading performance than the existing EC ESI could be constructed. Only by introducing an equal weighting system of the components included in EC ESI gave an indicator (ESI1) with better leading performance. However, the best performance was obtained when combining only two of the three components with the longest leads against the reference series.

Chart 12. United Kingdom: EC ESI, ESI1 and industrial production Ratio to trend

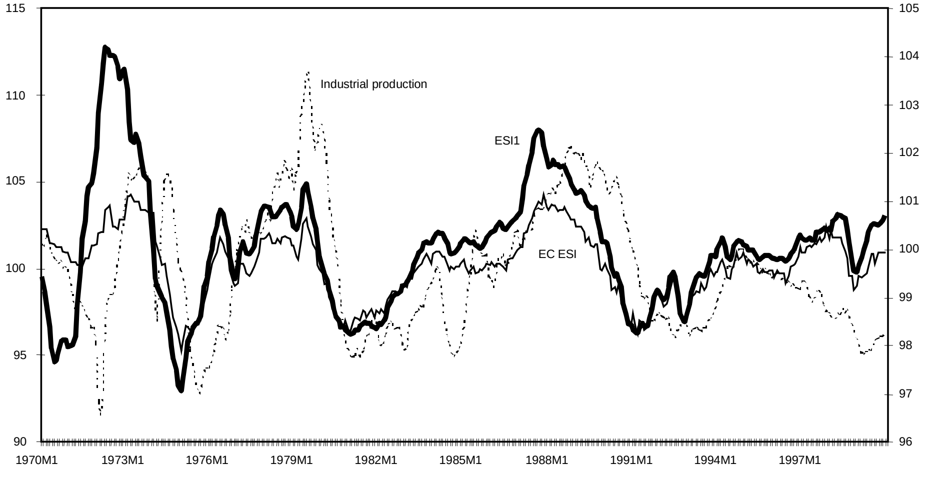


Table 3. Historical Performance of Different Sets of Sentiment Indicators

Country/ Sentiment Indicator	No. of series	Components	Weighting system	Performance 1970-1999: Turning point analysis: Median lag (+) in months at: Peak	Trough	All turning points	Mean absolute deviation around median	Cross-correlation: Test period	Months Lag (+)	Peak value
France										
EC Economic Sentiment Indicator	4	ICI, CCI, CSCI, SPI	Different weights	-3 0	0 3.5	-1 0	12.0 13.9	1970-99 1977-99	-3 -1	0.52 0.79
ESI1	4	ICI, CCI, CSCI, SPI	Equal weights	-4	3	-1	7.9	1970-99	-3	0.68
ESI2	3	ICI, CSCI, SPI	Equal weights	-3	4	-1	7.9	1970-99	-2	0.72
ESI3	2	ICI, SPI	Equal weights	-3	4	-1	7.7	1970-99	-3	0.70
ESI4	2	CSCI, SPI	Equal weights	-10 -10	1.5 -0.5	-4 -4	7.9 7.8	1970-99 1977-99	-6 -3	0.58 0.69
Germany										
EC Economic Sentiment Indicator	4	ICI, CCI, CSCI, SPI	Different weights	-7	-0.5	-2	5.8	1970-99	-6	0.62
ESI1	4	ICI, CCI, CSCI, SPI	Equal weights	-7	-1	-1	7.6	1970-99	-7	0.66
ESI2	3	ICI, CCI, CSCI	Equal weights	-7	0.5	0	7.9	1970-99	-4	0.67
ESI3	3	ICI, CCI, SPI	Equal weights	-6	-1	-1	7.0	1970-99	-5	0.70
ESI4	2	ICI,CCI	Equal weights	-5	-0.5	-1	5.6	1970-99	-2	0.75
ESI5	2	CCI, SPI	Equal weights	-6	-1.5	-2	7.5	1970-99	-8	0.61
Italy										
EC Economic Sentiment Indicator	4	ICI, CCI, CSCI, SPI	Different weights	-2	0	-1	7.1	1970-99	-3	0.38
ESI1	3	ICI, CCI, CSCI	Equal weights	-4	-5	-5	5.4	1970-99	-3	0.65
ESI2	2	ICI, CSCI	Equal weights	-10.5	-6.5	-7.5	5.7	1970-99	-4	0.62

Table 3. continued

Country/ Sentiment Indicator	No. of series	Components	Weighting system	Turning point analysis: Median lag (+) in months at Peak	Trough	All turning points	Mean absolute deviation around median	Cross-correlation: Test period	Months Lag (+)	Peak value
United Kingdom										
EC Economic Sentiment Indicator	4	ICI, CCI, CSCI, SPI	Different weights	-1 -0.5	-5 -3	-3 -1	4.9 5.3	1970-99 1976-99	-8 -8	0.68 0.68
ESI1	4	ICI, CCI, CSCI, SPI	Equal weights	-7 -1	-8 -5	-8 -3	7.1 6.4	1970-99 1976-99	-11 -9	0.71 0.73
ESI2	3	ICI, CSCI, SPI	Equal weights	-7.5 -1	-8 -5	-8 -3	7.2 7.1	1970-99 1976-99	-11 -10	0.66 0.65
ESI3	3	ICI, CCI, SPI	Equal weights	-7 -1	-8 -5	-8 -3	7.8 7.2	1970-99 1976-99	-11 -8	0.72 0.76
ESI4	2	ICI, CCI	Equal weights	-5	-9	-7.5	4.7	1976-99	-7	0.76
ESI5	2	ICI, SPI	Equal weights	-7.5	-11	-9.5	6.4	1976-99	-8	0.69
ESI6	2	ICI, CSCI	Equal weights	-5	-4	-5	7.8	1976-99	-10	0.59

Performance 1970-1999

ICI = Industrial confidence indicator
CSCI = Consumer confidence indicator
CCI = Construction confidence indicator
SPI = Share price index

5 Standard Set or Specific Components for Individual Countries

Two different strategies could be used for the selection of component series to be included in a composite indicator. A standard set of indicators across countries may be used or an individual set of indicators per country may be used. The use of a standard set of indicators across countries is a good approach for obtaining international comparability. However, cyclical indicators, which perform well in one country may not work well in another because of differences in economic structure and statistical system.

The EC ESI is calculated on a standard set of indicators. However, as shown in Section 3, some of the standard indicators used in the EC system did not work equally well in all countries and alternative country specific ESIs could be constructed with better leading performance.

The OECD CLI is based on individually selected leading indicators for each country. The performance of this indicator was compared with the performance of the EC ESI in Section 2 and the results pointed in favour of the OECD CLI. In this section, we will look into the component series used in individual countries for the construction of the OECD CLI and their cyclical characteristics in order to explain the results obtained in Section 2. The historical performance of OECD CLI and components for France, Germany, Italy and the United Kingdom for the period 1960-1996 are set out in Table 4.

France

The OECD CLI for France includes 11 component series: 4 business and consumer survey series, 4 financial series, one export related series (terms of trade), one series related activity in the United States (USA CLI) and a series related to consumption of durable goods (passenger cars registered).

Two series are identical to two of the components used in the EC ESI: the share price index and the consumer sentiment indicator. These two indicators are the components with the longest leads in the EC ESI. Two of the three business survey series are also included in the EC ESI, but as components in the industrial confidence indicator (production future tendency and finished goods stocks).

The performance of most components and in particular all financial components included in the OECD CLI show much longer leads in comparison to the industrial and construction confidence indicators included in the EC ESI which show coincident behaviour. All financial indicators show average leads at all turning points of over 10 months measured by the median lag, while the USA CLI shows a lead of 7 months. However, terms of trade and passenger cars show coincident behaviour according to the median lag.

The inclusion of financial indicators in the OECD CLI for France is the major factor behind the longer lead obtained for this indicator in comparison with the EC ESI. However, as shown in Section 3, the performance of the EC ESI can be improved, if only the two components with the longest leads in this indicator are combined. Such an indicator is on the other hand narrowly based which may reduce its reliability in a specific cycle.

Germany

The OECD CLI for Germany is very much restricted to components related to activity in the industrial sector of the economy. Out of the 6 components included, 4 refer to industrial activity as measured by the business survey in industry. The remaining two components refer to a volume series on new orders and the share price index.

Only the share price index is a component series that is used in common with the EC ESI. Two of the 4 business survey series are also included in the EC ESI, but as components in the industrial confidence indicator (order books and finished goods stocks). The other two survey series are the business climate indicator and order inflow/demand tendency where in particular the later indicator shows the best leading performance at turning points of all survey series.

The performances at cyclical turning points of all components show an average lead in the range 2-7 months measured by the median lag. This performance is not better than that registered for the components included in the EC ESI. In addition, the EC ESI is more broad based in that it includes construction and consumer confidence indicators.

Overall, the OECD CLI and the EC ESI show about the same coincident performance at cyclical turning points. The results in Section 3 also show that it is difficult to improve this performance by combining only the two components included in the EC ESI with the longest leads, that is the construction confidence indicator and the share price index.

Italy

The OECD CLI for Italy includes 6 components: 3 business and consumer survey series, one financial series, one export related series (terms of trade), and a volume series on new orders.

The consumer confidence indicator is the only component in common with the EC ESI. However, the two business survey series are also included in the EC ESI, but as components in the industrial confidence indicator. All survey series show good leading performance at cyclical turning points with an average lead in the range of 7-9 months measured by the median lag. About the same performance is registered for the industrial and consumer confidence indicators included in the EC ESI. However, the construction confidence indicator included in the EC ESI shows coincident behaviour at cyclical turning points.

The performances of the financial series (yield of long term government bonds) and the terms of trade series register the longest leads at cyclical turning points with an average lead of 12 and 14 months respectively. The share price index is not included in the OECD CLI because it shows no cyclical relationship with the reference series. This indicator is, however, included in the EC ESI and this shows the danger of using a standard set of components across countries.

The OECD CLI shows a much longer lead at cyclical turning points in comparison to the EC ESI and the inclusion of the financial series and the terms of trade series mainly explain this. However, as shown in Section 3, the performance of the EC ESI can be improved if the share price index is excluded and can be improved even more if also the construction confidence indicator is removed.

United Kingdom

The OECD CLI for the United Kingdom includes 9 components: 6 business survey series, 2 financial series and a series related to consumption of durable goods (passenger cars registered).

Only the share price index is a component series that is used in common with the EC ESI. Three of the six business survey series are also included in the EC ESI, but as components in the industrial confidence indicator (production expectations, order books and finished goods stocks). The other three survey series are the business climate, prospects for exports and raw material stocks expectations. All these other survey indicators show longer leads at turning points than the survey series included in the industrial confidence indicator.

The performance of the financial series (prime bank bills) register the longest lead at cyclical turning points with an average lead of 17 months, followed by the survey series on export prospects with an average lead of 9 months. However, two of the survey series, order books and finished goods stocks show coincident behaviour at turning points. These two series are also components in the industrial confidence indicator included the EC ESI

The OECD CLI shows a slightly better performance at cyclical turning points in comparison with the EC ESI. The inclusion of the financial series and the longer leading business survey series are the main factors behind this. However, as shown in Section 3, the performance of the EC ESI can be improved, by introducing an equal weighting system of components and can be improved even more if only two of the tree components with longest leads against the reference series are combined.

Table 4. Historical Performance of OECD Composite Leading Indicator and Components

	Performance 1960-1996							
			Turning point analysis				Cross-correlation	
Zone/Country/ Composite Indicator	Components	Extra (x) or missing (m) cycles	Median lag (+) in months at: Peak	Trough	All turning points	Mean absolute deviation around median[1)]	Months Lag (+)	Peak value
France								
OECD Composite Leading Indicator			-5.5	-5	-5	6.0	-8	0.75
	1. Bond yield granted by government	2x	-10	-10	-10	9.2	-14	-0.50
	2. Share price index	3x	-10.5	-6.5	-7	8.2	-7	0.34
	3. Inter-bank loans (3 months)	3x	-11	-10	-10.5	6.6	-14	-0.69
	4. Call money rate	2x	-15	-10.5	-11	11.8	-14	-0.50
	5. Terms of trade	3x, 1m	-1	1	0	12.6	-13	0.56
	6. Passenger cars registered		-3	0	-2	11.2	-1	0.44
	7. USA Composite leading indicator		-6.5	-7	-7	9.9	-8	0.44
	8. Prospects for industrial sector (BS)	3x	-5	-1	-3	7.1	-7	0.54
	9. Production future tendency (BS)	2x	-6	-2	-3.5	7.0	-6	0.70
	10. Finished goods stocks (BS)	1x	0	-1	-1	10.1	-3	-0.61
	11. Consumer sentiment indicator (BS)	1x	-6	-3	-4.5	2.1	-3	0.60

1) Standard deviation

Table 4. continued

Zone/Country/ Composite Indicator		Components	Extra (x) or missing (m) cycles	Performance 1960-1996: Turning point analysis: Median lag (+) in months at: Peak	Trough	All turning points	Mean absolute deviation around median[1)]	Cross-correlation: Months Lag (+)	Peak value
Germany									
OECD Composite Leading Indicator				-8	-4	-6	6.0	-8	0.80
	1.	Share price index, industrials		-6	-4	-5	6.0	-9	0.50
	2.	New orders, total volume		-9	-4	-6	4.0	-3	0.90
	3.	Order inflow/demand tendency (BS)	1x	-8	-4	-7	6.0	-9	0.50
	4.	Finished goods stocks (BS)		-5	-2	-3	3.0	-5	-0.70
	5.	Order books (BS)		-6	-2	-3	3.0	-3	0.80
	6.	Business climate (BS)		-4	-2	-2	3.0	-7	0.60
Italy									
OECD Composite Leading Indicator				-9	-8.5	-9	4.9	-7	0.76
	1.	Yield of long term government bonds	1x	-12	-15	-12.5	9.3	-12	-0.53
	2.	Terms of trade	1x, 1m	-11	-16	-14.5	12.5	-14	0.58
	3.	New orders, total volume	1x	-6	-7.5	-7	5.5	-3	0.43
	4.	Production future tendency (BS)	1x	-7	-8	-7.5	6.3	-3	0.61
	5.	Order books/demand tendency (BS)	1x	-11	-7.5	-9	5.9	-6	0.62
	6.	Consumer confidence indicator (BS)	2x	-12	-8	-8.5	6.8	-8	0.64
United Kingdom									
OECD Composite Leading Indicator				-10	-9	-9	7.2	-12	0.71
	1.	Prime bank bills (3 months)	2x	-13	-17	-17	8.4	-18	-0.55
	2.	Share price index	5x	-8.5	-7.5	-7.5	7.7	-10	0.57
	3.	New cars registered	1x	-4	-9	-6	10.3	-7	0.52
	4.	Production future tendency (BS)	2x	-2	-8	-4	6.4	-15	0.63
	5.	Order books/demand tendency (BS)		0	-5	0	6.3	8	0.64
	6.	Raw material stocks future tend. (BS)	1x	-2	-7	-5.5	6.8	-9	0.56
	7.	Finished goods stocks (BS)	3x	-2	-1.5	-2	13.2	-7	-0.72
	8.	Prospects for exports (BS)	5x	-8	-13.5	-9	6.7	-5	0.29
	9.	Business climate (BS)	1x	-6	-7.5	-6.5	6.6	-5	0.53

1) Standard deviation

6 Confidence and Composite Indicators: Methodological Issues

In the following paragraphs some of the basic steps for the calculation of a composite index or a confidence index are outlined and differences in methods applied by the OECD and the European Commission are discussed.

The basic principle to form a composite indicator consists of summing the individual indicator series included in the basket of component series while accounting for the component series relative importance and cyclical amplitude. The basic formula can be written as:

$$CI = \sum w_i \; s_i \; C_i$$

where:
CI = composite indicator
C = component series
i = is the number of component series
w = is the weight of the component series
s = is the standardisation factor of the component series

Trend estimation

The first consideration in the construction of a composite cyclical indicator is that of amplitude-stationarity. The first objective is to ensure that each individual indicator series included in the composite indicator is stationary in some way. Both the OECD and the EC indicator systems use the "growth cycle" or "deviation-from-trend" approach. Trend estimation is thus a crucial step in detecting cyclical movements and identifying turning points.

Long-term trends in the OECD system are estimated using a modified version of the Phase Average Trend method (PAT) developed by the US National Bureau of Economic Research (NBER). The PAT method requires an initial list of turning points, which define the cyclical phases, in order to estimate a trend, which cuts through the phases. A first list of turning points is obtained from the preliminary peaks and troughs identified by calculating a first trend estimate based on a 75-month moving-average. A series of tests are then executed on the deviations from the trend in order to eliminate extreme values to obtain a better identification of the final turning points. For this purpose, the program specifies a minimum duration of each phase (5 months) and the minimum duration of each cycle (15 months). The final trend is then calculated with a validated list of turning points as input to the PAT program. The detrended results (deviations from long-term trend) are then used as input to the composite calculation.

Trend estimation in the EC system systems is not performed directly but implied by using month-to-month changes either in percentage form or differences (balances in the case of survey series) as input to the composite index calculation. This method is used in the EC system for the three confidence indicators while the original NBER method is used to de-trend the share price index.

Smoothing

It is necessary to ensure that all component series have equal "smoothness". This is to ensure that month-to-month changes in the composite indicator are not unduly influenced by irregular movements in any one indicator series. The OECD procedure is to use the "Months for Cyclical Dominance" (MCD) moving average. This procedure ensures approximately equal smoothness between series and also ensures that the month-to-month changes in each series are more likely to be due to cyclical than to irregular movements. The data lost at the end of the series due to the moving average are restored with an extrapolation by regression over the end of the series.

The MCD moving averages used to smooth the component series in the OECD system are set out in Table 5. The MCD values for most series across investigated countries are in the range 1-3. A MCD value of 1 means that no smoothing is needed and this concerns only a few component series: the USA Composite Leading Indicator (France), one business survey series on order books (Germany) and three business survey series on raw material stocks, prospects for exports and business climate (United Kingdom). Most financial and other business survey indicators have an MCD of 2 or 3 while most quantitative statistical indicators show MCD values of 4 or 5 such as terms of trade (France), cars registered (France and United kingdom), volume of new orders (Italy). The smoothing performed on the component series in the OECD system ensures that the composite leading indicators across all countries are smooth with a MCD value of 1, which means that no further smoothing is needed for an easy identification of a cyclical turning point.

On the other hand, no smoothing of component series is performed in the EC system, but, as can been seen in Table 5, almost all component series used in the EC system show MCD values in the range 2-4. This is reflected in the ESIs, which show MCD values of 2 for all investigated countries except Germany (MCD=1). This means that only in the case of Germany is the ESI smooth enough for easy identification of cyclical turning points.

Standardisation (normalisation)

Standardisation or normalisation of component series is necessary in order to minimise the influence of series with marked cyclical amplitude to dominate the composite indicator. The method used in the OECD system to calculate normalised indices is first to subtract the mean and then to divide by the mean of the absolute values of the difference from the mean. The normalised series are then converted into index form by adding 100.

The method of normalisation used in the EC systems is to reduce each component series so that their average month-to-month changes are equal, i.e. by dividing the month-to-month changes with the average month-to-month change. This method however gives little weight to the more irregular series in the cyclical movement of the composite index, unless some prior ad-hoc smoothing is performed. In contrast in the OECD system the amplitudes of the cyclical movements are normalised but the relative magnitude of the irregular movements are unchanged.

Lagging

Finally, it may sometimes be necessary to lead or lag particular indicators. In the OECD system this is done in only one case, where the indicators selected for a particular country fall into two distinct groups of "longer-leading" and "shorter-leading" indicators. Combining the two types of indicators gave unsatisfactory results because of the interference between the two cycles. The alignment was improved by lagging the longer-leading group of indicators.

Weighting

Different weights may be assigned to component series in order to reflect their economic significance (coverage and economic reason), statistical adequacy, cyclical conformity, speed of availability of data, etc. The purpose of weighting is to improve reliability by giving higher weight to components with good quality i.e. indicators which correlate highly with each other and the resultant composite indicator.

A statistical method such as principal component analysis could be used to choose optimal weights. However, such a method would minimise the contribution of indicators, which do not move with the other indicators. This may reduce the reliability of the composite indicator because some indicators perform better in one cycle and others in a different cycle. Therefore, most indicator systems in operation use an equal weighting system after standardisation, once the components have been selected.

On the other hand, a weighting or scoring system is a valuable tool in the selection of indicators to be included in the composite index.

In the OECD system, equal weights are normally used to obtain each country's composite indicator. This does not mean that there is no weighting in the OECD system, because equal weighting implies, by default, a judgement on appropriate weights, and the normalisation process is itself a weighting system in reverse. However, when the composite indicators for individual countries are combined into indicators for country groups, each composite indicator is assigned the weight used in calculating group totals for the industrial production index.

Different weights are however used in the EC system. The components are divided into two groups with equal weights to components in each group. The first group contains the industrial confidence indicator and the consumer confidence indicator, and the second group includes the construction confidence indicator and the share price index. The components in the second group are given half the weight of the components in the first group.

Aggregation

In the OECD system, the raw composite index is obtained by averaging the normalised indices of each component series. A composite index calculated on an incomplete set of data is linked to the body of the index by use of a linking factor which is equivalent to applying the growth-rate of the "incomplete" index to the last point at which a full index is available.

Timeliness and absence of excessive revisions are obvious requirements of good cyclical indicators. These two issues are discussed in the following.

The timeliness or availability of component series at the time of the compilation of the OECD composite leading indicators is set out in Table 5. Availability is here measured in terms of months where one indicates that data for a component series is available for the month for which the composite leading indicator (CLI) is calculated. In the OECD system only the CLI for the United Kingdom is calculated with all component series available every month. However, the percentage of component series available for the calculation of the CLIs in France and Germany is over 80 per cent, but only 67 per cent in Italy.

What is interesting to notice is that all series not available for the month for which the CLI is calculated concerns quantitative statistical series such as terms of trade, cars registered and volume of new orders. On the other hand, all financial and business survey component series are always available for the month for which the CLI is calculated.

The advantage with the EC system is that all components are always available for the calculation of the composite index across countries. This is explained by the fact that all component series refer to business survey and financial indicators.

It is very important that the series are not revised to a significant extent in later periods if they are to be used for analysing the present economic situation and for forecasting. Business survey series rarely are revised whilst in many countries preliminary data for conventional statistics are released very quickly but later revised up to three times. For a few indicators - in particular indices of production and new orders - about 30-40 per cent of the forecasting errors are due to revisions of the first published data in some countries.

7 Summary and Conclusions

The EC Economic Sentiment Indicator (ESI) is calculated on a standard set of indicators while the OECD Composite Leading Indicator (CLI) is based on individually selected leading indicators for each country. The results presented in section two points in favour of the OECD CLI in comparison to the EC ESI and alternative combinations of component series for the construction of the EC ESI in order to improve the forecasting capacity in individual countries were investigated in sections three and four.

The results obtained in section three indicate that it may be possible to improve the cyclical performance of the EC ESI by selecting and combining only the best components for each country. The EC ESI uses different weights for the aggregation of component series. However, if only the best components are selected and combined it would be reasonable to give equal weights to the components and the alternative ESIs evaluated were constructed with equal weights for the different components.

The results in section four shows that the forecasting performance of the EC ESI could be improved in all investigated countries except Germany, if only the components with the longest lead in this indicator were combined. In the case of Italy, a better alternative ESI could be constructed if the share price index was excluded. This indicator shows no cyclical relationship with the reference series and to include it shows the danger of using a standard set of components across countries. The results for the United Kingdom shows that only by introducing an equal weighting system of the components included in EC ESI gave an alternative ESI with better leading performance.

Two different strategies could be used for the selection of component series to be included in a composite indicator. A standard set of indicators across countries may be used or an individual set of indicators per country may be used. The use of a standard set of indicators across countries is a good approach for obtaining international comparability. However, cyclical indicators, which perform well in one country may not work well in another because of differences in economic structure and statistical system.

The OECD Composite Leading Indicator (OECD CLI) is based on individually selected leading indicators for each country. The performance of this indicator was compared with the performance of the EC ESI in section 2 and the results pointed in favour of the OECD CLI.

In section five, we looked into the component series used in individual countries for the construction of the OECD CLI and their cyclical characteristics in order to explain the results obtained in section 2. The results showed that the inclusion of financial indicators in the OECD CLI for France and Italy was the major factor behind the longer lead obtained for this indicator in comparison with the EC ESI. In the case of the United kingdom, the inclusion of financial series and alternative longer leading business survey series were the main factors behind the better performance of the OECD CLI in comparison to the EC ESI.

The basic steps for the calculation of a composite index or a confidence index were outlined in section six and differences in methods applied by the OECD and the Commission of the European Communities (EC) were discussed. In particular, issues related to smoothing of component series, timeliness of components and revisions to component series and composite indicators were investigated.

In the OECD system, component series are smoothed by the "Months for Cyclical Dominance" (MCD) moving average. This procedure ensures approximately equal smoothness between series and also ensures that the month-to-month changes in each series are more likely to be due to cyclical than to irregular movements. On the other hand, no smoothing of component series is performed in the EC system. The effect of this is reflected in the Economic Sentiment Indicators (ESI), which show MCD values of 2 for all investigated countries except Germany (MCD=1). This means that only in the case of Germany is the ESI smooth enough for easy identification of cyclical turning points.

In the OECD system only the CLI for the United Kingdom is calculated with all component series available every month. However, the percentage of component series available for the calculation of the CLIs in France and Germany is over 80 per cent, but only 67 per cent in Italy. What is interesting to notice is that all series not available for the month for which the CLI is calculated concerns quantitative statistical series such as terms of trade, cars registered and volume of new orders. On the other hand, all financial and business survey component series are always available for the month for which the CLI is calculated.

The advantage with the EC system is that all components are always available for the calculation of the composite index across countries. This is explained by the fact that all component series refer to business survey and financial indicators.

From above summary the following conclusions may be drawn on how to improve the forecasting potential of the EC Economic Sentiment Indicators in individual countries:

- Select only best performing component series in individual countries;
- Use an equal weighting system for aggregation of component series;
- Perform smoothing of component series;
- Introduce more financial components.

Table 5. Timeliness, Irregular Variation and Smoothing of Components in the OECD System of Composite Leading Indicators and in the EC System of Economic Sentiment Indicators

OECD Composite Leading Indicators 1960-1999	Timeliness	Irregular variation MCD	EC Economic Sentiment Indicators 1970-1999	Timeliness	Irregular variation MCD
France			France		
OECD Composite Leading Indicator	1	1	*EC Economic Sentiment Indicator*	1	2
Components			Components		
1. Bond yield granted by government	1	2	1. EC Industrial Confidence Indicator	1	2
2. Share price index	1	3	2. EC Construction Confidence Indicator	1	3
3. Inter-bank loans (3 months)	1	2	3. EC Consumer Confidence Indicator	1	2
4. Call money rate	1	2	4. Share price index	1	3
5. Terms of trade	3	4			
6. Passenger cars registered	2	5			
7. USA Composite leading indicator	1	1			
8. Prospects for industrial sector (BS)	1	2			
9. Production future tendency (BS)	1	2			
10. Finished goods stocks (BS)	1	2			
11. Consumer sentiment indicator (BS)	1	3			

Table 5. continued

OECD Composite Leading Indicators 1960-1999	Timeliness	Irregular variation MCD	EC Economic Sentiment Indicators 1970-1999	Timeliness	Irregular variation MCD
Germany			Germany		
OECD Composite Leading Indicator	1	1	*EC Economic Sentiment Indicator*	1	1
Components			Components		
1. Share price index, industrials	1	2	1. EC Industrial Confidence Indicator	1	1
2. New orders, total volume	3	3	2. EC Construction Confidence Indicator	1	2
3. Order inflow/demand tendency (BS)	1	5	3. EC Consumer Confidence Indicator	1	2
4. Finished goods stocks (BS)	1	2	4. Share price index	1	2
5. Order books (BS)	1	1			
6. Business climate (BS)	1	2			
Italy			Italy		
OECD Composite Leading Indicator	1	1	*EC Economic Sentiment Indicator*	1	2
Components			Components		
1. Yield of long term gov. bonds	1	2	1. EC Industrial Confidence Indicator	1	2
2. Terms of trade	5	3	2. EC Construction Confidence Indicator	1	4
3. New orders, total volume	3	5	3. EC Consumer Confidence Indicator	1	2
4. Production future tendency (BS)	1	3	4. Share price index	1	3
5. Order books/demand tendency (BS)	1	3			
6. Consumer confidence indicator (BS)	1	2			
United Kingdom			United Kingdom		
OECD Composite Leading Indicator	1	1	*EC Economic Sentiment Indicator*	1	2
Components			Components		
1. Prime bank bills (3 months)	1	2	1. EC Industrial Confidence Indicator	1	2
2. Share price index	1	2	2. EC Construction Confidence Indicator	1	3
3. New cars registered	1	5	3. EC Consumer Confidence Indicator	1	2
4. Production future tendency (BS)	1	3	4. Share price index	1	2
5. Order books/demand tendency (BS)	1	2			
6. Raw material stocks fut. tend. (BS)	1	1			
7. Finished goods stocks (BS)	1	3			
8. Prospects for exports (BS)	1	1			
9. Business climate (BS)	1	1			

Bibliography

OECD, "OECD Leading Indicators and Business Cycles in Member Countries, Sources and Methods 1960-1985", No.39 January 1987.

OECD, "OECD Leading Indicators", Ronny Nilsson, Economic Studies, No. 9, Autumn 1987.

OECD, "Leading Indicators for OECD, Central and Eastern European Countries", Ronny Nilsson, Article published in the book "Is the Economic Cycle Still Alive?", edited by Baldassarri/Annunziato, St Martins Press Inc., New York 1994/The Macmillan Press LTD, London 1994.

OECD, "Cyclical Indicators in Poland and Hungary", I. Kudrycka, Research Centre for Economic and Statistical Studies, Poland, J. Hoos, G. Muszely, Ministry of Finance, Hungary and R. Nilsson OECD, Paper presented at the OECD Meeting on Leading Indicators, Paris October 1996.

OECD, "Cyclical Indicators and Business Tendency Surveys", OCDE/GD(97)58, General Distribution; 1997.

OECD, "An update of OECD Leading Indicators", Gérald Petit, Gérard Salou, Pierre Beziz and Christophe Degain, Paper presented at the OECD Meeting on Leading Indicators, Paris October 1996.

OECD, "The 1994 Mexican Crises: Were Signals Inadequate?", Pierre Beziz and Gerald Petit, Paper published in the Cambridge Review of International Affairs (Summer/Fall 1997).

OECD, "Calculation of composite leading indicators: A Comparison of Two Different Methods", Olivier Brunet, Paper presented at the 25th CIRET Conference, Paris October 2000.

OECD, "The OECD System of Leading Indicators: Recent Efforts to Meet User’s Needs", Benoit Arnaud, Paper presented at the 25th CIRET Conference, Paris October 2000.

National Bureau of Economic Research (NBER), "Programmed turning Point Determination", Gerhard Bry and Charlotte Boschan, Paper published in Cyclical Analysis of Time Series: Selected Procedures and Computed Programmes, 1971.

Chapter II

Leading Indicators

The OECD System of Leading Indicators: Recent Efforts to Meet Users' Needs

Benoît Arnaud

1 Introduction

The OECD system of composite leading indicators (CLI) consists of leading indicators for 22 Member countries and 7 country groupings such as, major seven countries or Europe. The Statistics Directorate (STD) of the OECD has been publishing CLI since 1981 using a modified version of the method developed by the US National Bureau of Economic Research (NBER).

CLI data are published in the *Main Economic Indicators* (MEI) publication as well as in a monthly note. Historical data are available on Leading Indicators and Business Surveys (LIBS) diskettes. Since 1996, information about CLIs has been available on the OECD web site. This has broadened the audience of OECD CLIs who have continuously provided STD with various feedback. STD has attempted to reflect these views in the compilation and dissemination of CLIs. STD also aims to convince potential users of the usefulness of the OECD system.

This paper aims to provide information on the recent developments within STD regarding the expectations of users of the OECD system of CLIs. For the past two years, STD has received more than hundred questions, requests or suggestions from users, or potential users, of CLIs in a variety of institutions including national banks, national statistical offices, universities and economic services. It should be emphasised that this is not the result of a user survey but voluntary feedback from users. Even given this limitation, the feedback provided by CLI users still gives valuable insight on how CLIs are used, by whom, and where changes in the OECD methodology should be incorporated, in particular, those which will expand the use of the CLI by both OECD internal and external users.

This paper is organised as follows. A description of the OECD system of CLIs is given in section 2. Section 3 discusses the information available in various publications or the web site of OECD. In section 4, the various users' requests are examined and the main conclusions concerning their expectations are drawn. Finally the directions in which STD proposes to concentrate its efforts are summarised in section 5.

2 The OECD System of CLIs

This section gives a brief description of the methodology used in the compilation of the OECD CLIs. This explanation is mainly extracted from the publication *"OECD Leading Indicators and Business Cycles in Member countries, Sources and Methods 1960-1985", No.39, OECD (1987)* and from an article available on the OECD Internet site[1] *"OECD Composite Leading Indicators: a tool for short-term analysis", OECD (1999).* The programs that are used to compile CLIs have been written by the OECD statistics directorate using the FAME language.

Composite leading indicators for individual countries

The total industrial production index has been chosen as reference series for each country.[2] The Gross Domestic Product (GDP) may have been a more obvious choice; however, GDP has a major disadvantage of only being available on a quarterly basis, while the industrial production index is compiled every month. Production in industry includes almost all the sectors susceptible to medium-term variations which are the most appropriate for analysis by the cyclical indicators technique. In addition, many non-industrial sectors such as associated services display cycles connected to that of industry.

The "growth-cycle" or deviation-from-trend approach has been retained. The point of departure is that essentially the same medium-term variations exist in many different economic indicators and that the timing differences which exist between them may be exploited for the purposes of forecasting. However, the variations are obscured by different rates of growth of long-term trend and therefore measuring deviations from trend is appropriate.

Long-term trends in time series are estimated using a modified version of the Phase Average Trend (PAT) method developed by the US National Bureau of Economic Research (NBER). The PAT procedure requires an initial list of turning points which define the cyclical phases. Initial turning points estimates are calculated using the Bry-Boschan routine.[3] The trend is eliminated from the series by dividing the original series by the trend to give a ratio-to-trend since the underlying structure of most series used is multiplicative. However, some component series are considered to have an additive structure. They are detrended by substraction to give a difference-from-trend.

The component series used in the compilation of the CLIs were selected according to several criteria. These include consistency and length of lead of the detrended indicator over the detrended reference series at cyclical turning points, absence of extra or missing cycles, smoothness, frequency of publication, economic significance, breadth of coverage and theoretical justification of the leading relationship. An evaluation is also necessary on the basis of a number of practical criteria: freedom from excessive revisions, timeliness of

---

1 http://www.oecd.org/std/cli

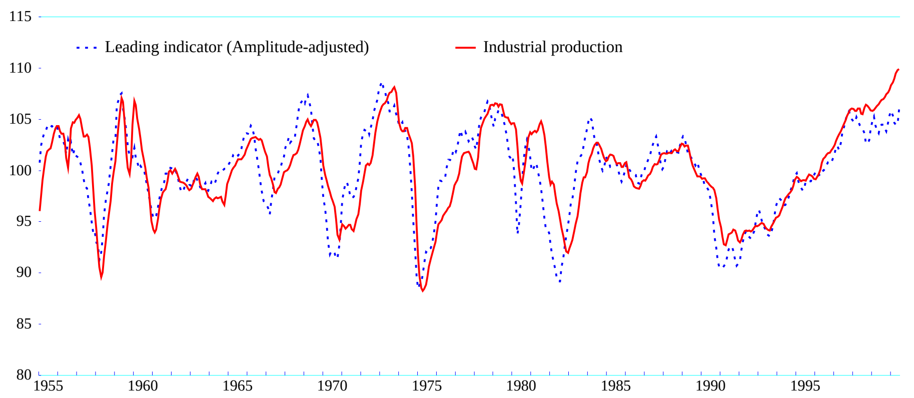
2 Exceptions are industrial production in manufacturing for the Netherlands and Norway and volume of manufacturing sales for Denmark.

3 See [1] and [2] for more details.

publication and availability of a long run of data of statistically satisfactory reliability. The main leading indicator series used in the OECD system are: business surveys results on production, stocks, order books and new orders; monetary and financial indicators such as monetary aggregates, interest rates and share prices; construction indicators; terms of trade; stocks.

Prior to aggregation into a single CLI, component series are smoothed according to their MCD[4], in order to reduce the irregularity of the final composite indicator. They are also normalised, to homogenise the cyclical amplitude.

Figure 1. Cyclical developments in the United States, Ratio to trend (smoothed)



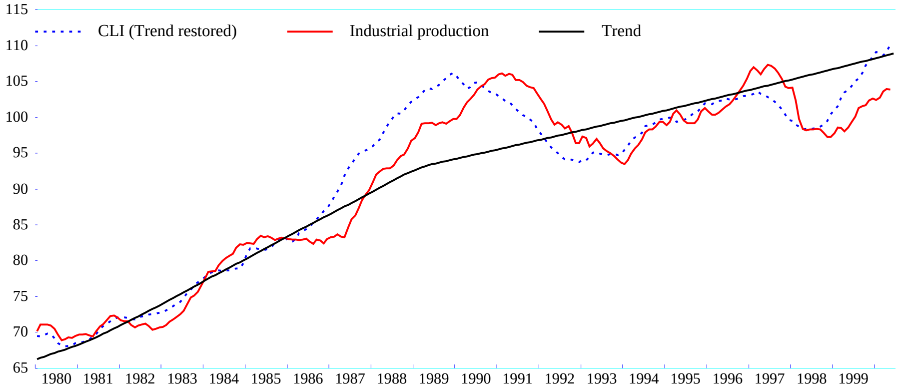
This form of the CLI is called the amplitude-adjusted and ratio-to-trend CLI. This can then be multiplied by the trend of the reference series to obtain the trend restored CLI.

The CLI is calculated by aggregating the detrended component series. For presentation, an adjustment of the cyclical amplitude to equal that of the reference series is carried out. In the following figure, we can easily compare the evolution of both series. In particular, this is a way of emphasising how well the CLI has predicted the turning points in the economy of a country (the United States in this example).

This trend-restoration enables direct comparison with the reference series. It gives an idea of the amplitude of future changes in reference series but this quantitative information should be interpreted very carefully, since component series are not selected according to a strict quantitative criteria based on the cross-correlation with the reference series. The following graph compares the trend restored CLI with the industrial production in Japan over the twenty past years.

4 Months for Cyclical Dominance. For details, see [1] page 42.

Figure 2. Industrial production and leading indicator for Japan



Composite leading indicators for zones

In addition to CLIs for 22 individual Member countries, the OECD publishes CLIs for the following groups of countries or zones: OECD total, Major seven countries, OECD Europe, European Union, Big four European countries, NAFTA (North American Free Trade Agreement) and since October 1999, Euro area.

The reference series for a zone is the weighted average of the individual countries reference series. The weights are derived from GDP originating in industry and the GDP Purchasing Power Parities for 1995. The amplitude-adjusted CLI for the zone is calculated by weighting together the amplitude-adjusted CLIs for individual countries, using the same weights as for the reference series. The same method is used to obtain the industrial production trend for the zone. Then, the trend restoration is done as for individual countries, by multiplying the amplitude-adjusted composite indicator by the trend of the reference series.

3 Information Currently Available

The information on OECD CLIs is currently made available monthly in the following three publications: *Main Economic Indicators (MEI)* which is disseminated as an electronic product and as a paper publication, *Leading Indicators and Business Surveys (LIBS)* which is disseminated electronically and the *Monthly note of leading indicators* which is available as a printed edition and on the Internet (http://www.oecd.org/std/cli).

In addition to the monthly note, the OECD Internet web site contains general information on OECD CLIs such as methodological information, how they can be used, etc.

Two other publications should also be mentioned: the very latest figures for the trend restored CLI are available in the weekly *OECD Hot File* and graphs of historical CLI data for some area totals are published in the annual *OECD Historical Statistics*.

Note that the information disseminated on the web site is available free of charge, unlike information disseminated in paper publications and other electronic media.

This section describes in detail for each type of information how different issues are dealt with in the different publications. A distinction is made between the *historical data* that are little revised and that give information on the past cycles, and the *recent data* that may be more subject to revisions from one month to another and give information on the future developments.

Methodological information

This information is disseminated via the OECD Internet site. The primary purpose is to show how CLIs can be used in practice.

Regarding the general methodology, a brief description of the procedures of CLI calculation is given on this site. Reference is also made to different articles and publications related to the OECD methodology.

On the OECD Web site users can find updated statistical characteristics of the different OECD CLIs. Some of these characteristics are summarised in the last page of the monthly note. These are the MCD value, the CLI median lead, the mean deviation from median, the total number of components as well as the number of components available for the computation of the index for the last month.[5] Some of these characteristics are likely to be modified from one month to another. Other characteristics that are rarely modified are available on the web site, independently from the monthly note. They comprise the list of component series and the sequence of turning points for the reference series.[6]

Users are also provided with an example of analysis using OECD CLIs. This is done in section 4 of *"OECD Composite Leading Indicators: a Tool for Short-term Analysis".*

Links are also made to related papers.

Historical data

Historical data for the reference series, composite leading indicators and component series are available in the LIBS monthly diskettes. In this publication, currently, no explanation is given on the way the leading indicators are compiled. This means implicitly that only users who already have a good knowledge of OECD CLIs are able to use the data in an optimal way.

5 For individual countries, 40 % of the components is required to compile the CLI for a given month.

6 The sequence of turning points is provided only for the countries whose CLI has been recently revised.

Recent data

Recent data on trend restored CLI are published in the three main monthly publications cited above; a little more than one year is provided in MEI and the monthly note, and the whole time series in LIBS. In the monthly note, additional information is available to assist users in their analysis and interpretation of recent developments in CLIs. This consists of the monthly growth rate of the trend restored CLI, the 12 month percentage change at annual rate of the trend restored CLI along with the reference series long-term trend rate.

Recent data for the ratio to trend CLI as well as for the trend of the reference series and for the different forms (original, ratio to trend or difference from trend) of component series are disseminated in LIBS.

Summary of information available on each aggregation level

Table 1 summarises the information available. Each level is the aggregation of leading indicator series of the lower level. As one can see, individual country level is the level at which the information is the most developed, in particular, the statistical characteristics. For the most aggregated levels, that is, the area totals and OECD total levels, recent data are available on the three main dissemination media. Lastly, data for component level are available only on LIBS.

Table 1. Information available on each aggregation level

Level	Aggregation level	Type of information	Available on
4	Total OECD CLI	Statistical characteristics	Not available
		Historical data	LIBS
		Recent data	LIBS, MEI, Internet
3	Area totals CLIs	Statistical characteristics	Not available
		Historical data	LIBS
		Recent data	LIBS, MEI, Internet
2	Individual countries CLIs	Statistical characteristics	Internet (partially available)
		Historical data	LIBS
		Recent data	LIBS, MEI, Internet
1	Component series	Statistical characteristics	Not available
		Historical data	LIBS
		Recent data	LIBS

4 Users' Requests

Users are encouraged on the OECD web site, to request additional information and provide STD with feedback on the dissemination of CLIs. Most of these requests are sent by email to a generic account (stat.contact@oecd.org). Some users also contact STD by fax. These can be precise questions on a particular issue or requests for general methodology. Some users only send suggestions. Once they have been answered, these requests are stored in a database. In order to check the relevance of the CLI dissemination and to improve the product, a detailed examination of these requests has been undertaken. The most recent feedback has been taken into account in the preparation of this paper. This encompasses more than hundred requests received during the past two years.

The analysis of user's requests has been done according to the following two dimensions: by subject and by geographical area to which the request is related. It should be noted that the financial crisis in Asia as well as the introduction of the European single currency occurred during the period during when these requests were received. This explains why so many requests are related to these two subjects.

By subject

The requests are classified in Table 2 according to the subject mentioned. Note that in this paper, "general methodology" means main principles used in the compilation of the CLIs for every country, in contrast to "statistical characteristics" which are specific to each CLI, such as, for example, the list of component series or component countries, their weights or the average lead at turning points.

Table 2. Users' requests classified by subject

Subject			%
Methodology	General Methodology		**31.3**
	Of which	Request for program (FAME calculations)	7.8
	Statistical characteristics and aggregation methods		**35.9**
	Of which:	Area total	12.5
		Country	14.0
		Component	9.4
Data	Data (figures, turning points)		**29.7**
	Of which:	Recent data	9.4
		Historical data	20.3
Others	Statistical tools		**3.1**
	Of which	12 months percentage change	3.1
Total			**100.0**

About two thirds of requests (67.2%) concern methodology and are divided more or less equally between questions on general methodology (31.3%) and on the statistical characteristics of the CLIs (35.9%). Nonetheless, a number of requests (3.1%) are related to the statistical tools provided with CLI data to assist users in their interpretation.

The main reasons for users to request information on general methodology are for re-producing the OECD composite indicators; developing their own set of leading indicators, in particular for non-Member countries; teaching; comparing their own way of determining turning points with the OECD method. This emphasises the fact that many users contact the OECD in order to acquire OECD expertise and experience on the CLIs. In that context, dissemination of precise information about the compilation of the indicators and, above all, interpreting leading indicators within the framework of economic analysis is fundamental.

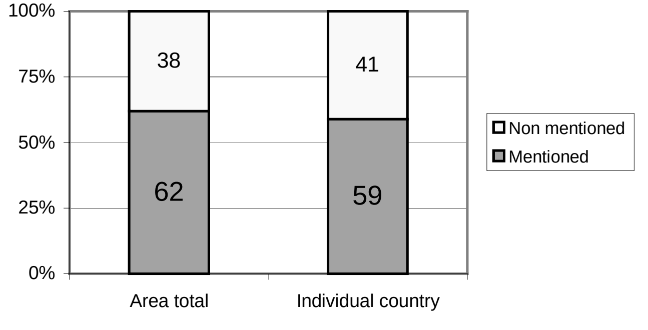
A large number of requests concerning statistical characteristics relate to the component series. This represents 9.4% of the total number of requests. This shows that users require regular information about the evolution of component series in order to explain developments in the CLI. For example, some users tried to determine the weight of share prices in the total CLI for Europe, in the context of the risk of the Asian crisis propagation from one region to the next.

Requests for data constitute only 29.7% of the requests received. This percentage is rather low compared to the proportion of questions on methodology but can be explained by the fact that a great deal of recent data is already available on the web site, as shown in Table 1. This is confirmed by the fact that two thirds of the data requests relate to historical data (20.3%). Questions about the current data often relate to recent or future turning points. Many users appear to be using data as input into a composite indicator or for studying reference cycles.

By type of geographic area

Figure 3 shows the percentage of requests for both categories of region (area total or individual country). It should be noted that some requests mention CLI from both categories.

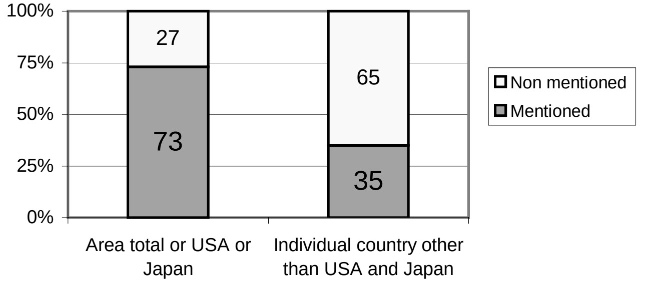
Figure 3. Users' requests by category of region



As mentioned in section 2 above, in addition to CLIs for 22 Member countries, CLIs are computed for the following area totals: OECD total, OECD Europe, European Union, Euro area, Four large European countries, Major 7 countries and NAFTA. 62 % of the requests mention at least one of these area totals. This is slightly higher than the proportion of requests mentioning at least one individual country (59 %).

Furthermore, the composite indicator for United States (30%) is clearly preferred to the one for NAFTA (3%) to illustrate the cyclical developments in North America. In addition, Japan is the only Asian country for which OECD provides a CLI. From this we can deduce that the CLI for these two countries are mostly used for comparison between the large world markets and can be equated with the area totals listed above. Taking this into account, the figure becomes:

Figure 4. Users' requests by category of region when ranking USA and Japan with area totals



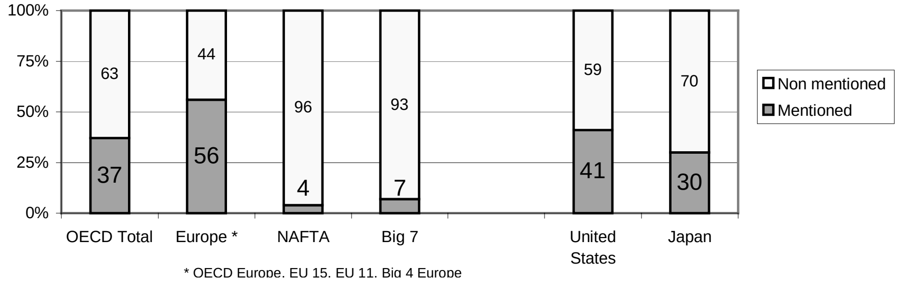
The proportion of requests mentioning an individual country other than United States and Japan is 35 %. Thus one can see the growing importance of international comparisons and the significance of economic analysis of geographic areas in the use of the OECD system of leading indicators.

Since the creation of the system of CLIs in the beginning of the 1980s, world trade has increased dramatically with economies becoming more and more open. Convergence of economies is the current line of development in each area. So, the evolution in area totals became more easily interpretable. This convergence has been reinforced by the creation of large regional markets such as NAFTA and the European Union in the beginning of the 1990s. Users are thus more and more interested in global developments in each area. They look for explanations for these developments as much in the various parameters of the economy (inflation, interest rates, share prices, production, employment, inter-area trade) as in the developments in each individual country of the area.

Since this is a system of leading indicators, information is provided for four different levels (OECD total, area total, country, component series). Therefore in addition to their turning point forecasting function, leading indicators, when formed in a system, can also be used as explanatory variables for the CLI aggregated at the higher level.

Figure 5 shows the proportion of requests concerning each group of countries, as percentages of the requests mentioning at least one group of countries (or the United States or Japan).

Figure 5. Users' request for large geographical area



The proportion of requests concerning the total of the Major 7 countries is very low (7%). This may be due to the fact that this group of countries is not a regional market and, on the other hand, that the total OECD (37%) is preferred as a proxy representation of the world economy. Europe is the area whose composite leading indicator has been mentioned most in recent years. This is mainly due to the creation of the European Monetary Union on 1 January 1999 and the introduction of the European single currency (Euro) now used in 11 European countries. At the same time, a composite leading indicator has been created for the Euro area.

Summary

What emerges from this examination can be summarised as follows. An attractive feature of the OECD work on leading indicators is the fact that they are based on a system with four levels of aggregation. It is possible to explain developments in the more aggregated CLI by studying developments at more detailed level. To allow this, users need detailed methodological information on the way each level is computed from the lower one. Data also need to be provided at each level. When individual countries economies are in a process of converging, CLIs for large regions are becoming more useful. This is the case, in particular, of the CLI for Europe when making comparisons between larger regional markets.

5 Efforts to Meet Emerging Users' Needs

In the light of the analysis presented in sections 3 and 4 above, STD proposes to review existing dissemination of information on CLI through consideration of the following options:

1. To develop new tools for comparison between large regional markets. This has begun to be implemented with the recent creation of a CLI for the Euro area.
2. To improve the information available on component series. This will be done by revising regularly the composition of CLIs for each individual country, and then keeping the users informed of the relevance of each component series.
3. To modify the presentation of the monthly note on leading indicators. In particular, the new note should enable better cross-country and cross-area comparisons as well as the interpretation of recent developments of the CLIs. The aim of this is to expand the audience of users of CLIs and to ensure that CLIs are used optimally.

This section details the efforts undertaken by STD regarding these three issues.

A CLI for the Euro area

Section 4 shows that users are more and more interested in CLIs for large areas, provided these CLIs are representative of a regional market. With monetary union, the Euro area is an economic entity in its own right. A CLI for this area has been created and has been included in the monthly note since October 1999. This paragraph describes its characteristics.

The method used to compile the CLI for Euro area is the same as for the other groups of countries. This means that the CLI is obtained by aggregating together the CLIs for individual countries. An alternative method would have been to select Euro area component series and compile the aggregated CLI using the same approach as for individual countries. The method chosen presents the advantage of being in accordance with the framework of the OECD system. Developments in the Euro area CLI can be interpreted by analysing developments observed at country level. This analysis could be complemented with analysis of the cyclical evolution in the European Central Bank (ECB) indicators.

The reference series is Total industrial production. It is calculated by weighting together the seasonally adjusted industrial production indices expressed on 1995 = 100 for the eleven individual countries. Weights in Euro area, derived from GDP originating in industry and the purchasing power parity for 1995 are as follows:

Table 3. Composition of the Euro area CLI

Country	% weight	Country	% weight
Austria	3.0	Ireland	1.7
Belgium	3.7	Italy	22.5
Luxembourg	0.3	Netherlands	5.0
Finland	1.7	Portugal	3.0
France	19.5	Spain	9.1
Germany	30.5	Total	100.0

Table 4 gives the sequence of turning points that has been retained. Turning points of minor cycle are in parenthesis. Note that initially, a trough was detected in May 1968 but since the reference series is significantly affected by strikes, the procedure was re-run using an estimate for this date instead of the original value. The actual trough is in August 1967.

Table 4. Cyclical characteristics of reference series

Turning point dates (months)		Amplitude	
Peaks	Troughs	(% of trend)	
	3/63	-5.47	
1/64		6.23	
	8/67	-10.54	
2/70		11.45	
	12/71	-7.27	
1/74		10.65	
	7/75	-12.62	
(1/77)			(8.61)
	(3/78)		(-6.42)
3/80		12.42	(10.27)
	12/82	-7.13	
1/91		5.93	
	7/93	-10.07	
12/94		7.44	
	12/96	-4.35	

The amplitude of a cycle at peak or trough is measured by the percentage deviation of the original series from the trend.

The amplitude-adjusted CLI for the Euro area is derived from the combination of the amplitude-adjusted CLIs of each individual country using the same weights as used to compute the industrial production index. 60% of the total weight of countries is required to compute the CLI for any given month.

The characteristics of this CLI are displayed in the following table. The median and mean lead at turning points as well as the mean deviation of the lead from the median are expressed as a number of months.

Table 5. Characteristics of composite leading indicator

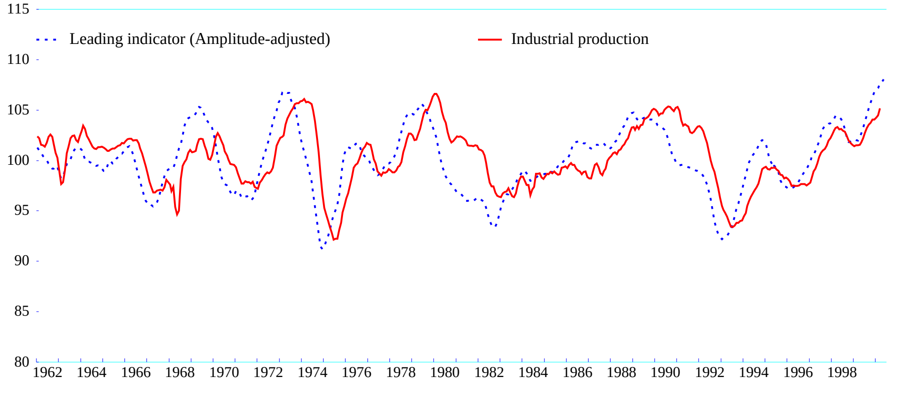
	Delay in release	Extra cycles	Start date	Mean lead at			Median lead at			Mean devia-tion from median	Cross-correla-tion (Lead: 6)
				All TP	Peaks	Troughs	All TP	Peaks	Troughs		
CLI	2	2	1962	5.8	7.7	4.3	6	9	3	3.4	0.81

This emphasises that this CLI is timely, the delay in release is two months. It predicts turning points in reference series 6 months ahead on average and the lead is significantly higher at peaks (median lead: 9 months) than at troughs (median lead: 3 months). The mean deviation from median is 3.4 months: i.e., the lead is not too dispersed. Two extra cycles are signalled during the period 1982-1992.

The trend of the reference series is obtained using the same aggregation method as for reference series and CLI.

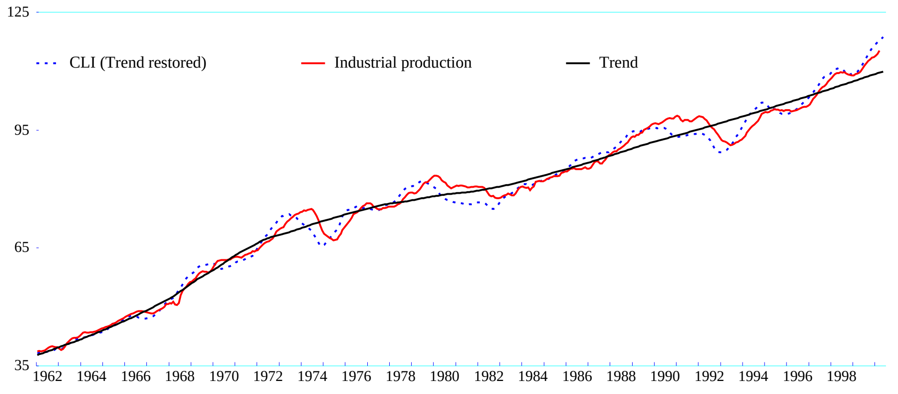
The following graph shows the MCD smoothed, trend-eliminated series of the reference series against the amplitude adjusted leading indicator.

Figure 6. Cyclical developments in Euro area, Ratio to trend (smoothed)



The following graph consists of reference series in original, seasonally adjusted and MCD smoothed form, and its trend, together with the trend-restored CLI.

Figure 7. Industrial production and leading indicator for Euro area



The examination of the cyclical pattern of Euro area can be divided into four periods:

1 Before 1972, amplitudes are relatively low.

2 The period 1972-82 is marked by two cycles of large amplitudes, the two peaks being in January 1974 and March 1980.

3 The 1980's are years during which amplitudes are not very pronounced. No major cycle occurred before the downturn of 1991.

4 In the recent period (from 1992) the reference series tend to have a more cyclical behaviour with medium amplitudes.

Improving information on component series

In 1996-97 the CLIs for the major seven countries as well as for Belgium and Norway were modified to take into account the transformations in national economies. At the same time, a CLI was created for Mexico.[7] The CLIs for the 12 remaining individual countries have not been modified since the creation of the OECD system in the beginning of the 1980's. For some of these countries, the composition of the CLI may have become less relevant than at the time the CLI was set up. STD plans to revise these indicators on a cycle of about 3 years.

Each modification to an individual country's CLI would be the opportunity to publish, along with the monthly note, a short study explaining the revision. The cyclical

[7] See [6] for more details. CLI for Mexico is an interesting illustration of how a CLI could have given clear and reliable warnings of a severe recession.

characteristics of the different component series would be presented along with the performance of the revised CLI. The comparison between the former and the new list of component series would give useful information on the evolution of this country's economy.

Regular revisions would also reflect the convergence between countries in the same area. We can expect the list of component series to become more harmonised from one country to another. This would facilitate the interpretation of developments in CLIs for area totals.

Facilitating the interpretation in developments in CLIs

The propositions developed in the previous paragraphs of this section were aimed at improving on an on-going basis the quality and the quantity of information available in order to provide the users with a set of tools as complete as possible. Another important matter is the presentation of the information. In recent months, discussions have been held within STD, and with staff in other Directorates of the OECD to determine a way of disseminating the information in a concise as well as pedagogical way. The aim of the discussions was to extend use of leading indicators into a broader audience of internal OECD and external users, including non-specialists. STD proposes to modify the presentation of the monthly note.

In its present version, the monthly note provides a very standard format. A short text summarising the main developments in CLIs is followed by a set of graphs and tables in which CLIs for every countries and area totals are dealt with in the same way. The last page of the monthly note gives the main statistical characteristics of the CLIs.

It is proposed to reduce the standard section of the note. The graphs showing cyclical evolutions would be published every month only for the large regions. The latest data would be summarised in an overview table containing only the very latest figures (less than 6 months) as well as some statistical characteristics for all countries and groups of countries.

At the same time, we would attach more importance to the description of the behaviour of the CLI of one particular country or group of countries. The choice of the country described could be in relation to the revision of the CLI as described in the previous paragraph, but could also be linked to current economic events. What STD feels is important is the necessity of providing the user with a reasonably detailed statistical analysis of the development of an economy in a given region. While providing detailed information, this would also provide an analytical approach which users may apply to other countries or areas.

6 Concluding Remarks

The number of users that have taken the effort to contact the OECD to provide useful feedback and the diversity of their questions is a good indication of the extent to which OECD is well known for providing useful information. In particular, the fact that OECD provides CLIs for a large number of countries and groups of countries appears to be very attractive. This emphasises that OECD makes good use of the fact that its statistical activities are based on international comparison.

In order to keep the OECD system of CLI relevant, STD needs to monitor users' requirements on an on-going basis. In particular, STD should continue to encourage feedback from users and to develop accordingly the information available. Precise elements for the interpretation of the results should also be disseminated to allow non-specialist to use this system in an optimal way.

STD will also have to incorporate into the CLIs developments in the world economy, by regularly checking the relevance of the series used in the compilation of CLI and by calculating new CLIs, in particular for new OECD Member countries.

The assessment of the methodology currently used to compile OECD CLIs is also an important issue for the coming years.

Bibliography

[1] OECD (1987) "OECD Leading Indicators and Business Cycles in Member countries, Sources and Methods 1960-1985", No. 39 January 1987.

[2] National Bureau of Economic Research (1971) "Cyclical Analysis of Time Series: Selected Procedures and Computer Programs", Gerhard Bry and Charlotte Boschan, NBER technical paper 20.

[3] OECD (1997) "Cyclical Indicators and Business Tendency surveys", OCDE/GD(97)58, General Distribution.

[4] OECD (1987) "OECD Leading Indicators", Ronny Nilsson, Economic Studies, No. 9.

[5] OECD (1996) "An update of OECD Leading Indicators", Gérald Petit, Gérard Salou, Pierre Beziz and Christophe Degain, Paper presented at the OECD Meeting on Leading Indicators, Paris October 1996.

[6] OECD (1997) "The 1994 Mexican Crisis: Were Signals Inadequate?", Pierre Beziz and Gérald Petit, Paper published in the Cambridge Review of International Affairs (Summer/Fall 1997, Vol. XI, No. 1).

[7] OECD (1994) "Leading Indicators for OECD, Central and Eastern European Countries", Ronny Nilsson, Article published in the book "Is the Economic cycle still alive?", edited by Baldassarri/ Annunziato, St Martins Press Inc., New York 1994/The Macmillan Press LTD, London 1994.

[8] OECD (2000) "Confidence Indicators and Composite Indicators", Ronny Nilsson, Paper presented at the CIRET conference, Paris October 2000.

[9] OECD (2000) "Calculation of composite leading indicators: A Comparison of Two Different Methods", Olivier Brunet, Paper presented at the CIRET conference, Paris October 2000.

[10] OECD, Main Economic Indicators, monthly publication.

[11] OECD, Monthly Note of Economic Leading Indicators.

Calculation of Composite Leading Indicators: A Comparison of Two Different Methods

Olivier Brunet

1 Introduction

This study compares the German and the Japanese composite leading indicators using two methods of trend estimation and two methods of data normalisation. These two composite indicators have been selected in order to show that the nature of the component series is important in choosing the calculation method of a composite leading indicator.

The indicator of the first country, Germany, is composed mainly of series from business tendency surveys (qualitative series). For the second country, Japan, the indicator is based mainly on series in level form (quantitative statistical series).

2 Presentation of the Indicators

Sixty-six percent of the German composite leading indicator is composed of business tendency surveys series: Order books: level, Order inflow of flow demand: tendency, Business climate indicator and Finished goods stocks: level. The two other component series are Share prices and Total new orders.

Table 1. Component series of the German CLI

	Starting date	Ending date
Qualitative series		
Stock finished goods: level	January 1959	June 2000
Order books: level	January 1963	June 2000
Orders inflow: tendency	January 1960	June 2000
Business climate: current	January 1961	June 2000
Quantitative series		
Net new orders	January 1962	June 2000
Share prices	January 1960	July 2000

The study is conducted on data from January 1964 to June 2000 in order to compile the composite leading indicator with all data figures available for all component series for the whole period. All series are seasonally adjusted.

The German composite leading indicator is very much restricted to components related to activity in the industrial sector of the economy. Out of the six components four refer to industrial activity as measured by the business survey in industry. The remaining two components refer to a volume series on new orders and the share price index.

The Japanese composite leading indicator only contains of two business tendency surveys series: Finished goods stocks: level and Business situation: prospects. The rest of the series (77%) are series in level form and include: Producer inventory ratio to shipment, Excess of imports over exports, Total stocks in manufacturing, New loans for equipment (commercial banks), Ratio loans to deposits, Share prices Tokyo stock exchange and New vacancies.

The study is conducted from January 1974 to April 2000 in order to compile the composite leading indicator with all data available for all component series for the whole period. From April 1998, the Japanese composite leading indicator is only calculated using eight components because the series New loans for equipment is only available until March 1998. All series are seasonally adjusted.

Table 2. Component series of the Japanese CLI

	Starting date	Ending date
Qualitative series		
Business climate: future tendency	2nd Quarter 1967	2nd Quarter 2000
Stock finished goods: level	1st Quarter 1963	2nd Quarter 2000
Quantitative series		
Inventory shipment ratio	January 1955	May 2000
Total stocks in manufacturing	January 1955	April 2000
New vacancies	January 1957	June 2000
Share prices	January 1959	June 2000
New loans for equipment	January 1955	March 1998
Ratio loans to deposits	January 1955	May 2000
Excess of imports over exports	January 1973	June 2000

3 Estimation of the Trend: Presentation of the Two Methods

Phase-Average Trend (PAT)

This first method estimates a trend in a direct way and removes it from the series. For the trend estimation, we use the Phase-Average Trend method.

The first consideration in the construction of a composite cyclical indicator is that of amplitude-stationarity. The first objective is to ensure that each individual indicator series included in the composite indicator is stationary in some way. The "growth cycle" or "deviation-from-trend" approach is used. Trend estimation is thus a crucial step in detecting cyclical movements and identifying turning points.

The PAT method requires an initial list of turning points, which define the cyclical phases, in order to estimate a trend which cuts through the phases. A first list of turning points is obtained from the preliminary peaks and troughs identified by calculating a first trend estimate based on a 75-month moving-average. A series of tests are then executed on the deviations from the trend in order to eliminate extreme values; and to specify a minimum duration of each phase (5 months) and the minimum duration of each cycle (15 months). The final trend is then calculated with a validated list of turning points as input to the PAT program. The detrended results (deviations from long-term trend) are then used as input to the composite calculation.

Period to Period Changes (PPC)

This second method removes the trend by transformation of data to stationary form by applying the period to period changes to data. This method has been applied to all quantitative series:

$$S_t = \frac{X_t - X_{t-1}}{(X_t + X_{t-1})/2}$$

with: S : Period to period changes series

X : Quantitative series

t : Time

Concerning the qualitative tendency business survey series, it is not necessary to apply the period to period changes or first difference method to this kind of series because data are first differences by nature. For business tendency surveys series, data fluctuate over a level which is a constant figure and therefore series do not have a trend. The balance i.e. the difference between positive and negative answers is used for the business tendency survey series.

4 Smoothing

Before the normalisation stage, we need to smooth both the quantitative series and the period to period change series. The method of smoothing is explained as follows:

It is necessary to ensure that all component series have equal "smoothness". This is to ensure that period-to-period changes in the composite indicator are not unduly influenced by irregular movements in any one indicator series. The "Months for Cyclical Dominance" (MCD) moving average is used. This procedure ensures approximately equal smoothness between series and also ensures that the period-to-period changes in each series are more likely to be due to cyclical than to irregular movements. The data lost at the end of the series due to the moving average are restored with an extrapolation by regression over the end of the series.

Table 3. MCD moving average results

		MCD (PAT)	Transformation	MCD (PPC)
Japan				
	Business climate: future tendency	1	No	1
	Stock finished goods: level	1	No	1
	Inventory shipment ratio	2	Yes	6
	Total stocks in manufacturing	1	Yes	6
	New vacancies	3	Yes	6
	Share prices	2	Yes	6
	New loans for equipment	4	Yes	6
	Ratio loans to deposits	2	Yes	6
	Excess of imports over exports	6	Yes	6
Germany				
	Stock finished goods: level	2	No	2
	Order books: level	1	No	1
	Orders inflow: tendency	5	No	5
	Business climate: current	2	No	1
	Net new orders	3	Yes	6
	Share prices	2	Yes	6

The MCD values for most series across investigated countries are in the range 1-3. A MCD value of 1 means that no smoothing is needed and this concerns only a few components: for Germany: Order books: level with exception of Order inflow: tendency, and for Japan: Business climate, Finished goods stocks: level and Total stocks in manufacturing. Most financial and other business survey indicators have an MCD of 2 or 3 while most quantitative statistical indicators show MCD values of 4, 5 or 6 such as Excess of imports over exports and Loans for equipment for Japan. The smoothing performed on the component series ensures that the composite leading indicators across all countries are smooth with a MCD value of 1, which means that no further smoothing is needed for an easy identification of a cyclical turning point.

In table 3, the MCD moving average results with the two different methods are presented for all series. The column "transformation" indicates whether the series has been transformed to period to period change series. The transformation has been made only for the quantitative series and the results show how the irregular component has increased for the change series which now have MCD moving average of 6.

5 Normalisation: Presentation of the Two Methods

Normalisation or standardisation of component series is necessary in order to minimise the influence of series with marked cyclical amplitude on the composite indicator.

Normalisation of Phase-average Trend series (NPT)

In this first method normalisation is performed on detrended series and involves two steps, first the mean of the MCD moving average series is subtracted to the MCD moving average series itself. Then this series is divided by the mean of the absolute values of the differences of the MCD moving average series from its mean. The normalised series are then converted into index form by adding 100. This method standardises the amplitudes of the cyclical movements but leaves the relative magnitudes of the irregular movements unchanged. We are going to apply this method to the method PAT. In the rest of the paper, the composite leading indicator compiled by the methods PAT then NPT is called the First method.

Normalisation of Period to Period changes series (NPP)

This second method normalisation is performed on the period to period change series and qualitative business tendency survey series. For period to period changes series, it is carried out by dividing the series by the mean of the absolute values of the period to period change data. For the business tendency surveys series, period to period changes method is not applied. Therefore, normalisation is performed by dividing the series by the mean of the absolute values of the qualitative series. The MCD moving average method (see part ??) is then applied to each component series (qualitative series and period to period changes series) in order to get a smoothed series. We are going to apply this method to the method PPC. In the rest of the paper, the composite leading indicator compiled by the methods PPC then NPP is called Second method.

composite leading indicators issued from the second method but the results were very difficult to interpret due to the irregularity of the composite series. Applying the MCD to each series really gives better results for interpretation.

6 Aggregation

The aggregation method (in order to calculate the composite leading indicators) contains three steps: lagging, weighting and then aggregation.

Lagging

It may sometimes be necessary to lead or lag particular indicators. In this study, no lagging is undertaken.

Weighting

Different weights may be assigned to component series in order to reflect their economic significance (coverage and economic reason), statistical adequacy, cyclical conformity, speed of availability of data, etc. The purpose of weighting is to improve reliability by giving higher weight to components with good quality i.e. indicators which correlate highly with each other and the resultant composite indicator.

In this study, equal weights are used to obtain each country’s composite leading indicator. Equal weighting implies, by default, a judgement on appropriate weights, and the normalisation process is itself a weighting system in reverse.

Aggregation

In this study, the raw composite index is obtained by averaging the normalised indices of each component series. A composite index calculated on an incomplete set of data is linked to the body of the index by use of a linking factor which is equivalent to applying the growth-rate of the "incomplete" index to the last point at which a full index is available. For the German (January 1964) and Japanese (January 1974) composite leading indicator, beginning dates are the same for all component series.

For Japan and Germany, two composite leading indicators by country are compiled with the two different methods described. The following section presents the graphical and empirical results of the study.

7 Country Results

Germany

Chart 1 shows the industrial production and the German composite leading indicator (ratio to trend series) compiled with the first method.

Chart 1. Industrial production and Composite leading indicator, First method

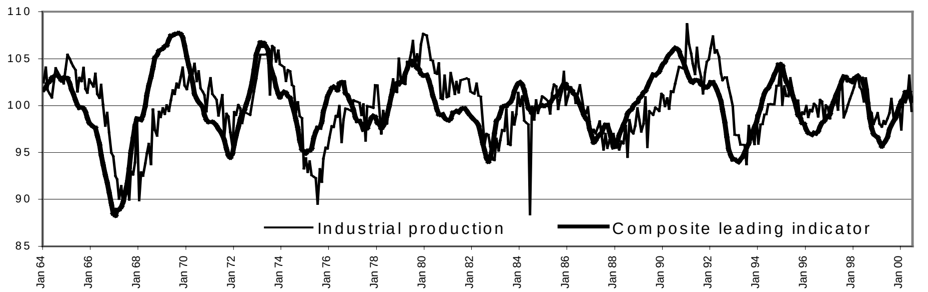
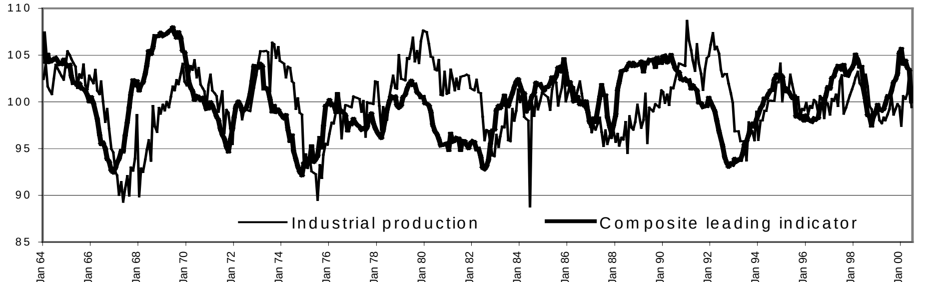


Chart 2 shows the industrial production and the German composite leading indicator (ratio to trend series) compiled with the second method.

Chart 2. Industrial production and Composite leading indicator, Second method



A comparison of the two leading indicators shows that the patterns are quite similar except for a brief period between 1988 and 1990.

In table 4, three composite leading indicators are presented and the study is conducted over the period 1963-1999. The first is the OECD composite leading indicator compiled with fixed turning points and the second is calculated with automatically selected turning points. These two composite leading indicators are compiled with the first method. The third one is compiled with the second method.

Table 4. Performance of German CLIs

	Performance 1963-1999 Against ratio to trend of industrial production						
	Turning point analysis					Cross-correlation	
	MCD	Median lag (+) in months at			Mean absolute deviation around median	Months Lag (+)	Peak value
Germany Composite Indicator		Peak	Trough	All turning points			
OECD (fixed turning points)							
Composite Leading Indicator: **OECD method**	1	5.5	5	5	6.0	8	0.75
Automatic turning points							
Composite Leading Indicator: **First method**	1	6	2	3	6.2	5	0.74
Composite Leading Indicator: **Second method**	2	9	5	6	7.7	8	0.54

The composite leading indicator compiled with the first method and automatic turning points has a lead of five months against the industrial production whereas the composite leading indicator compiled by the OECD shows a lead of eight months. The only difference between these two indicators is the method of determination of the turning points which points in favour of the OECD method. On the other hand, the composite leading indicator compiled with the second method shows almost the same good results as the OECD one except that the MCD equals 2. This value of MCD is due to the two quantitative series.

The second method improves the lead at turning points and in particular at peaks. But the cross correlation result is quite low (0.54). The first method does not bring any improvements to the two other ones for the criteria shown in table 4. In addition, the lead that gives the best cross-correlation is significantly lower than for the OECD method and the median lead at the troughs is only two months.

Japan

Chart 3 shows the industrial production and the Japanese composite leading indicator (ratio to trend series) compiled with the first method.

Chart 3. Industrial production and Composite leading indicator, First method

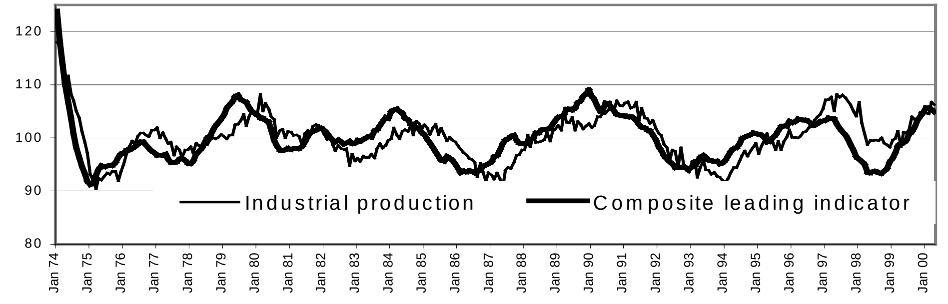
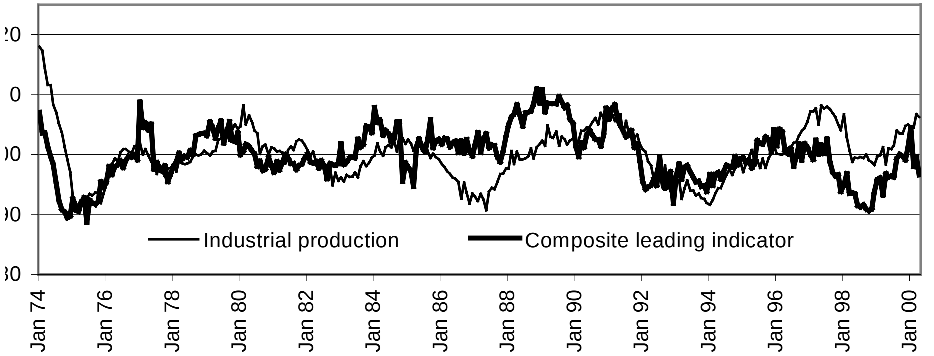


Chart 4 shows the industrial production and the Japanese composite leading indicator (ratio to trend series) compiled with the second method.

Chart 4. Industrial production and Composite leading indicator, Second method



For Japan, unlike the composite leading indicator for Germany, the differences between the two methods are significant. The second method gives a more irregular composite leading indicator and does not define troughs as well as the first method.

Table 5. Performance of Japanese CLIs

	Performance 1974-1999 Against ratio to trend of industrial production						
	Turning point analysis					Cross-correlation	
	MCD	Median lag (+) in months at			Mean absolute deviation around median	Months Lag (+)	Peak value
Japan Composite Indicator		Peak	Trough	All turning points			
OECD (fixed turning points)							
Composite Leading Indicator: **OECD method**	1	6	4	5	3.8	6	0.81
Automatic turning points							
Composite Leading Indicator: **First method**	1	9	4.5	6	4.7	4	0.77
Composite Leading Indicator: **Second method**	3	12	0	2	8.2	5	0.47

First of all, the results given by the second method are bad compared to the two other methods. Indeed, the composite leading indicator calculated using the second method is quite irregular (MCD equals 3). The high MCD value is explained by the fact that seven of the nine turning points are quantitative series with MCD equals 6. The cross-correlation is low (0.47) and the median lead to trough is 0.

On the other hand, the first method improves the forecast ability at turning points, especially at peaks (9 months against 6 months for OECD method) without improving too much the deviation around median. At the same time, the cross-correlation is slightly lower (at 77% for a four months lead).

8 Conclusions

This paper has examined the use of two different calculation methods for two countries, Germany and Japan, which each have different types of components series. The German composite leading indicator is mainly composed of business tendency survey series while the Japanese composite leading indicator mainly has quantitative series. The main conclusions are:

- The second method (normalised period to period changes) generates a rather irregular composite indicator with MCD of 2 or 3. This makes it difficult to identify promptly coming turning points. This behaviour is related to the type of component series used and the need for transformation to stationary series. The second method works well with qualitative business tendency survey series where no transformation is used.

- With fixed turning points, the first method (normalised PAT) seems to be the better method to compile a composite leading indicator whatever the set of components (quantitative or qualitative).
- However, trend estimation as used in the first method is more difficult to perform and revisions at the end of the series may change the trend slightly. This will however in general not change the dating of turning points.

Overall, this paper shows that choice of calculation method depends on the nature of the component series.

Bibliography

OECD, "OECD Leading Indicators and Business Cycles in Member Countries, Sources and Methods 1960-1985", No.39 January 1987.

OECD, "OECD Leading Indicators", Ronny Nilsson, Economic Studies, No. 9, Autumn 1987.

OECD, "Leading Indicators for OECD, Central and Eastern European Countries", Ronny Nilsson, Article published in the book "Is the Economic Cycle Still Alive?", edited by Baldassarri/Annunziato, St Martins Press Inc., New York 1994/The Macmillan Press LTD, London 1994.

OECD, "Cyclical Indicators and Business Tendency surveys", OCDE/GD(97)58, General Distribution; 1997.

OECD, "An update of OECD Leading Indicators", Gérald Petit, Gérard Salou, Pierre Beziz and Christophe Degain, Paper presented at the OECD Meeting on Leading Indicators, Paris October 1996.

OECD, "The 1994 Mexican Crises: Were Signals Inadequate?", Pierre Beziz and Gerald Petit, Paper published in the Cambridge Review of International Affairs (Summer/Fall 1997).

OECD, "Procedures for constructing composite indexes: a re-assessment", J. Cullity and A. Banerji, Paper presented at the OECD Meeting on Leading Indicators, Paris October 1996.

OECD, "Confidence indicator and composite indicator", Ronny Nilsson, Paper presented at the 25th CIRET conference, Paris October 2000.

OECD, "The OECD System of Leading Indicators: Recent Efforts to Meet User’s Needs", Benoit Arnaud, Paper presented at the 25th CIRET conference, Paris October 2000.

National Bureau of Economic Research (NBER), "Programmed turning Point Determination", Gerhard Bry and Charlotte Boschan, Paper published in Cyclical Analysis of Time Series: Selected Procedures and Computed Programmes, 1971.

Leading Indicators of Financial Vulnerability: Demonstrated in the Case of the Asian Crisis

Gernot Nerb and Markus Taube

1 Introduction and Summary

The outbreak of the "Asian crisis" during the first half of 1997 was unexpected for most observers. The crisis affected the economies that had shown the fastest growth and whose future had not been called into question by economists.[1] To be sure, in 1994 Krugman pointed out that the growth rates would have to decline,[2] but the danger of a far-reaching crisis was overlooked by the vast majority of observers.[3] Indeed, even in the mid-1990s proclamations lauded the beginning of an "Asian-Pacific century".

The surprising onset of the crisis and the collapse of the monetary and financial systems of several Southeast Asian economies raised the question of the predictability of such occurrences. Two opposing positions emerged:

- The Asian crisis was fundamentally unpredictable since there was no actual cause for the crisis. According to this position, certain market corrections were necessary in 1997. These, however, were carried out in such a poor way that a panic was touched off among market participants by policy mishaps and hasty government reactions, and that it was this that led to the actual crisis.[4]
- From the other standpoint, the Asian crisis was only the last logical link in a chain of misguided developments which were basically foreseeable.[5] The surprising onset of the crisis was thus the result of policy-makers, economists and corporations having observed the wrong indicators.

In the final analysis, however, only a synthesis of these two positions will be able to explain the complex events during the build-up to the outbreak of the crisis, its ensuing contagion and the final collapse of currencies and financial markets in the region.[6]

---

1 The most prominent account of economic development in the area is: World Bank (1993).

2 Krugman, Paul (1994).

3 For a notable exception see: Park, Yung-chul (1996).

4 This argument is brought forward by e.g.: Radelet, Steven; Sachs, Jeffrey (1998).

5 This view is best represented by: Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998:a and 1998:b).

6 The necessity of such a synthesis has been outlined by: Corbett, Jenny; Vines, David (1999).

In this article we argue that the Asian crisis was the result of a longer process which, through an accumulation of unfortunate developments and institutional mismatches, led the affected economies to stray far from a sustainable growth path, which is what made a fundamental correction urgently necessary. At its core the Asian crisis was more than merely a small market correction or a strong cyclical plunge. In terms of its predictability and avoidability, the following points hold true:

- The *structural imbalances in the crisis-affected countries* were diagnosable and these imbalances could have been corrected by political measures. The outbreak of the crisis was by no means unavoidable, but ultimately the result of an inadequate and too slow an economic-policy and institutional alimentation of the economic process in general and the actual problems of 1997 in particular.
- The ultimate *triggering and timing* of the crisis was not foreseeable, although the susceptibility to crisis of the affected economies and their constitution, in which even a slight – additional – misguided development could lead to a crisis out-break, could have been recognizable. In view of the interpretation difficulties which occur for a number of hard indicators based on the fundamentals, this would have been possibly achievable with the simultaneous observation of soft indicators in the form of business survey results, since the latter provided interpretation assistance with regard to the fundamentals.
- The *gravity of the crisis* could not have been prognosticated. Indeed, panic and a self-intensifying escape reaction seems to have worsened the crisis to the extent that was not explainable in terms of the strength/weakness profile of the affected economies. Mass psychology was involved to a considerable extent, a phenomenon that can be grasped best by so-called soft indicators such as data from business surveys. The occurrence of such an overreaction indicates that prior to the crisis the economic developments in the region had become severed from the fundamentals, and far-reaching risk analyses had received little attention. The collective panic from being caught unawares and the overreaction were thus already anchored in the run-up to the crisis. Accordingly, it is incorrect to maintain that the crisis was not real but only a panic reaction, since the building-up of a crisis potential and its panic-like realization are two sides of the same coin. This of course does not change our conclusions that, at the same time, the crisis was not inevitable. Even if the factor that ultimately triggered the crisis cannot be identified with absolute certainty, in our opinion there are clear indications that it was the combination of structural imbalances and the occurrence of a pessimism based in mass psychology that caused a cyclical weakening, which is common in market economies, to turn into a currency and banking crisis.

The conclusion that can be drawn for an early indicator system is that in addition to the traditional economic indicators for determining imbalances also the so-called soft indicators, i.e. the assessments and expectations of the economic decision-makers, must be followed intensively.[7] The business surveys, introduced by the Ifo Institute 50 years ago, can provide a major contribution, especially when the surveys are conducted on a representative and regular basis, preferably monthly. If such surveys are only done quarterly or less frequently, there is the danger that in the case of a strong weakening of the business climate, the suspicion first turns towards a statistical irregularity, and the subsequent survey is needed to confirm the development. In a quarterly survey a relatively long period of time is lost that would otherwise be useful for economic-policy measures. In the case of monthly surveys, this recognition lag is accordingly shorter.

2 Economic Fundamentals to be Monitored in a System of Vulnerability Indicators

In the following we try to identify some potential fundamental weaknesses that could make countries vulnerable for a crisis of their financial and currency systems.

Current account imbalances

If we accept Larry Summers' benchmark of a current account deficit in excess of 5% of GDP as dangerous[8], then the crisis countries did not make a very bad impression – with the notable exception of Thailand. As shown in chart 1 only Thailand featured a current account deficit of consistently more than 5% of GDP and was the only country in 1996 that lay beyond this threshold. Indonesia and Korea never overstepped the threshold while Malaysia and the Philippines experienced a rather volatile development covering both sides of the benchmark.

Even if the danger zone may be defined as lying below 5% of GDP, the existing current account deficits appeared to be sustainable if put into perspective with the exceptionally high growth rates the countries featured during the 1990s (chart 2).[9]

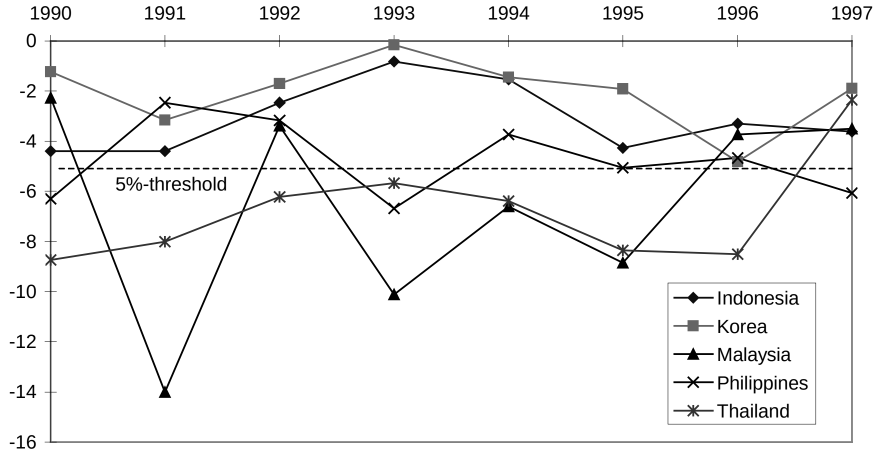
---

[7] The IMF itself points out that "[e]conomic theory, while relatively good at characterizing equilibrium situations, tends to be less informative about the dynamics that could lead from one equilibrium to another. To predict the timing of rare events such as financial crises, which may critically depend on factors that are hard to capture such as structural features of the economy, institutional developments, changes in the political landscape, *and expectations of domestic and foreign players in various markets, is likely to be even more demanding*. More important, the process of policymaking and the policy responses themselves have a crucial bearing on whether situations of stress degenerate into crises." International Monetary Fund (1998), p. 174. Italics added by the authors.

[8] In his Economist article on the Mexican crisis L. Summers claims "that close attention should be paid to any current-account deficit in excess of 5% of GDP". Summers, Lawrence (1996), p. 49.

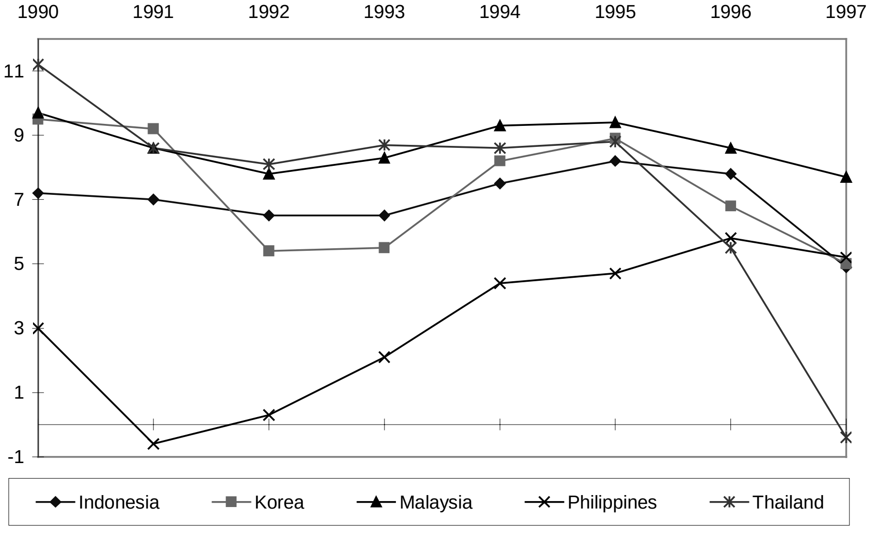
[9] In this respect the Philippines showed the weakest performance of the countries analysed, but the Philippines showed a strong upward trend and were eventually able to withstand the crisis comparatively well.

Chart 1. Current Account Deficits in East Asia, 1990-1997 (NIA Definition, in % of GDP)



Source: IMF (various): International Financial Statistics Yearbook, Washington.

Chart 2. Real GDP growth in East Asia, 1990-1997



Source: IMF (various): International Financial Statistics Yearbook, Washington.

Therefore the current account deficits by themselves did not indicate any imminent dangers, although Corsetti et al. are perfectly right to point out that "the observed high rates of growth may have contributed to downplaying the riskiness and costs of a strategy of excessive reliance on foreign capital and current account imbalances."[10]

In order to evaluate the sustainability of the current account deficits, it is therefore necessary to add some further variables to the analysis. Once again, a brief look at the investment ratios gives no reasons for concern. During the 1990s Indonesia, Korea, Malaysia and Thailand all featured high investment ratios of more than 30% or even 40% of GDP. The Philippines had a slightly lower investment ratio, but with more 20% of GDP theirs was also quite respectable. Therefore we have to take one further step in order to identify the Achilles heel of the crisis countries: The quality of investment was deteriorating. Rising incremental capital output ratio (ICOR) values at the macroeconomic level and low profitability of major players in the region at the firm level were clear indicators that the efficiency of investment came down and endangered the ability to meet the obligations accepted when building-up the current account deficits.[11]

Adverse real exchange rate developments

The development of the real exchange rate 1990-1996 is shown in chart 3. With the exception of Korea, which experienced a substantial depreciation of the real exchange rate in the early 1990s, all countries observed went through a significant real appreciation in 1996. We tend to interpret this real appreciation as a further misalignment and movement away from equilibrium and not as a movement towards a new equilibrium real exchange rate brought about by differential productivity developments in the economies observed and the USA.

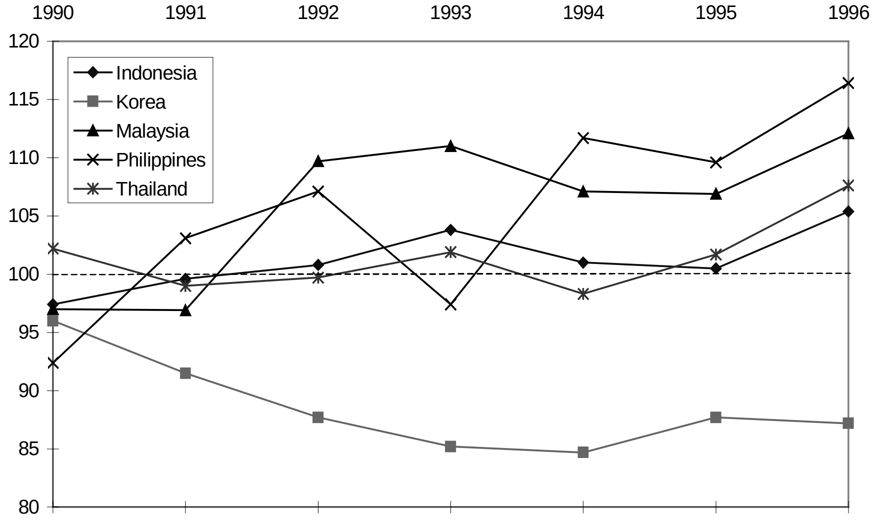
The countries observed had chosen to peg their currencies in a more or less strict fashion to the US dollar, allowing the nominal exchange rate not to fully accommodate inflation rate differentials. While this latter aspect of exchange rate policies was certainly asking for trouble in the long run, the policy of stabilizing the national currencies against the dollar was well put and for some time rather successful. Stable (and credible) exchange rate regimes made these economies credible for a reduction of currency risk premiums and furthered large capital inflows while a real appreciation allowed for a widening of interest differentials in their favor. All this helped make these economies highly attractive for foreign capital. And large inflows of foreign direct investment and comparatively cheap external financing were supposed to enable these economies achieve rapid capital accumulation, and translate into high productivity growth due to close interaction with and competitive pressure from the world market.

---

10 Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998:b), p. 12.

11 For a detailed account of the deterioration of investment efficiency and corporate profitability see Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998:b), pp. 12-16.

Chart 3. Real Exchange Rate Development, 1990-1996 (End of year data)



The base figure 100 is the average for the year 1990.

Source: JP Morgan, as quoted in: Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998:b).

Seen from this perspective the real appreciation did have a positive effect on the attraction of foreign capital, but its side-effects included a deterioration of export competitiveness and therefore posed a burden for the current account. The real appreciations brought about were eventually deviations from the equilibrium that were prone to external shocks. Slight changes in the external environment could push the real appreciation and in its wake the current account deficit to unsustainable levels and from there enforce a correction – either by political means or by crisis.

As a matter of fact, that is exactly what happened. In the years and months preceding the crisis a whole string of external events changed the external environment for the East Asian growth strategy and subjected it to corrective pressure: (a) the world business cycle slowed down putting a burden on exports; (b) in 1996 the world market prices of some of the main export goods of the region (especially semi-conductors) came under pressure inflicting negative terms of trade shocks on the economies; (c) the US dollar, to which the currencies of the countries observed were pegged, experienced a strong appreciation, consequently towing their currencies upwards vis-à-vis currencies outside the dollar bloc; (d) economic activity in the region was depressed by the ailing Japanese economy. The Japanese economy had been stuck in the doldrums for years and aborted its renaissance in spring 1997 by the badly timed introduction of a consumption tax, thereby quenching hopes for an upswing in economic activity in the whole region.

At the bottom-line we may conclude that the exchange rate policy followed during the early 1990s had made the region vulnerable for external shocks. In the middle of the decade and in the months preceding the crisis, worsening external parameters increased the pressure on the economies. And as one parameter after the other turned negative the crisis potential was slowly building up.

Structural deficiencies in regard to credit allocation and foreign debt

The real exchange rate development described above and the strong economic performance of the former years gave rise to a massive inflow of foreign capital and increasing foreign debt. The result was an explosion of bank credit. During 1990-1997 the growth rates of credit expansion were well in excess of those of GDP.[12] This lending boom was accompanied by unfortunate developments both on the side of credit allocation for specific uses (investment projects) and on the side of the credit structure.

A lending boom generally implicates over-investments, i.e. the creation of production capacities in excess of the equilibrium level, and an asset-price inflation, in the case of East Asia primarily in terms of real estate. This development of a worsening of the quality of capital investments and an increasingly large risk burden in the credit portfolios of financial intermediaries was additionally exacerbated in the crisis-affected countries because of an insufficiently regulated financial system and because of the moral hazard that business practices were increasingly based on due to explicit and implicit guarantees of state agencies.[13] In the previous years of rapid economic growth, the institutional construction of national financial systems in the region was not able to keep pace with the demands of growing liberalization of cross-border economic activity.[14] Ultimately the financial intermediaries alimented only an unsustainable economic growth. The automatically accompanying accumulation of non-performing loans finally posed the core of the financial crisis.[15]

That this financial crisis was able to develop into a currency crisis[16] lies in the fact that the national financial intermediaries had refinanced a large portion of their domestic loan grants on the international capital market – and what is worse to a growing portion on short term. The second problem therefore concerns the composition of foreign debt. As shown in chart 4, there was an enormous build-up of short-term debt in the years and months leading to the crisis. In Korea, Indonesia and Thailand short-term debt was significantly in excess of the foreign exchange reserves of the country. The lowest ratio was to be found in Malaysia, but here the picture is tainted by a strong acceleration in the growth of this indicator. All in all the best picture was offered by the Philippines, with a rather high but decreasing ratio.

---

[12] Dickie, Paul (1998), p. 7.

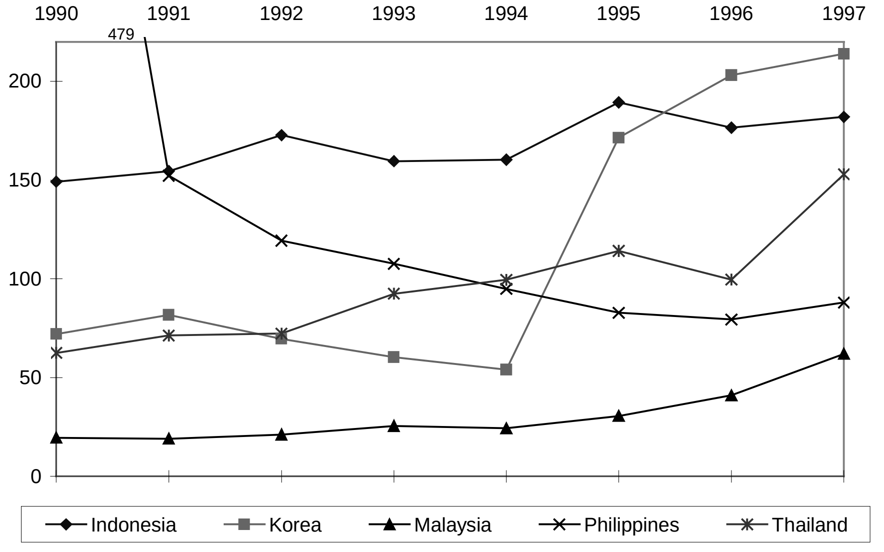
[13] Krugman, Paul (1998) and: Corbett, Jenny; Vines, David (1999), pp.160-161.

[14] IMF Staff (1998).

[15] See e.g.: Kindleberger, Charles (1978).

[16] Corbett, Jenny; Vines, David (1999), pp.156-157.

Chart 4. Short-term debt in % of foreign reserves



Source: World Bank, as quoted in: Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998:b); Data for Summer 1997 from: Bank für Internationalen Zahlungsausgleich (1998).

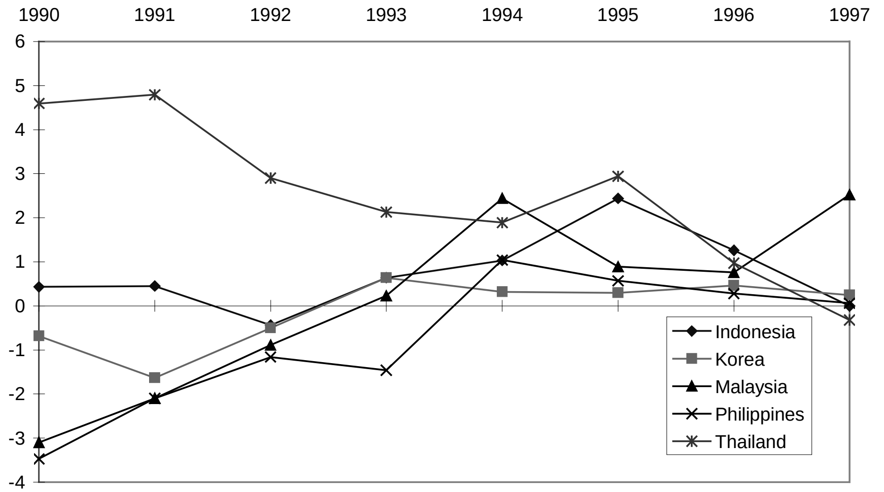
The region had become extremely vulnerable for a change of sentiment on behalf of foreign lenders. In case they refused to roll-over their short-term credits – and that is exactly what happened eventually – Korea, Indonesia and Thailand would default as their liabilities surpassed their foreign exchange holdings. A financial crisis would automatically turn into a currency crisis, which in turn would aggravate the financial crisis. Now the prerequisites for a total economic collapse were all in place.

Fundamental data that did not provide any clue of the imminent danger of crisis include the governments' fiscal balances and the absolute amount of foreign debt accumulated in the region.

Unsustainable fiscal deficits?

A priori strongly growing fiscal deficits should be characteristic for economic "crisis"-candidates. But as can be seen from chart 5 the fiscal balance was positive during the three years preceding the crisis in all countries.

Chart 5. Government Fiscal Balance, 1990-1997 (in % of GDP)



Source: IMF (various): International Financial Statistics Yearbook, Washington.

This positive picture of fiscal position is qualified, of course, by the accumulation of non-performing debt in the financial system,[17] insofar as here a latent burden for the government budgets was in position which in the case of an outbreak of crisis with the accompanying bail-outs would have to lead to a massive worsening of the fiscal balance. This burden was only latent however and did not necessarily need to become reality.

Excessive foreign debt?

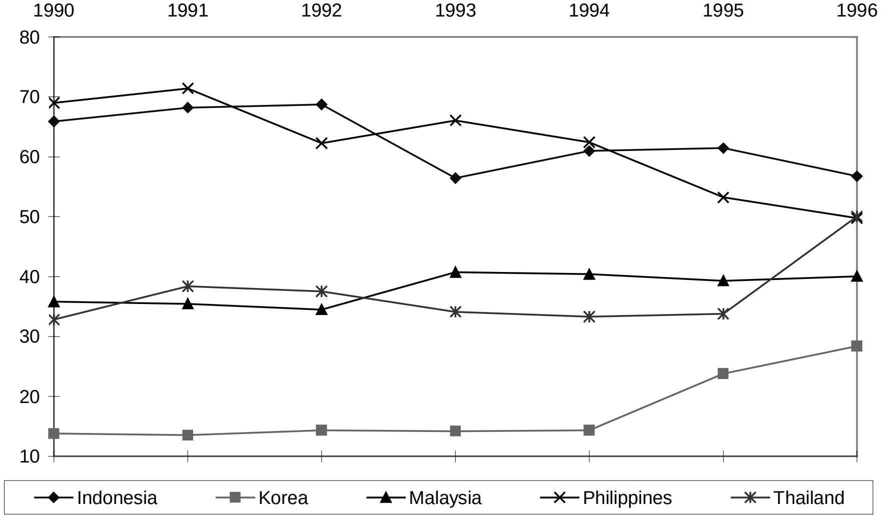
When looking at the World Bank data on the volume of foreign debt available at the outbreak of the crisis in Summer 1997, there is hardly a sign of imminent danger (chart 6). The ratio of foreign debt to GDP had been relatively stable since 1990. Indonesia and the Philippines, which had seen a rather high level of foreign debt in 1990, had been moving on a decisively downward directed trend. Adverse signals were coming only from Korea, which had seen a steep increase in the ratio of foreign debt to GDP in 1995 and 1996 and

---

17 Non-performing loans as a proportion of total lending amounted in 1996 to 13% in Indonesia, 8% in Korea, 10% in Malaysia, 14% in the Philippines and 13% in Thailand. Source: Bank für Internationalen Zahlungsausgleich (1997): Jahresbericht, Basel.

from Thailand, where the ratio was jumping upwards by 50% from 1995 to 1996. But even in these countries the ratios were still in an acceptable range.[18]

Chart 6. Foreign debt as % of GDP



Source: World Bank, as quoted in: Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998:b).

3 Soft Factors Triggering the Crisis

Up to now we have been trying to identify the extent to which the crisis-affected economies had become vulnerable for a financial or currency crisis. The weak (and in some respects deteriorating) fundamentals were a necessary precondition for the outbreak of the crisis. But they alone cannot explain why the crisis struck in July 1997 or why the crisis developed to the disaster observed in the following months.

The indicators shown above are possibly nothing but rather fuzzy pointers with respect to the well-being or vulnerability of an economy. They have to be interpreted and transformed into decisions/actions by managers who are influenced by various sets of information, value systems and schools of thought. The threshold value of a particular indicator, beyond which a company retreats from a certain market, still has to be determined in terms of subjective assessments of the overall situation in the local and the

---

18 It should be noted that on the basis of their research Corsetti et al. suggest that these World Bank estimates may have been seriously downwardly biased. Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998:b), p. 33.

world economy as well as with regard to company strategies and objectives. For example, the decisive factor in the evaluation of current account deficits is not the size of the deficit in absolute figures but rather the subjective assessment of the extent to which this deficit is sustainable and can be turned around in the future. Soft factors, such as business confidence, are therefore of prime importance for the outbreak of a crisis.

At this stage the Ifo-type business surveys may enter the picture. These surveys are mainly based on the experience that economic actions, in particular investments in real estate, machinery and equipment but also portfolio investments depend more on the perception of future trends than on ex-post facts. The same type of economic data may trigger at one stage of the economic process an increase, in another stage a decrease of investment activity. What causes the difference is the way decision makers assess future trends. To give an example: International investors in Southeast Asia should have realized at least since the middle of the 1990s that asset prices – in particular for real estate and equities – have risen in dramatically. Nevertheless, as long as the perception prevailed that the Asian tiger countries offered an enormous economic growth potential, these extraordinarily high asset price levels did not prevent the further inflow of new capital. Economic theory gives no hint of how long such overoptimistic perceptions of future trends will last. However, the Ifo-type of qualitative surveys focusing on the assessments of economic trends – rather than the exact figures – has proved to be a good instrument for monitoring the waves of optimism and pessimism, as reflected in such perceptions.

Within qualitative business surveys there are two main approaches to measure the impact of more psychological factors on economic decisions. One can be characterized as the bottom-up, the other one as the top-down approach. The bottom-up approach is the main line of these surveys. The underlying philosophy is the fact that economic decisions within a company are not a quasi automatic function of hard economic data such as present incoming orders but the assessment of future demand trends. Or to give another example: The same level of unsold finished goods may be regarded at one stage of the business cycle as too high and trigger a cut in production and a reduction of the number of persons employed, whereas in another stage of the business cycle the same level of stocks may be regarded as appropriate or even too low, thus triggering an increase in production and in the number of persons employed. The usefulness of these indicators based on managers' assessments and expectations to forecast short-term economic trends within the next six to twelve months has been demonstrated in many research papers.[19] An example of the usefulness of these indicators is given in chart 7, which shows the Ifo Business Climate for the overall economy and GDP. The lead of the business climate is not perfectly stable but is as a rule never shorter than about three months and reaches six months on average. (When taken into account that the official GDP data are published with a time lag of at least three, very often up to six months – not to mention the revisions afterwards – the effective lead of the business climate indicator is even significantly longer.) The generally agreed usefulness of these business surveys in constructing leading indicators is the reason that these surveys are in the meantime spread

[19] A good summary of the usefulness of business surveys is given in: EU Commission (1997). A link of business survey data to the financial development is given by: Nerb, Gernot; Stock, Kurt (2000).

throughout the world. As an example of business survey data (bottom-up approach) in Southeast Asia chart 8 depicts the Business Climate Indicator (all industries) for the Republic of Korea. This indicator is based on about 2900 replies coming from about 1700 manufacturing and 1200 non-manufacturing companies. (The indicator is calculated as the mean of the balances of the positive and negative replies to the two questions on the current and expected business situation of the company; to avoid negative figures 100 is added.)

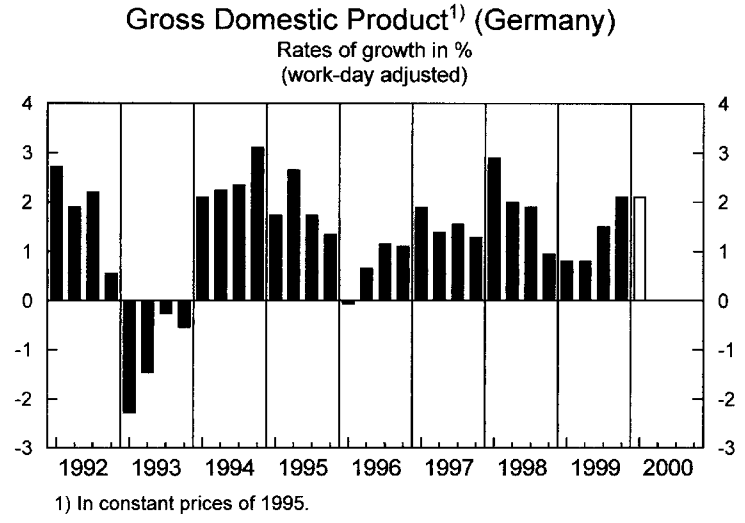
As demonstrated in chart 8 the indicator peaked already in the 3rd quarter 1995 and displayed thereafter a clearly downward trend till the 4th quarter 1997. Thus, one can argue that early signals for a deterioration of the economic trend were available long before the crisis started in autumn 1997. However, one has to admit that the leading indicator only signaled a pronounced economic downturn but that it was difficult to assess at this early stage the intensity of the crisis.

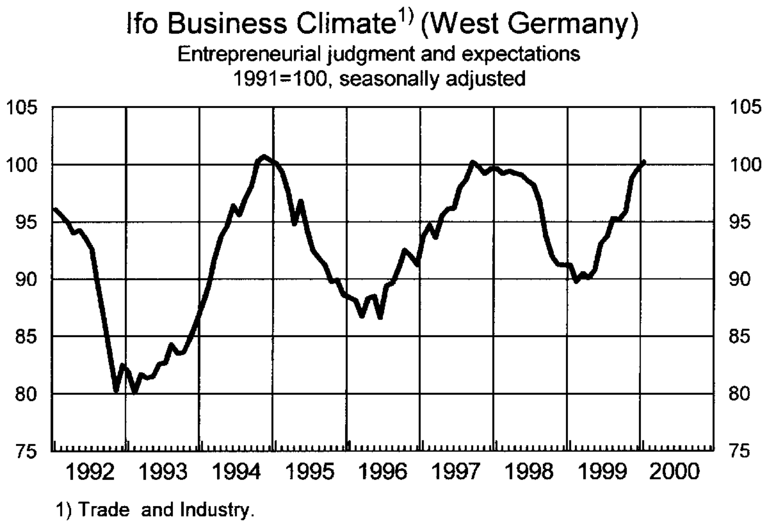
The second type of qualitative business surveys, labeled above as top-down approach, is demonstrated in the lower part of chart 8 and in chart 9. Here the replies come from managers of multinational companies who assess the global economic situation of the country under consideration. As shown in chart 9 the assessment of the overall economic situation in the case of Korea peaked in autumn 1995 and revealed a clearly downward trend, reaching a trough in early 1998 and staying there for almost the full year 1998 before recovering in the first quarter of 1999. The expectations series (columns in chart 9) gave signals even as early as the beginning of 1994 that economic growth in the region was slowing down and showed a sharp decline in early 1996. However, as in the case of the discussed bottom-up indicator (upper part of chart 8) it has to be admitted that though the direction of the business trend was indicated with a long lead there was no clear hint for the intensity of the downswing resulting in a financial crisis.

In a similar way as in Korea the bottom-up indicator of the Ifo Economic Survey International (ESI) gave early warning signals in Thailand and in Singapore. However, the lead period was rather short in Malaysia and in the Philippines: The indicator measuring the assessment of the current economic situation plunged in both cases dramatically between the 4th quarter 1997 and the 1st quarter 1998 without showing clear signs of a beginning slowdown before that period.[20]

[20] An analysis of the prognostive power of the ESI in respect to the Asian Crisis is provided in: Hilpert, Hanns Günther; Taube, Markus (1999).

Chart 7. Business Cycle Indicators





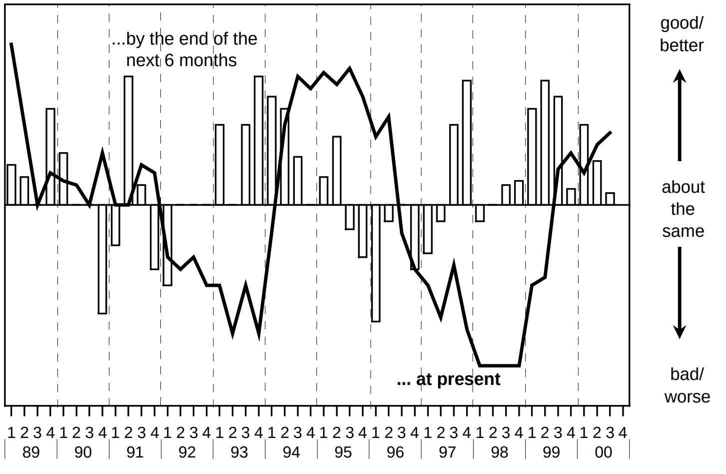
Sources: Federal Statistical Office, Ifo Business Survey.

Chart 8. Korea Republic

Business Climate Indicator (all industries).
Balance of positive and negative answers, concerning current business situation.

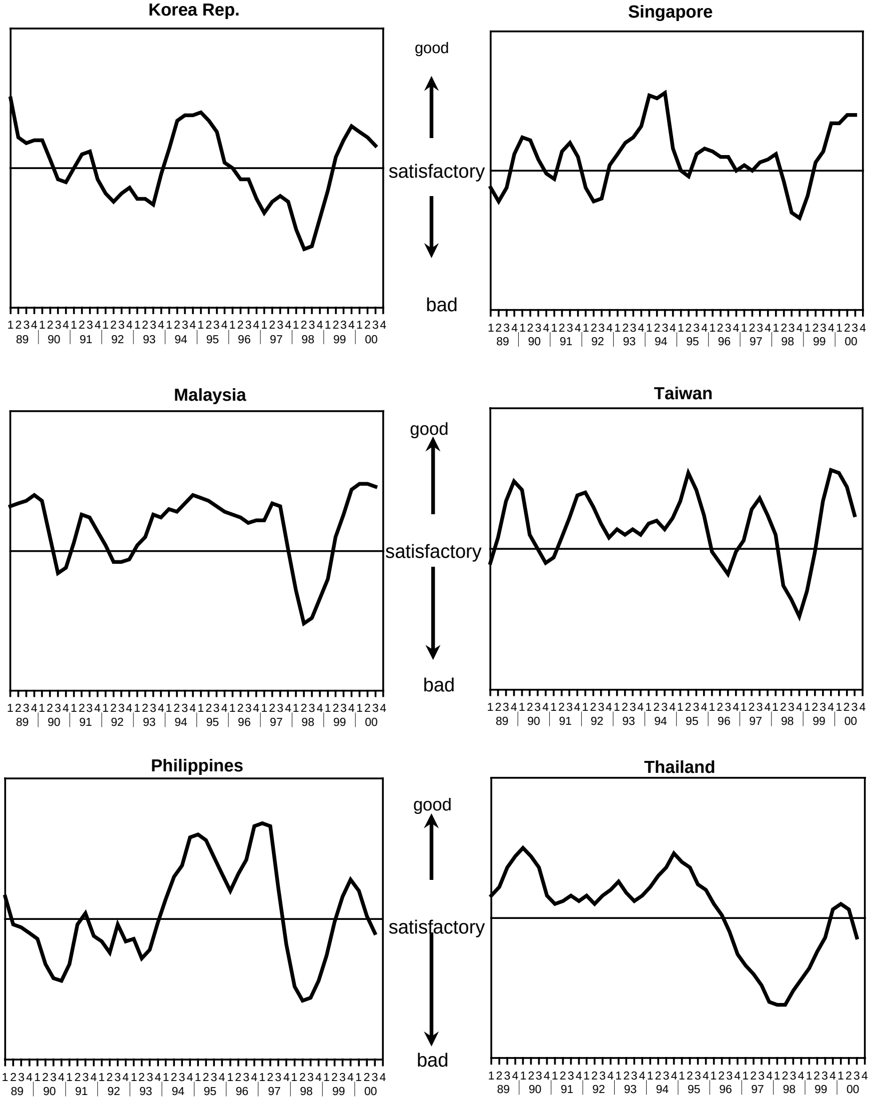
Source: Bank of Korea.

Economic situation ...



Source: ifo Economic Survey International, 3/2000.

Chart 9. Overall Economic Climate[1)]



[1)] geometric mean of current and expected economic situation, seasonally adjusted and smoothed

Source: ESI 3/2000

4 Conclusion

A major factor in the sudden withdrawal of trust in the future development of the region was a shift in the evaluation of some characteristics of the East Asian economies. The close relations and concerted actions between politics and entrepreneurs, which until then had been regarded as something not really understood but very successful nonetheless, as an empirical look demonstrates, were suddenly reinterpreted as crony capitalism and highly detrimental for long-term economic development.

That this change of sentiment occurred precisely in July 1997 cannot be attributed to individual occurrences. Instead, it was a chain of negative responses. Finally a critical mass was reached and fundamental positions that had not previously been questions were subject to close scrutiny.

This change in attitude led to a comprehensive re-evaluation of economic data. And as one indicator after the other moved beyond the revised threshold, a gold-rush mentality turned to panic, followed by a stampede out of the region. International speculators certainly speeded up and aggravated this process. But they did not wreak havoc on economies that had been perfectly sound until then. Speculators can only "attack" those economies (better: their currencies) that have built up inconsistencies in their economic setup and are prone for corrections.

In an ex post analysis it is quite possible to identify the underlying weaknesses of the economies struck by the crisis and to show what forces led to the outbreak and contagion of the crisis. But the question remains why the crisis was allowed to break out in the first place. Was it not possible to identify the state of affairs before the disequilibrium could lead to a crisis? Which indicators had been overlooked or were there no indicators at all that pointed at the fundamental weaknesses underlying the economic success of these countries? As a matter of fact, the picture is rather bleak. Most common measures for the well-being of an economy gave no clear indication of a lingering crisis.[21]

As shown above, the current account by itself was no feasible indicator for the lingering crisis, just as well as the investment ratio or the volume of foreign debt per se. Also no indication of a looming crisis was to be seen with respect to the fiscal position of the countries concerned. Only when these individual observations are seen in connection with additional indicators does a potential for danger emerge. It is such combinations of individual indicators that provide the best picture of the growing danger behind the facade of unclouded growth euphoria since in this way it can be tested how robust the superficially positive signals of individual indicators are.

In the context of the Asian crisis, informative composite indices are above all those that juxtapose non-performing loans with the phenomenon of the lending boom, and those that relate the current account position with real exchange rate developments.

An indicator that pointed to a fundamental worsening of the economic situation in the crisis-affected countries at a very early stage is the price earnings ratio determined for the

[21] Shirazi, Javad K. (1998).

local stock markets. This ratio reached its highest values in 1993-1994 and fell thereafter continuously. But a general observation of this indicator is deceiving since, on the one hand, it was unable to predicate the exact beginning of the crisis and, on the other hand, can be interpreted to say just the opposite, namely that a lowering of the price earnings ratio corrects previously unrealistically overvalued stocks thus reducing the danger potential. Nevertheless, based on Krugman's interpretation of the factors that touched off the crisis, it was precisely this indicator which is the key for understanding (and forecasting) the Asian crisis. From this perspective, severe moral hazard led to excessively risky lending operations which created asset price inflation. These inflated asset prices allowed the books of the financial intermediaries to appear better than they actually were. When, however, asset prices slowly declined, the true situation of the financial services sector and the misguided developments on the macroeconomic level became apparent. Analysts and observers were induced to rethink their assessment of the available indicators of the economic situation.

But the earliest warning signals came from the business survey results. Though also not a perfect leading indicator for economic crises, these qualitative data appear to be well-suited to capture changes in economic assessments of local but also international managers, not least the international fund managers, who appear to be especially vulnerable to irrational trends and prone to stampede in the other direction. Such changes in sentiment do not come overnight, however, but are expressed some time in advance in assessments and expectations. This makes business surveys so important for an early detecting of looming crises which, despite the importance of structural imbalances, frequently have their roots in mass psychology. Early recognition with the help of such qualitative indicators is all the more effective if the times between surveys is short. Especially advantageous are monthly surveys.

References

Bank für Internationalen Zahlungsausgleich (Various): Jahresbericht, Basel.

Corbett, Jenny; Vines, David (1999): Asian Currency and Financial Crises: Lessons from Vulnerability, Crisis, and Collapse, in: The World Economy, Vol. 22, No. 2, S. 155-177.

Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998a): Fundamental Determinants of the Asian Crisis: A Preliminary Empirical Assessment; http://www.stern.nyu.edu/globalmacro/jimf06.pdf.

Corsetti, Giancarlo; Pesenti, Paolo; Roubini, Nouriel (1998b): What Caused the Asian Crisis? Part I: A Macroeconomic Overview; http://www.stern.nyu.edu/globalmacro/AsianCrisis.pdf.

Dickie, Paul (1998): The Real Culprit Behind the Asian Crisis, In: ADB Review, Vol. 30, No. 4, pp. 5-9.

EU Commission (1997): The Joint Harmonised EU Programme of Business and Consumer Surveys, European Economy, No. 6, Brussels.

Hilpert, Hanns Günther; Taube, Markus (1999): Crisis and Transformation Reflected in the Changing Sentiment of Economic Experts in East Asia - What can be Expected for 1999?, Paper Prepared for the 14th LVMH Conference "Crisis and Transformation in Asia: Implications for Western Corporations?", Fontaine-Bleau, February 5-6, 1999 (Mimeo).

IMF Staff (1998): The Asian Crisis: Causes and Cures, in: Finance & Development, Vol. 35, No. 2, here download from: www.imf.org.

International Monetary Fund (1998): International Capital Markets. Developments, Prospects and Key Policy Issues, Washington.

International Monetary Fund (Various): International Financial Statistics Yearbook, Washington.

Kindleberger, Charles (1978): Manias, Panics and Crashes. History of Financial Crises, London.

Krugman, Paul (1994): The Myth of Asia's Miracle, in: Foreign Affairs, Vol. 73, 6 (November/ December), pp. 62-73.

Krugman, Paul (1998): What happened to Asia, http://web.mit.edu/krugman/www/disinter.html.

Nerb, Gernot; Stock, Kurt (2000): The Usefulness of Ifo Business Survey Results for Financial Market Analysis, in: Use of Survey Data for Industry, Research and Economic Policy, Selected papers presented at the 24th CIRET Conference, Wellington, New Zealand, 1999, Ashgate 2000.

Park, Yung-Chul (1996): East Asian Liberalization, Bubbles and The Challenges from China, Brookings Papers on Economic Activity 2, Washington.

Radelet, Steven; Sachs, Jeffrey (1998): The Onset of the East Asian Financial Crisis, http://www2.cid.harvard.edu/hiidpapers/eaonset2.pdf.

Shirazi, Javad K. (1998): The East Asian Crisis: Origins, Policy Challenges, and Prospects, http://www.worldbank.org.

Summers, Lawrence (1996): Ten Lessons to Learn, in: The Economist, December 23rd 1995 - January 5th 1996, pp. 48-50.

World Bank (1993): The East Asian Miracle. Economic Growth and Public Policy, Washington.

The Belgian Industrial Confidence Indicator: Leading Indicator of Economic Activity in the Euro Area?

Jean-Jacques Vanhaelen, Luc Dresse and Jan De Mulder

Abstract

The international press has recently reported on the widely-held view in the financial markets that the movement of the Belgian industrial confidence indicator might precede the euro area business cycles. The initial purpose of this paper is to assess whether this market perception is more than a simple optical illusion, resulting from the inspection of graphical representations of the data. For that, explicitly formalised methods are used to identify the timing of turning points in the industrial confidence indicators for Belgium and for the euro area, and the statistical significance of the differences in timing has been assessed using the Randomization Test proposed by Banerji. We conclude that the turning points in Belgium do in fact significantly lead turning points in the euro area from 1993 onwards.

The leading nature of the Belgian industrial confidence indicator is not really surprising, as changes in the business cycle stages in Belgium seem to have been ahead of changes in the euro area during the period from 1985 to the first quarter of 2000. Among the three different reference series used to compare the business cycle movements in Belgium and in the euro area, the null hypothesis that turning points in Belgium do not lead those in the euro area is rejected at a confidence level above 90 p.c. in the case of GDP and of the degree of utilisation of production capacity in manufacturing industry. The leading nature is more pronounced for the sub-period beginning with the first quarter of 1993, especially in the case of GDP. However, the comparison of the movements of the industrial production indices does not confirm these conclusions.

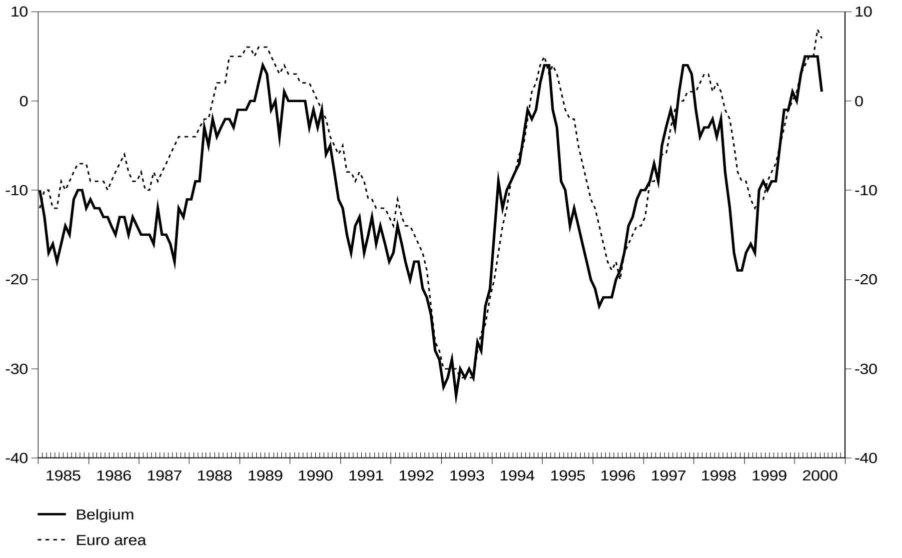
Due to the lack of sufficiently long time series for the euro area it was not possible to check whether differences in the economic structure could explain the leading nature of the activity in Belgium. However, using partial industrial confidence indicators, three factors (specialisation in intermediate goods, openness and high representation of small and medium-sized enterprises) that might explain why the Belgian business indicator and Belgian activity seem to lead their euro area counterparts were investigated, but could not be validated by the data.

As it seems to be impossible to identify one or more sectors or groups of enterprises accounting for the leading nature of Belgian economic activity when looking at turning points, at least when using the business survey data, it looks as if this leading nature is a kind of general feature of the Belgian economy.

Introduction

In order to conduct its monetary policy, the Eurosystem (composed of the ECB and the eleven members of the euro area) needs reliable and up-to-date information on the business climate in the euro area. The outcomes of the monthly business surveys are drawn upon heavily by the ECB as a leading indicator of economic activity that is readily available.

Chart 1. Industrial confidence indicator



Source: EC.

In an attempt to anticipate monetary policy decisions, players in the international financial markets are paying more and more attention to the business indicators for the euro area. Recently, some of them have discovered that the Belgian business indicator gives a good approximate idea of the business climate in the euro area and, more importantly, that it is a few months ahead of the euro area indicator.[1]

In the first part of the paper, we shall assess the robustness of the relationship between the industrial confidence indicators for Belgium and for the euro area. The analysis will focus on turning points. These are identified by using two procedures: the Bry

[1] See for example articles in the Wall Street Journal (14 July 1999), Le Monde (13 July 1999) or the Handelsblatt (24/25 September 1999).

and Boschan procedure and a procedure based on a moving median transformation of the series.

A simple non-parametric test, presented by Banerji at the 1999 CIRET Conference[2] , is used to calculate the confidence level of the difference in timing of the turning points in the series for Belgium and for the euro area.

The most obvious reason which might explain the leading nature of the Belgian indicator lies in a leading nature of the movements in economic activity itself, compared to the rest of the euro area. In the absence of an official, widely agreed dating of the business cycles for Belgium and for the euro area, the second part looks for systematic timing differences in the cyclical movements of three reference variables: GDP, industrial production indices and the degree of utilisation of production capacity.

In the third part, some breakdowns of the global industrial confidence indicators are investigated, in order to validate or reject the explanations often put forward for the leading nature of the Belgian industrial confidence indicator and of Belgian economic activity in relation to the euro area. These explanations are related to the structure of the economy, in terms of sectors, of openness to foreign markets and of size of enterprises.

Finally, some conclusions are drawn. They are to be interpreted bearing in mind the limitations of the study. Some approaches for future work are also presented.

1 Turning Points in the Belgian and the EUR 11 Industrial Confidence Indicators

Turning points are of particular importance in macroeconomic analyses, since it is quite obvious that rapid detection of changes in the direction of the movements of an economic variable can increase the effectiveness of economic policy decisions. As far as economic activity is concerned, the turning points are referred to by the so-called "peaks" (when the acceleration in the growth rate of economic activity comes to an end) and "troughs" (when the growth rate of economic activity stops decreasing). These points are of specific interest, because they delimit the stages (expansion, slowdown) in the business cycle .

In this paper we focus on the turning points in order to determine whether or not the industrial confidence indicator for Belgium leads the indicator for the euro area. Over the period from 1985 to mid-2000[3], only a few points in the time series can be regarded as turning points and consequently taken into account for the analysis.

Typically, two kinds of problems arise when a comparison of the timing of turning points between two series is attempted. First, the turning points in the business cycle component of the series have to be identified. Second, the robustness of the differences in timing has to be assessed. The following two sections present the methods used in this

2 Banerji (2000).

3 The cut-off date for the data was July 2000.

paper to address these issues. There are two different industrial confidence indicators for Belgium: the NBB synthetic curve and the EC confidence indicator. We briefly present these in section 3, before comparing the timing of the turning points of the EC industrial indicator for Belgium and for the euro area.

1.1 Identification of the turning points

The dynamic of most of the variables measuring economic activity, such as production or value added, sales or expenditures, includes a trend component which needs to be excluded in order to extract the business cycle. This should not be the case with confidence indicators, since the questionnaire is designed to elicit answers expressed as a deviation from the normal evolution. Nonetheless, seasonal factors and irregular variability also affect responses. Some signal-extracting methods have to be used to identify the turning points in the business cycle component of the series.

In this paper, two methods were applied to detect turning points. They both try to adhere to the following general principles:

- the choice of turning points should not be affected by aberrant observations or outliers;
- irregular movements in the series should be excluded;
- the identification method should minimise the risk of false signals or missing turning points in the business cycle;
- the application of the method should not delay the identification of the turning points for too long;
- some additional requirements may be imposed in accordance with stylised facts characterising business cycle movements, such as a succession of peaks and troughs, a minimum length for the phases and for the cycles, and possibly a minimum amplitude for the movements.

The first method we use is the well-known Bry-Boschan algorithm that has been extensively applied in the analysis of turning points for much of the last three decades.[4] It involves, first of all, the detection of extreme values and their replacement by a moving average. Second, an initial set of turning points is identified in a smoothed series, by applying a moving average filter. Using gradually shorter moving averages, the analysis shifts back to the original data in the immediate vicinity of the potential turning points. The turning points detected at the different stages are checked for the alternation of peaks and troughs, for a minimum span of 15 months between two successive peaks or two successive troughs, and of 5 months between two successive turning points. Finally, turning points in the first or the last 6 months of the series are rejected.

The originality of the second method, the NBB method, lies mainly in the smoothing method used. First, a centred moving median over five months is used to exclude outliers. Second, the resulting series is smoothed by a centred 5-month moving average with

---

[4] Bry and Boschan (1971).

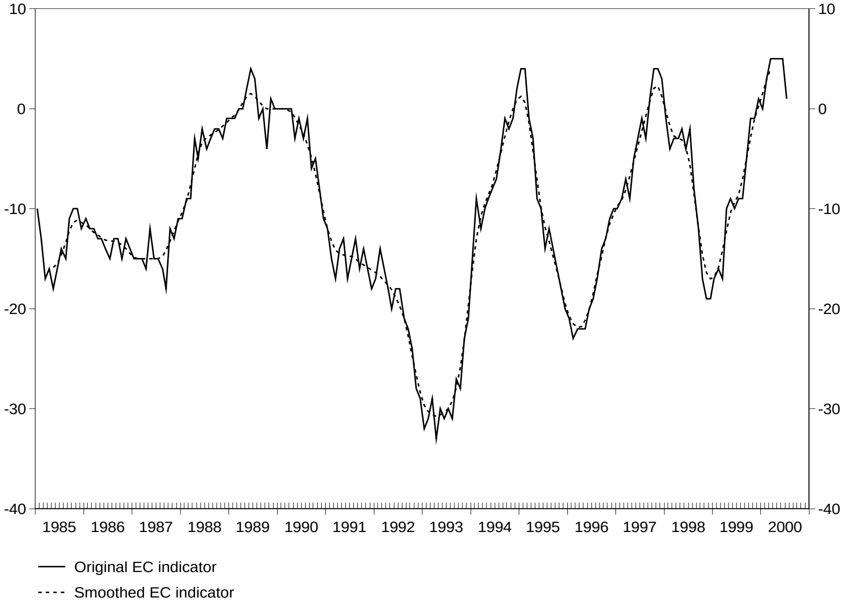
weights of 1/8, 1/4, 1/4, 1/4, 1/8. In addition, three restrictions are imposed on the local extremes of the smoothed series in order to identify the turning points:

1 alternation of peaks and troughs;

2 minimum of 5 months between two successive turning points;

3 minimum of 15 months between two successive peaks or two successive troughs.

Chart 2. NBB smoothing method - Industrial confidence indicator for Belgium



Source: EC; own calculations.

Applied for ten years in the field of business survey indicators in Belgium, the NBB smoothing procedure has proved its value in providing a useful representation of the cyclical movements, even for series where there are considerable irregularities in the short run.[5] Unlike the Bry-Boschan method, the NBB method does not shift back to the original series in order finally to identify the turning points.

5 NBB (1990).

1.2 Testing for leads in the turning points

Comparing the turning points identified in two series is not sufficient for properly assessing the leading nature of one of these series. As the direction or the length of the lead may vary from one turning point to the next, we need some method for measuring the statistical significance of leads. However, standard statistical techniques do not apply to this problem, because of the small number of cycles (and thus of turning points) usually covered by the series.

In a paper for the 1999 CIRET Conference, Banerji (2000) presented a non-parametric test for the differences in timing at cyclical turns, applicable to matched pairs of samples. In the so-called Randomisation Test, the null hypothesis that the differences are not statistically significant is to be tested against the alternative hypothesis that the leads are significant. The test can be conducted for increasing lead span, giving the confidence level at which the null hypothesis is rejected in favour of the alternative hypothesis that the difference in timing at turns significantly exceeds one, two, ... months, the full set of confidence levels thus resulting in a "lead profile chart", to use Banerji's terminology.

1.3 NBB definition versus EC definition of the Belgian industrial confidence indicator

There are two industrial confidence indicators for Belgium. In addition to the confidence indicator calculated by the EC according to the specification of the joint harmonised EU programme of business and consumer surveys[6] and published in the European Economy Supplement B, the National Bank of Belgium calculates and publishes its own synthetic indicator of confidence in manufacturing industry every month. Actually, both the EC industrial confidence indicator for Belgium and the NBB synthetic indicator are derived from the same business survey, conducted monthly by the Bank with a panel of some 1800 enterprises. The sources of differences between the two series are threefold:[7]

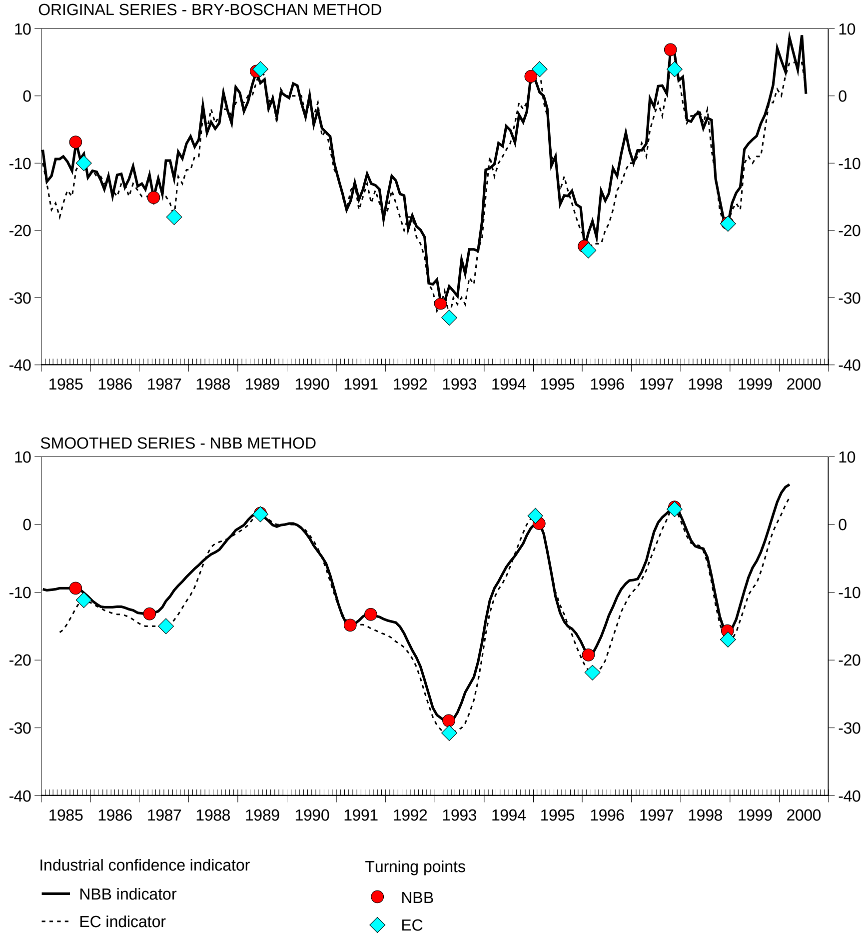
1. the choice of partial indicators (questions) taken into account for calculating the aggregate confidence indicator: the NBB includes eight questions in its synthetic confidence indicator, whereas the EC considers only three questions, namely on production expectations, assessment of order books and assessment of stocks of finished products (with inverted sign);
2. the methods used for seasonal adjustment;

[6] European Commission (1997).

[7] The convention for date labelling of the results is another source of difference. The National Bank of Belgium labels as results for the month t-1 the outcomes of the survey concerning the situation in month t-1. This survey is held in the first days of the month t, and published at the end of month t. The European commission labels the same outcomes as results for the month t, with publication at the beginning of t+1. In this paper, the EC's dating method has been applied to the NBB indicator.

3 the published series: in addition to the gross synthetic curve, the NBB publishes a series smoothed by the method described in section 1.1.

Chart 3. Comparison of the NBB and the EC industrial confidence indicators for Belgium



Sources: EC, NBB; own calculations.

Overall, the movements in the NBB synthetic curve and in the EC confidence indicator are very close to each other. The NBB curve seems to reach turning points a little in advance in the case of the original gross series, but the advance is less than one month in the case of the smoothed series.[8]

Among the five different partial indicators included in the NBB's synthetic indicator but not in the EC confidence indicator, it seems that the question concerning the movement of export orders systematically leads.[9] Conversely, the questions concerning the assessment of the export order book and concerning employment expectations lag, while the question about the movement of production and of domestic orders coincides with the global synthetic indicator.

For the sake of comparability with the euro area series, this study focuses on the industrial confidence indicator data for Belgium published by the European Commission. This is probably the series used by the financial markets, too. In any case, if the examination of the timing of turning points in the EC indicator shows a leading nature for Belgium, this conclusion certainly applies for the NBB synthetic curve as well.

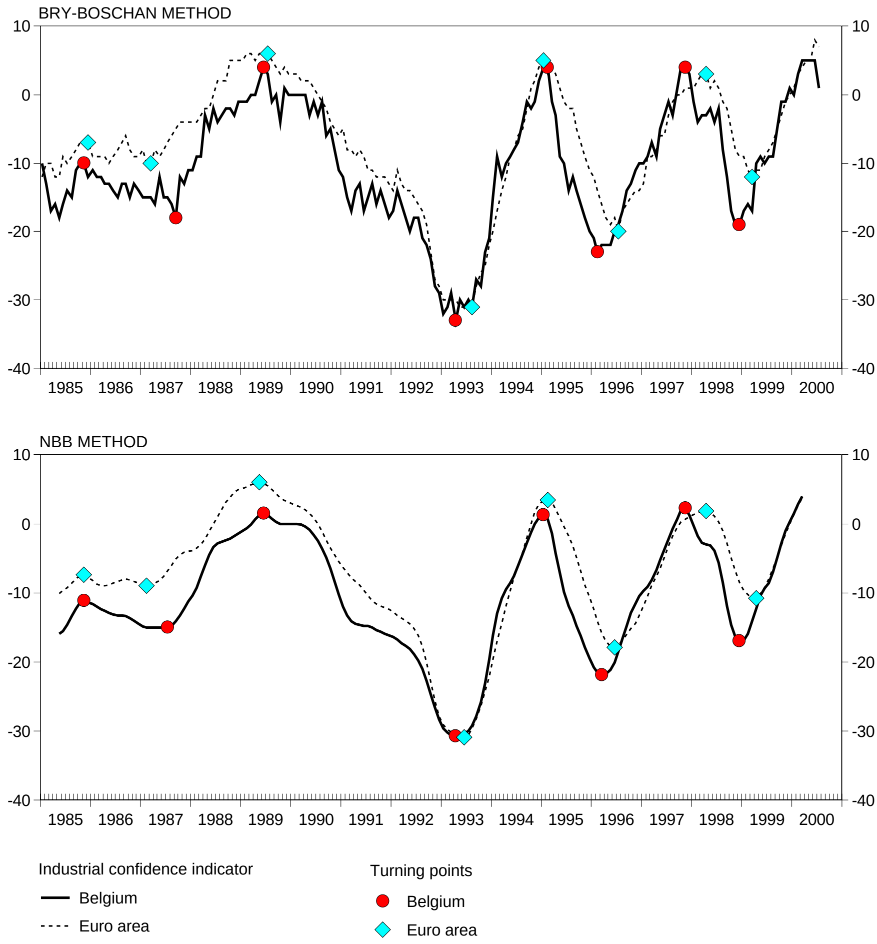
1.4 EC industrial confidence indicators for Belgium and the euro area

The timing of the turning points in the EC industrial confidence indicators for Belgium and for the euro area is compared using both the Bry-Boschan and the NBB methods of turning-points identification and applying the Banerji test for timing differences. Chart 4 shows that the cyclical movements of the industrial confidence indicators are very similar in Belgium and in the euro area. In both cases a distinction can be made between two sub-periods: since the beginning of 1993 the business cycles have been pronounced, with a rapid succession of upward and downward phases of large amplitude, while between 1985 and 1992 the cyclical movements were less sharp. In particular, in that period the irregular movements are substantial in the original series for Belgium; they are filtered by the NBB smoothing method.

8 The confidence level for the rejection of the null hypothesis of no leading nature for the NBB indicator is 99% for the original series, but 81% for the smoothed series. The respective confidence levels are 84% and 26% for the null hypothesis of no more than one month lead.

9 It would be worthwhile investigating whether the introduction of a question concerning the movement of export orders in the European Commission's harmonised survey, and hence in the calculation of the industrial confidence indicator, would not result in a quicker identification of turning points.

Chart 4. Industrial confidence indicators in Belgium and in the euro area



Source: EC; own calculations.

Over the period as a whole both the Bry-Boschan and the NBB procedures identify eight turning points, delimiting four cycles for Belgium and also for the euro area. For each of the economies, the results of both identification methods are quite similar. Comparing the timing of the turning points between Belgium and the euro area, and considering the results of the two methods together, Belgium leads in 11 cases out of 16 and lags in 4. The differences range from -6 to +5 months. Since the beginning of 1993 turning points in Belgium have preceded those in the euro area 9 times out of 10.

Table 1. **Industrial confidence indicators for Belgium and the euro area: Turning points**

	Bry-Boschan method			NBB method		
Turning points	Belgium	Euro area	Lead (+) / Lag (-) [1)]	Belgium	Euro area	Lead (+) / Lag (-) [1)]
Peak 1	Nov. 1985	Dec. 1985	1	Nov. 1985	Nov. 1985	0
Trough 1	Sept. 1987	Mar. 1987	-6	July 1987	Feb. 1987	-5
Peak 2	June 1989	July 1989	1	June 1989	May 1989	-1
Trough 2	Apr. 1993	Aug. 1993	4	Apr. 1993	June 1993	2
Peak 3	Feb. 1995	Jan. 1995	-1	Jan. 1995	Feb. 1995	1
Trough 3	Feb. 1996	July 1996	5	Mar. 1996	June 1996	3
Peak 4	Nov. 1997	Apr. 1998	5	Nov. 1997	Apr. 1998	5
Trough 4	Dec. 1998	Mar. 1999	3	Dec. 1998	Apr. 1999	4

Source: EC; own calculations.

1) Lead (+) or lag (-) of the Belgian turning points, in months.

According to the Randomisation Test, the confidence level for the rejection of the null hypothesis of no leading nature for the industrial confidence indicator for Belgium is around 80% for the whole period 1985-2000[10], but above 90% for the period 1993-2000.

Table 2. **Test for the leading nature of the Belgian industrial confidence indicator**

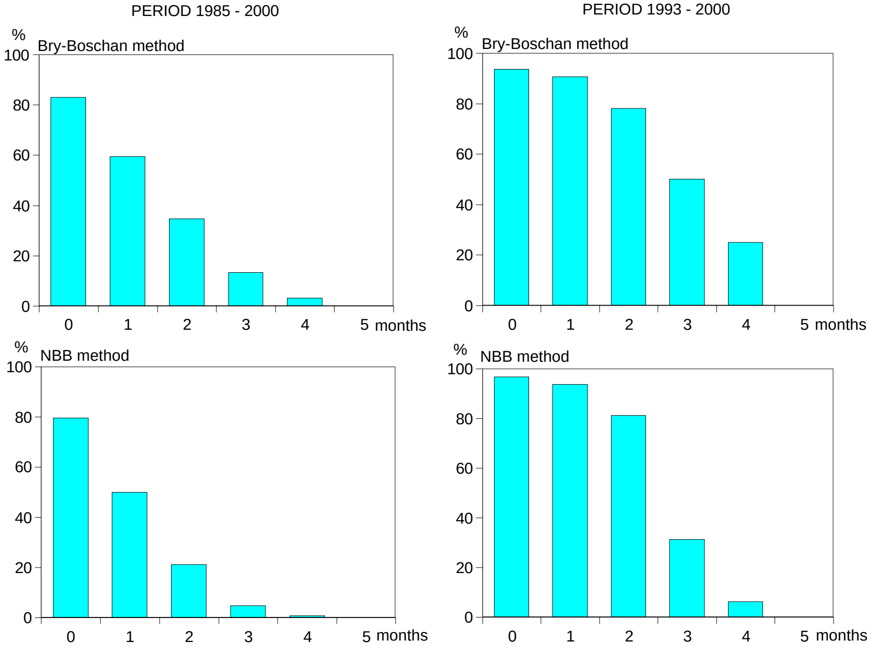
	Period 1985-2000	Period 1993-2000
Bry-Boschan method	83.2%	93.8%
NBB method	79.7%	96.9%

Source: EC; own calculations.

From the lead profile chart giving the respective confidence levels for the rejection of the null hypothesis of no leading nature for increasing timing differences, it appears that the confidence level decreases rapidly in the case of the period 1985-2000. For the period from 1993 to 2000 it remains above 90% for a difference of one month and close to 80% for a difference of two months. Lead profile charts show the confidence level in terms of different durations of lead. In Chart 5 this implies, for instance, that for the period 1993-2000, using the Bry-Boschan method, the null hypothesis of no leading nature can be rejected at a confidence level of 93.8%, the null hypothesis of a lead of at most 1 month can be rejected at a confidence level of 90.6%, the null hypothesis of a lead of at most 2 months can be rejected at a confidence level of 78.1%, etc.

10 For the period 1985-2000, the hypothesis of a lag of at least one month is rejected with a confidence level above 90%.

Chart 5. Industrial confidence indicator: Lead profile of Belgium in relation to the euro area[1)]

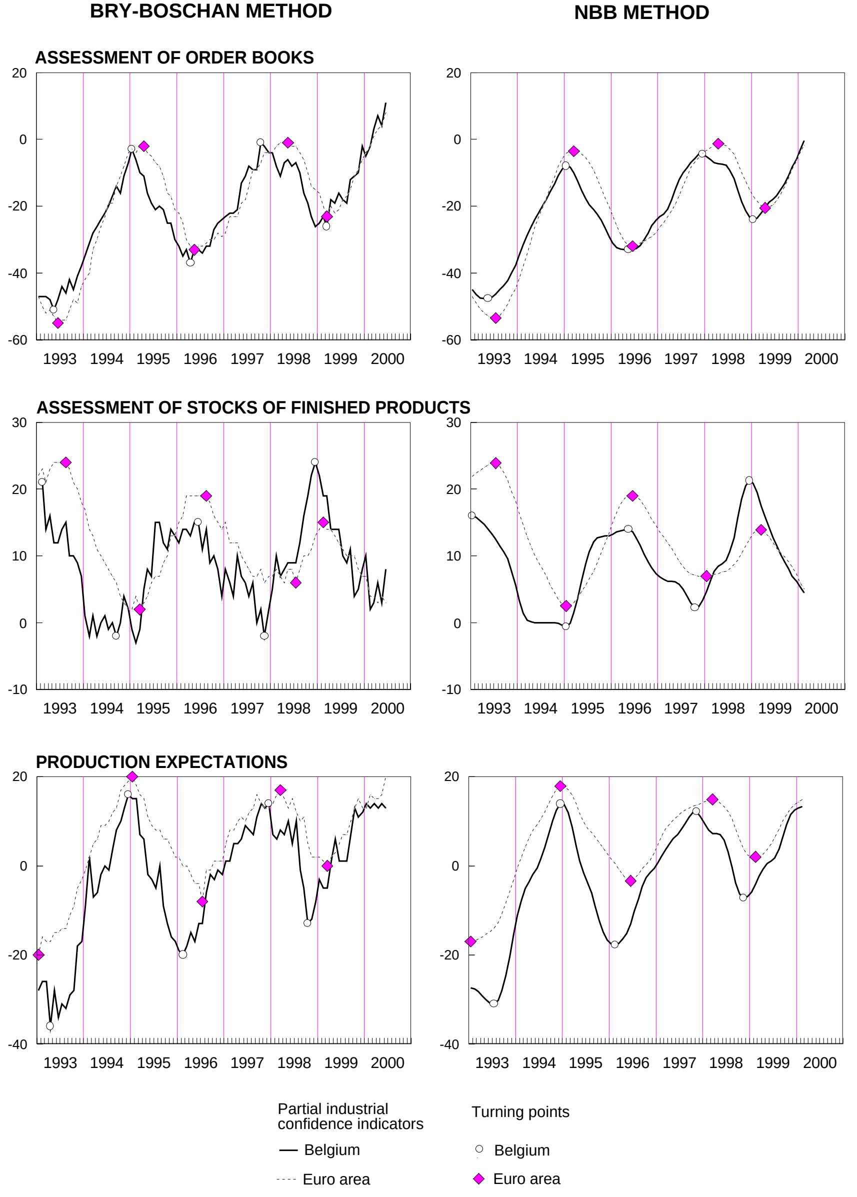


Source: EC; own calculations.

1) The chart shows the confidence level for different durations of lead, i.e. the probability that the null hypothesis of no lead can be rejected for the indicated number of months.

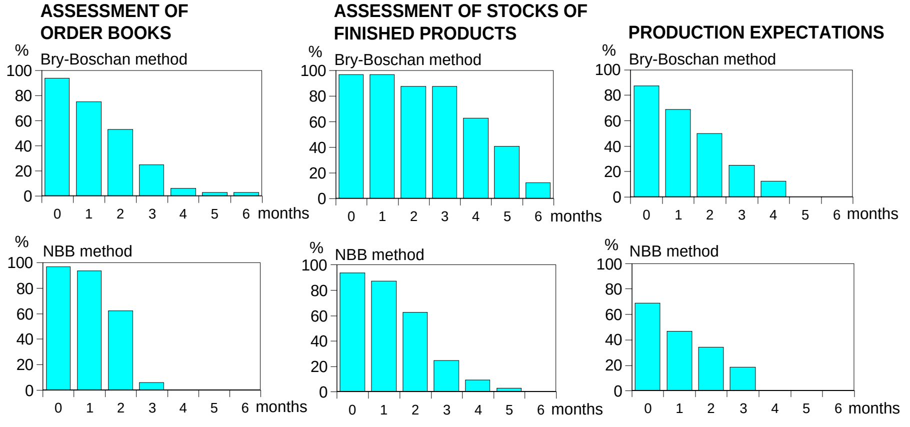
Confirming financial market sentiment, this first part has shown that since 1993 the turning points in the industrial confidence indicator for Belgium have significantly preceded those for the euro area with a difference usually larger than one month. The leading nature of the indicator for Belgium was far less pronounced, or even absent, before 1993. The reasons why the leading nature of the Belgian indicator was less pronounced before 1993 have not been investigated so far. Some structural changes may have occurred within the Belgian economy - for example in enterprises' stocking behaviour or in labour market flexibility, after the severe economic downturn of 1993 - or at euro area level, for instance the reunification of Germany.

Chart 6. Industrial confidence indicators in Belgium and in the euro area: Breakdown by question (Period 1993-2000)



Source: EC; own calculations.

Chart 7. Industrial confidence indicators - Breakdown by question: Lead profile of Belgium in relation to the euro area (Period 1993-2000)



Source: EC, own calculations.

As the industrial confidence indicator is composed of three questions, i.e. the assessment of order books, the assessment of stocks of finished products and expectations concerning production, the leading nature of the Belgian business indicator might actually be attributable to a sharply leading nature in one or two of the three questions considered, while the other question(s) have (has) no leading properties.

According to the data, however, and especially during the most recent period, Belgium seems, for all of the three questions, to lead in relation to the corresponding questions for the euro area as a whole, although the rejection of the null hypothesis of no leading nature is somewhat less pronounced for the question on production expectations.

It should be stressed that the leading nature of the Belgian indicators in relation to the euro area holds good only for timing differences in the turning points. Other methods employed in business cycle research, such as correlation analysis, structural VAR, causality tests or unobserved component methods consider the entire time series. It is important to bear in mind the difference in starting point when comparing the results obtained by one method or another. Preliminary research conducted at the NBB, using the Granger causality test and the SVAR approach, does not find any evidence that the Belgian industrial confidence indicator helps to explain the euro area indicator. However, these findings are not necessarily incompatible. Even if the Belgian indicator does not help to explain the euro area indicator, it is still interesting to know that turning points in Belgium usually precede turning points at euro area level.

2 Comparison of the Business Cycles

The leading nature of the industrial confidence indicator for Belgium in relation to the same indicator for the euro area might simply be due to the fact that the turning points in the business cycle movements of activity in Belgium are also ahead of those in the euro area.

In the absence of an official, widely agreed dating of the business cycles for Belgium[11] and for the euro area, three variables have been used as reference for the business cycle, namely real GDP, the index of industrial production and the degree of utilisation of production capacity in the manufacturing industry. The choice of the reference variables was restricted by the limited availability of sufficiently long detailed statistics for the euro area. In particular, the value added of manufacturing industry would have been a candidate if it had been available.

Applying the growth cycle definition[12], the turning points have been identified in the deviations from the trend for the different reference series. The trends have been calculated using the Hodrick-Prescott filter, with standard value for λ.

The whole period 1985-2000 has been considered, as well as the sub-period 1993-2000.

2.1 Turning points in the deviation from the GDP trend

The movements of the deviation from GDP trend are quite similar in Belgium and in the euro area. In both cases, nine turning points are identified by the Bry-Boschan procedure[13] over the period from 1985 to the first quarter of 2000, delimiting four complete cycles. However, both in Belgium and in the euro area, the cycle with a peak in the first quarter of 1990 and a trough in the first quarter of 1991 should be regarded as minor: the amplitude of the movement during this cycle is far smaller than during the other cycles, and the actual value of GDP remained above the trend value throughout the cycle, even in the trough.

Looking at the difference between Belgium and the euro area in the timing of the turning points, the same sub-periods as before can be distinguished: during the period from the first quarter of 1985 to the fourth quarter of 1992, the turning points coincided exactly in both series, but since the first quarter of 1993, the turning points in the Belgian series of deviation from trend GDP systematically lead those in the euro area series.

---

11 For alternative dating of the business cycle in Belgium, see Lebrun (1999), OECD (1998) or Bodart and Candelon (1999).

12 Niemira and Klein (1995), p. 6.

13 The NBB procedure is not suitable for quarterly data.

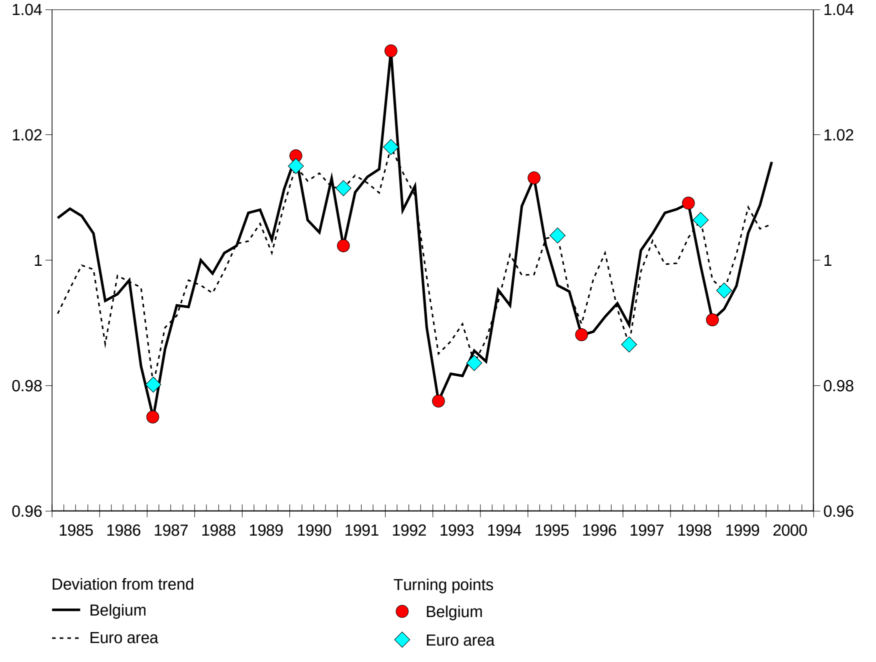
Table 3. GDP deviation from trend in Belgium and in the euro area: Turning points

Turning points	Belgium	Euro area	Lead (+) / Lag (-) [1)]
Peak 1			
Trough 1	1987 Q1	1987 Q1	0
Peak 2	*1990 Q1*	*1990 Q1*	*0*
Trough 2	*1991 Q1*	*1991 Q1*	*0*
Peak 3	1992 Q1	1992 Q1	0
Trough 3	1993 Q1	1993 Q4	3
Peak 4	1995 Q1	1995 Q3	2
Trough 4	1996 Q1	1997 Q1	4
Peak 5	1998 Q2	1998 Q3	1
Trough 5	1998 Q4	1999 Q1	1

Sources: ECB, Eurostat; own calculations.

1) Lead (+) or lag (-) of the Belgian turning points, in quarters.

Chart 8. GDP deviation from trend and turning points in Belgium and in the euro area

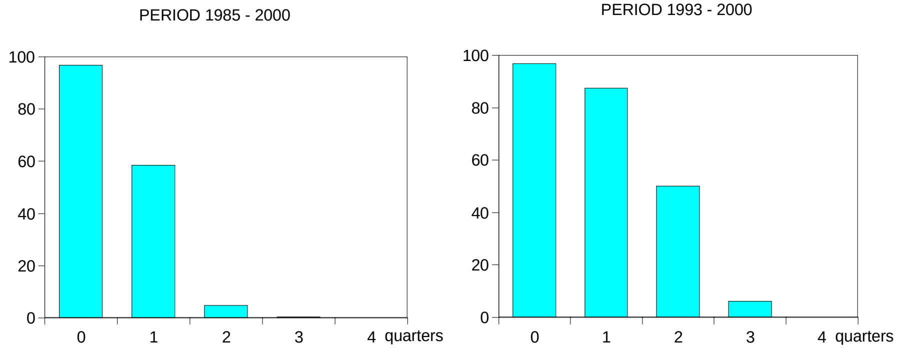


Sources: ECB, Eurostat; own calculations.

As a matter of fact, the period of lead between Belgium and the euro area is longer for the deviation from trend GDP than for the industrial confidence indicator. From 1993 onwards Belgium's lead ranges from 1 quarter to 4 quarters for GDP, against a lead of -1 to 5 months for the confidence indicator.

The null hypothesis of no timing difference can be rejected at a confidence level of 97% both for the period 1985-2000 and for the period 1993-2000.

Chart 9. GDP deviation from trend: Lead profile of Belgium in relation to the euro area



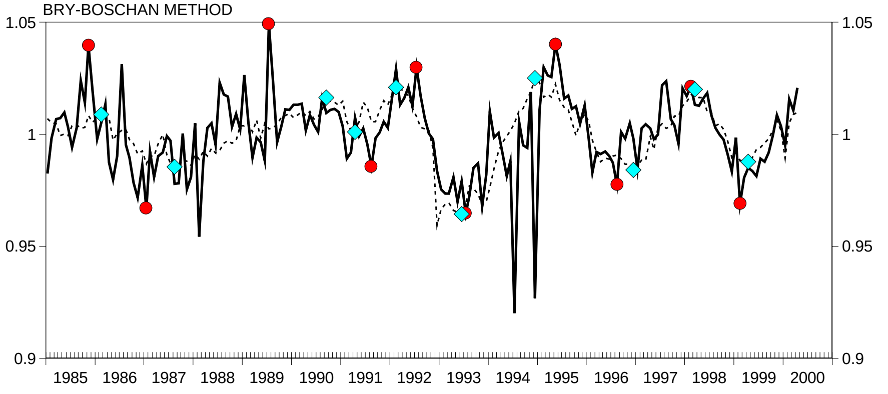
Sources: ECB, Eurostat; own calculations.

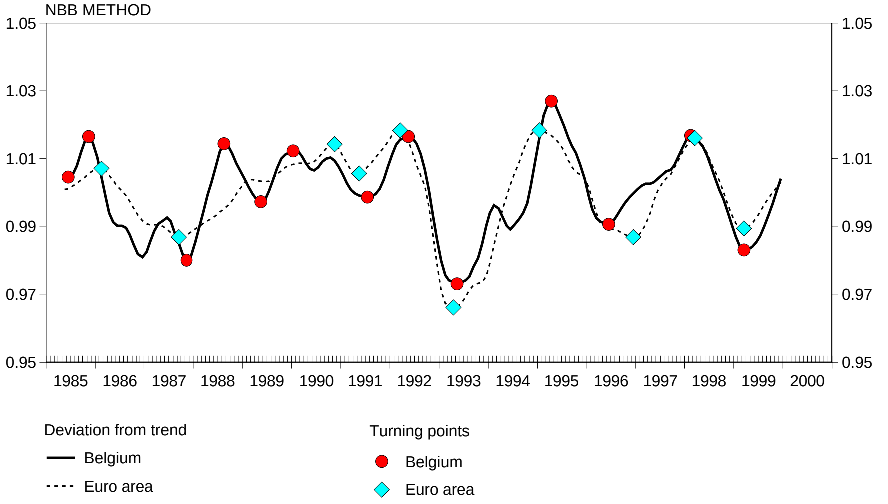
2.2 Industrial production

For different reasons, the industrial production index can be helpful for assessing the business cycle situation of Belgium and the euro area. First, its coverage of branches corresponds to the coverage of the industrial confidence indicator. Second, it is a monthly measure of the volume of activity. Moreover, indices of this kind are used for reference by the OECD.

Unfortunately, the deviation from trend of the industrial production indices exhibits a high degree of volatility, at least in the case of Belgium. Even after application of the NBB smoothing technique, the movements of the series and the dating of the turning points differ appreciably between Belgium and the euro area during the period 1985 to 1992. Since 1993 the business cycles have been more similar, but without any systematic leading nature for either of the series in terms of the timing of the turning points.

Chart 10. Industrial production: Deviation from trend and turning points in Belgium and in the euro area





Source: OECD; own calculations.

Table 4. Industrial production: Deviation from trend in Belgium and in the euro area: Turning points

	Bry-Boschan method			NBB method		
Turning points	Belgium	Euro area	Lead (+) / Lag (-) [1)]	Belgium	Euro area	Lead (+) / Lag (-) [1)]
Peak 1	Nov. 1985	Feb. 1986	3	Nov. 1985	Feb. 1986	3
Trough 1	Jan. 1987	Aug. 1987	7	Nov. 1987	Sept. 1987	-2
				Aug. 1988 [2)]		
				May 1989 [2)]		
Peak 2	July 1989	Sept. 1990	14	Jan. 1990	Nov. 1990	10
Trough 2	Aug. 1991	Apr. 1991	-4	July 1991	May 1991	-2
Peak 3	July 1992	Feb. 1992	-5	May 1992	Mar. 1992	-2
Trough 3	July 1993	June 1993	-1	May 1993	Apr. 1993	-1
Peak 4	May 1995	Dec. 1994	-5	Apr. 1995	Jan. 1995	-3
Trough 4	Aug. 1996	Dec. 1996	4	June 1996	Dec. 1996	6
Peak 5	Feb. 1998	Mar. 1998	1	Feb. 1998	Mar. 1998	1
Trough 5	Feb. 1999	Apr. 1999	1	Mar. 1999	Mar. 1999	0

Source: OECD; own calculations.

1) Lead (+) or lag (-) of the Belgian turning points, in months.

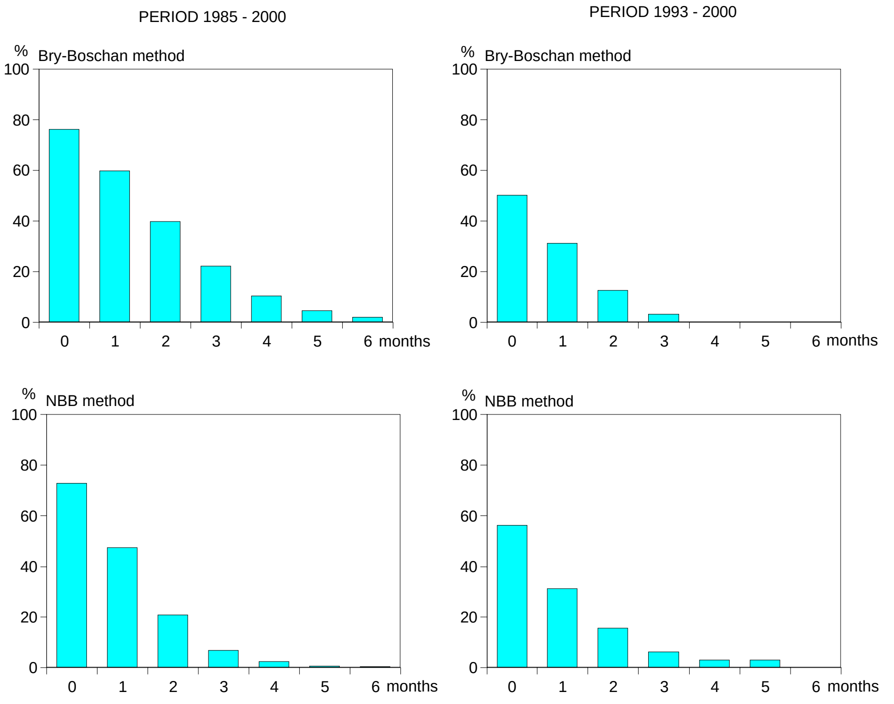
2) Turning points not included in the comparison.

Table 5. Test for the leading nature of the deviation from trend of Belgian industrial production

	Period 1985-2000	Period 1993-2000
Bry-Boschan method	76.2%	50.0%
NBB method	72.9%	56.3%

Source: OECD; own calculations.

Chart 11. Industrial production: Deviation from trend: Lead profile of Belgium in relation to the euro area



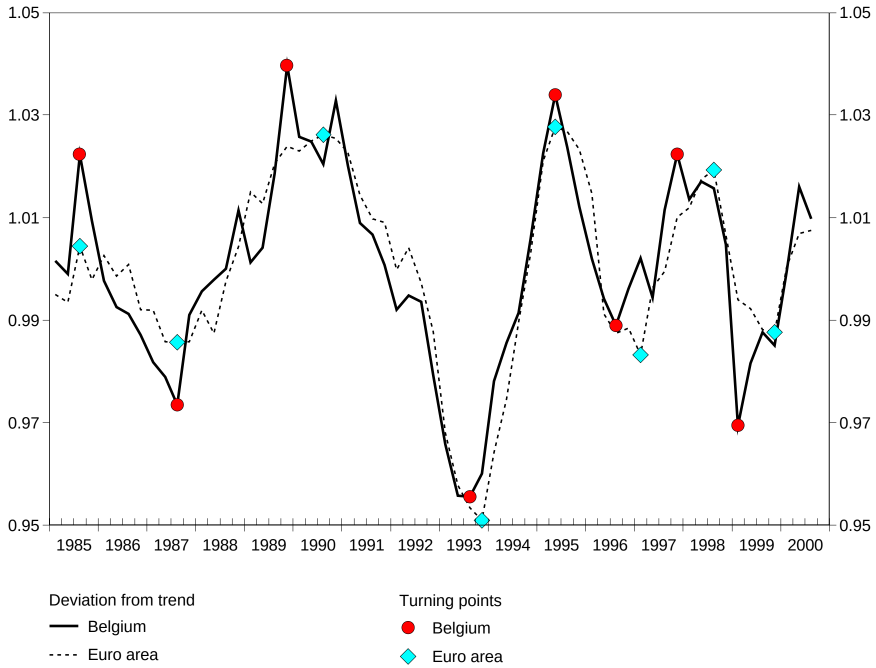
Source: OECD; own calculations.

2.3 Degree of utilisation of production capacity in manufacturing industry

Within the framework of the harmonised EU programme of business surveys, heads of enterprises in manufacturing industry are asked to report every three months the level of their capacity utilisation, expressed as a percentage of total capacity. Even if it is usually agreed that the movements in the degree of utilisation mainly reflect short-term demand fluctuations and that the level should be stable across the business cycle, it seems that some structural factors tend to raise it in Belgium.[14] Therefore, the degree of utilisation of production capacity series has been detrended.

14 See Vanhaelen and Dresse (2000).

Chart 12. Deviation from trend in degree of utilisation of production capacity and turning points in Belgium and in the euro area



Source: EC; own calculations.

Table 6. Deviation from trend in degree of utilisation of production capacity: Turning points

Turning points	Belgium	Euro area	Lead (+) / Lag (-) [1)]
Peak 1	1985 Q3	1985 Q3	0
Trough 1	1987 Q3	1987 Q3	0
Peak 2	1989 Q4	1990 Q3	3
Trough 2	1993 Q3	1993 Q4	1
Peak 3	1995 Q2	1995 Q2	0
Trough 3	1996 Q3	1997 Q1	2
Peak 4	1997 Q4	1998 Q3	3
Trough 4	1999 Q1	1999 Q4	3

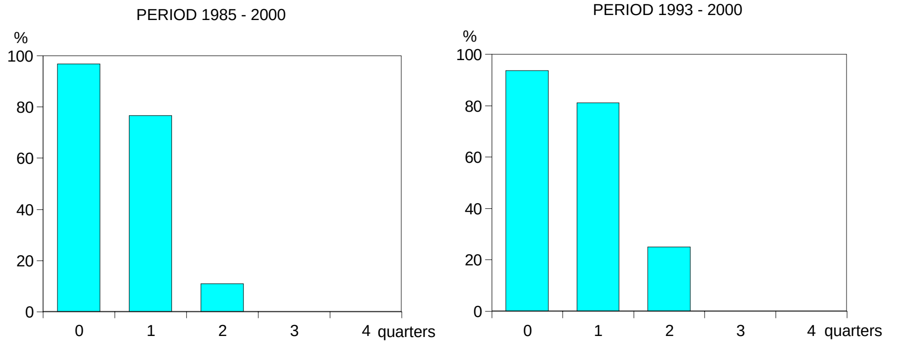
Sources: EC, own calculations.

1) Lead (+) or lag (-) of the Belgian turning points, in quarters.

Over the period 1985 to mid-2000, turning points in the degree of utilisation of production capacity in Belgium coincided with those of the euro area on three occasions and were leading in five cases, the maximum difference being three quarters.

Unlike in the case of the industrial confidence indicators or the deviation from GDP trend, the lead profile for the degree of utilisation is comparable for the period 1985 to 2000 and for the period since 1993. In both cases the confidence level for rejection of the hypothesis of no lead is above 90%, and around 80% for a difference of one quarter.

Chart 13. Deviation from trend in degree of utilisation of production capacity: Lead profile of Belgium in relation to the euro area



Source: EC; own calculations.

3 A More Detailed Analysis

Structural features might explain the leading nature of the Belgian economy in relation to the euro area. In this connection we have tried to investigate some possible explanations. The Belgian economy is characterised by a relative specialisation in intermediate goods, by its openness and by the importance of small and medium-sized enterprises.

In the absence of harmonised and internationally comparable statistics providing reliable information and data on structural features of the economies, an alternative way of looking for an explanation of the leading nature of the Belgian business cycle is to investigate the business survey data in greater detail.

In this section we concentrate on the period 1993-2000, as this is the period during which the Belgian industrial confidence indicator led that of the euro area.

3.1 Breakdown by sector

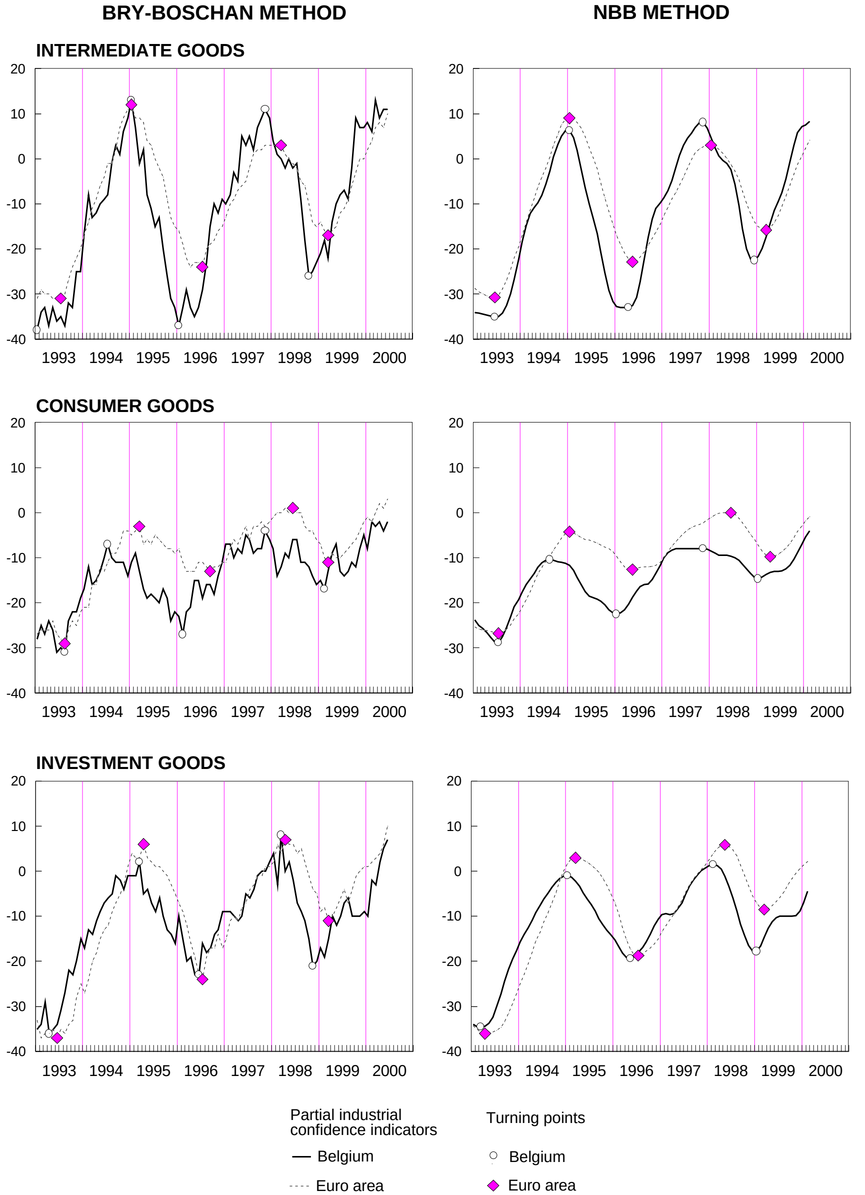
An explanation often put forward is that the product structure of the Belgian economy, i.e. a presumed relative specialisation in intermediate goods, explains at least part of its leading nature. As intermediate goods, being an input for other production activities, constitute the goods category which is affected first by impending upswings or downturns, a relative specialisation in this kind of product could possibly explain a tendency to lead. For this assumption to be the main explaining factor, it would be necessary for the turning points of the Belgian business indicators at more detailed levels not to be significantly ahead of the turning points of the corresponding indicators for the euro area as a whole, which would mean that the leading nature at the aggregated level would be due to different weights of the different sectors in Belgium and in the euro area.

The investigation to determine whether Belgium is also in the lead at more detailed levels of production was carried out for intermediate, consumer and investment goods.[15]

It appears from the lead profile chart that in most cases the null hypothesis of no leading nature of the turning points in Belgium can be rejected at a confidence level above 90%. Only in the case of the turning points for the intermediate goods, identified by using the NBB method, the confidence level for rejection of the no lead hypothesis is below 90%, and is around 50% for a lead by one month. All in all, this suggests that the Belgian product structure is not sufficient to explain the leading nature of Belgian economic activity.

[15] Data are available to enable the same to be done at more detailed levels, but the different curves become much more volatile, and the risk of arriving at the wrong conclusions becomes very great.

Chart 14. Industrial confidence indicators in Belgium and in the euro area: Breakdown by sectors (Period 1993-2000)



Source: EC; own calculations.

Chart 15. Industrial confidence indicators - Breakdown by sectors: Lead profile of Belgium in relation to the euro area (Period 1993-2000)

Source: EC; own calculations.

3.2 Breakdown into export-oriented and other companies

As Belgium's economy is characterised by a high degree of openness, the leading nature of Belgian activity could be due to its being more sensitive to economic developments abroad. In order to test this hypothesis, in the absence of euro area data, the Belgian manufacturing sector was broken down into two sub-sectors: export-oriented companies, defined as companies which achieve at least 50% of their turnover abroad, and other companies. If this hypothesis were correct, one could expect that the turning points of the business indicator for the former group of companies would significantly lead those of the indicator for other companies.

It is apparent from the charts that this was not the case. The assumption that the leading nature of Belgian activity is attributable to the economy's relative degree of openness is not supported by the data available in Belgium. However, it could be that the respective behaviour of export oriented companies and other companies exhibits more differences in the euro area than in Belgium.

Chart 16. Industrial confidence indicators in Belgium: Breakdown into export-oriented and other companies (Period 1993-2000)

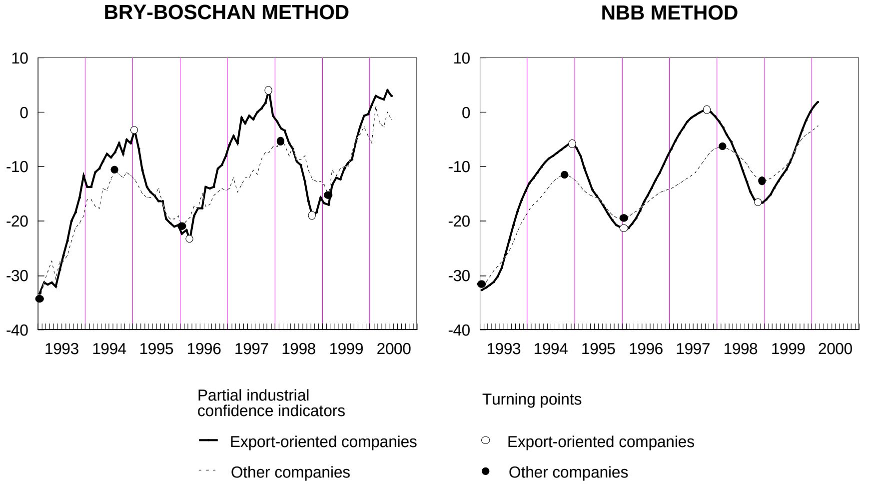
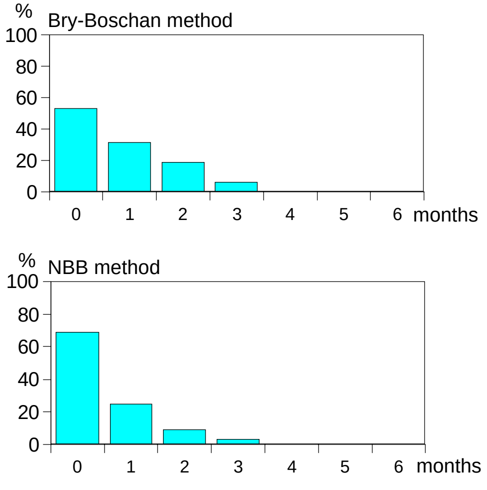


Chart 17. Industrial confidence indicators - Breakdown into export-oriented and other companies: Lead profile of export oriented companies in relation to other companies (Period 1993-2000)

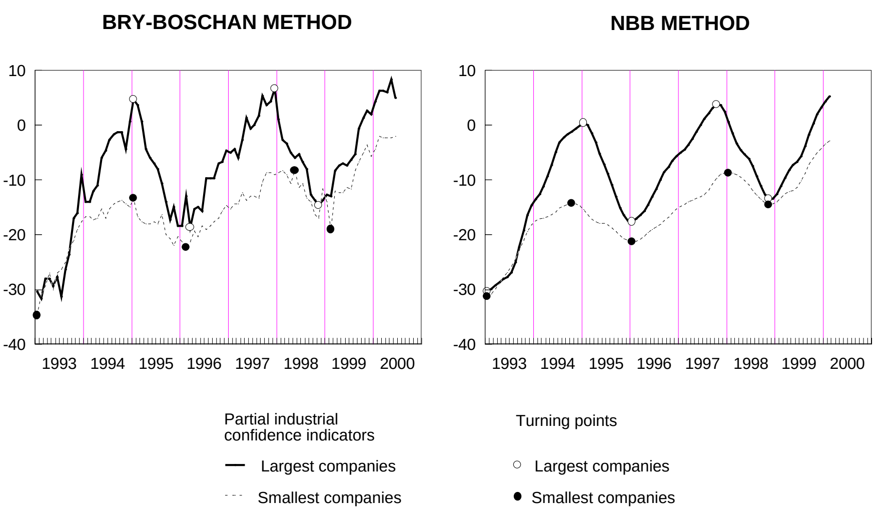


Source Chart 16 and 17: NBB.

3.3 Breakdown into large and small companies

A third possible explanation is based on the fact that production in Belgium takes place to a large extent in small and medium-sized enterprises. Although no figures are available concerning the importance of this sector in other countries, a different composition as regards the size of companies could explain a leading position occupied by Belgium. The reasoning is that the managers of smaller companies are more closely involved with real production activities in the enterprise, and that they therefore indicate changes in activity earlier in their replies to the questions of the business surveys. If this is so, the turning points of the business indicator for small companies should lead those of the large companies. For this reason, all companies participating in the Belgian business surveys in the manufacturing sector were subdivided into four size groups, based on turnover. In order to test the hypothesis, the largest and smallest groups were compared.

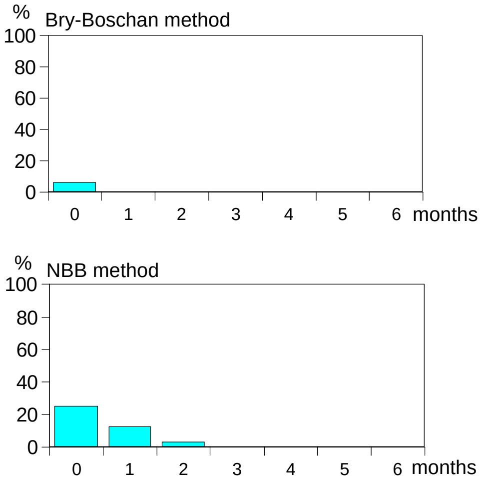
Chart 18. Industrial confidence indicators in Belgium: Breakdown into large and small companies (Period 1993-2000)



Source: NBB.

According to the data, there is no evidence of a significant lead in the timing of the turning points of small companies in relation to large ones. Thus, possible explanations referring to the relative importance of small and medium-sized enterprises are not confirmed by the data.

Chart 19. Industrial confidence indicators - Breakdown into large and small companies: Lead profile of the smallest companies in relation to the largest companies (Period 1993-2000)



Source: NBB.

4 Conclusions

The international press has recently reported on the widely-held view in the financial markets that the movement of the Belgian industrial confidence indicator might precede the euro area business cycles. The initial purpose of this paper was to assess whether this market perception is more than a simple optical illusion, resulting from the inspection of graphical representations of the data. In the first part, the timing of turning points in the industrial confidence indicators for Belgium and for the euro area has been identified using two alternative methods: the Bry-Boschan method and the NBB method, which is based on a combined median moving average smoothing of the series. Applying the Randomisation Test and the lead profile of the significance of the differences in timing, proposed by Banerji, we conclude that the turning points in Belgium do in fact significantly lead turning points in the euro area from 1993 onwards.

The leading nature of the Belgian industrial confidence indicator is not really surprising, as changes in the business cycle stages in Belgium seem to have been ahead of changes in the euro area during the period from 1985 to the first quarter of 2000. Among the three different reference series used to compare the business cycle movements in Belgium and in the euro area, the null hypothesis that turning points in Belgium do not lead those in the euro area is rejected at a confidence level above 90 p.c. in the case of GDP and of the degree of utilisation of production capacity in manufacturing industry. The leading nature is more pronounced for the sub-period beginning with the first quarter of 1993, especially in the case of GDP. However, the comparison of the movements of the industrial production indices does not confirm these conclusions.

At this point it would have been interesting to look for the causes of this leading nature for Belgium using value added statistics by sectors or comparing the components of the expenditure account in Belgium and the euro area. The degree of availability of sufficiently long time series for the euro area has hampered this work. However, using partial industrial confidence indicators, three factors (specialisation in intermediate goods, openness and high representation of small and medium-sized enterprises) that might explain why the Belgian business indicator and Belgian activity seem to lead their euro area counterparts were investigated, but could not be validated by the data.

As it seems to be impossible to identify one or more sectors or groups of enterprises accounting for the leading nature of Belgian economic activity when looking at turning points, at least when using the business survey data, it looks as if this leading nature is a kind of general feature of the Belgian economy.

However, further research could be carried out along different lines. Other methods for identifying turning points could be applied to the data in order to reject or validate the results obtained so far. To overcome the lack of historical data for the euro area, Belgium could be compared to the larger euro area countries. This might contribute to finding the cause of the leading nature of the Belgian economy within the components of the expenditure account (do changes in inventories or investments by enterprises or exports play a specific role in the Belgian business cycle?). The fact that the leading nature of the Belgian indicator and of the Belgian business cycle was weak or non-existent before 1993 also needs to be explained.

References

Banerji, A. (2000), "The lead profile and other non-parametric tools to evaluate survey series as leading indicators", in Use of Survey Data for Industry, Research and Economic Policy, selected papers presented at the 24th CIRET Conference, Wellington, New Zealand 1999, Oppenländer, Poser and Schips (ed) Ashgate.

Bry, G. and Boschan, Ch. (1971), Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, NBER Technical Paper 20.

Bodart V. and Candelon B. (1999), "L'indicateur IRES, un nouvel indicateur synthétique de la conjoncture en Belgique", in Note du Service d'analyse économique, IRES.

European Commission (1997), "The joint harmonised EU programme of business and consumer surveys", in European Economy, 1997-6.

Lebrun, I. (1999), "Le système d'indicateurs avancés du BfP - Un nouvel outil pour l'analyse conjoncturelle", Working paper 99-2, Bureau fédéral du Plan.

National Bank of Belgium (1990), "Révision de la courbe synthétique de conjoncture", Bulletin de la Banque Nationale de Belgique, Août-Septembre 1990, pp. 50-64.

Niemira M. and Klein Ph. (1995), Forecasting Financial and Economic Cycles, John Wiley and Sons, New York.

OECD (1998), "Revision of leading indicator for G7 and Belgium"

Vanhaelen, J.-J. and Dresse L. (2000), "The degree of utilisation of production capacity in Belgian manufacturing industry", in Use of Survey Data for Industry, Research and Economic Policy, selected papers presented at the 24th CIRET Conference, Wellington, New Zealand 1999, Oppenländer, Poser and Schips (ed) Ashgate.

Chapter III

Innovation and Technology

Explaining Innovative Activity in Service Industries: Micro Data Evidence for Switzerland

Spyros Arvanitis

1 Introduction

In earlier work we analyzed empirically the innovative behavior of Swiss manufacturing firms building on the wide consent in economic literature that demand prospects, type and intensity of competition, market structure, factors governing the production of knowledge (appropriability, technological opportunities), innovation and production costs as well as firm size (as a variable controlling for further unspecified influences) are the main determinants of a firm's innovative activity (see Arvanitis/-Hollenstein 1994, 1996). In this paper[1] we applied the same model to analyze innovative activities in the service sector. Several innovation variables referring to the input as well as the output side of the innovation process (seven indicators covering both product and process innovation) served as endogenous variables of the innovation model. Mostly due to lack of appropriate data at the firm level, innovative activity in the service sector remains still a rather underexplored area of research. In view of the increasing significance of this growth-leading sector more insights in the mechanism of the innovation process in this sector are needed in order to better understand the specific character of its growth process.[2]

In his study we used firm data from nine service industries (including also growth sectors like banking, insurance, software and other business services) collected by means of the Swiss Innovation Survey 1999 which was based on a questionnaire practically identical with the one used in the 2nd Innovation Survey of the European Community. Our sample considered firms with more than 5 employees. The final data set contained usable data for 595 service firms. Owing to the nature of our dependent variables (ordinal measures of innovation) we mostly used a Probit model for the econometric estimations.

The set-up of the paper is as follows: section 2 sketches briefly the underlying theoretical background and gives the necessary information with respect to the specification of the innovation equations respectively used in the empirical part of this

---

1 This research was supported by the Swiss Federal Ministry of Economic Affairs.

2 For recent studies analyzing, some of them descriptively, the innovative activity at the firm level based on survey data for the service sector see Gellatly, 1999, Gellatly and Peters, 1999 for *Canada*, Licht et al., 1997 and Janz and Licht, 1999 for *Germany*, Brower and Kleinknecht, 1996, 1997 for *Holland* and Sirilli and Evangelista, 1998 for *Italy*. For some earlier exploratory work for *Switzerland* see Arvanitis et al., 1998 and Donzé and Lenz, 1999.

study. In section 3 we present some descriptive information on various types of innovation indicators by industry and firm size class. Section 4 gives some information on data and method. In section we present the empirical results. Finally, in section 6 we draw some conclusions.

2 Framework of Analysis

Theoretical Background

The theoretical setting builds on a simple static deterministic model of a firm optimizing its innovation output separately for new products and new processes under monopolistic competition in a market of N identical firms; the model comprises the most important determinants of innovative activity as seen in the literature.[3] The formal exposition contains constant-elasticity functions for production costs and product demand complemented by a linear innovation cost function and a linear knowledge production function with intra- and extramural knowledge as inputs; in this context the extent of the appropriability of knowledge - or viewed the other way round - the existence of know-how spillovers can be taken into account. The innovation output is conceptualized as cost-reducing in the case of new processes or demand-creating for new products. On the basis of the first-order conditions of the usual optimization calculus with respect to production output q and intramural knowledge x and after performing the necessary comparative statics, we obtain equation (1) which is interpreted as a causal relationship between innovation output k and the factors on the right side:[4]

$$k = k(\underset{+}{\sigma},\ \underset{+/-}{\varepsilon},\ \underset{+}{1/N},\ \underset{+}{\lambda},\ \underset{-/+}{\beta},\ \underset{+}{\alpha},\ \underset{-}{\varphi},\ \underset{-}{c'}) \qquad (1)$$

with

σ : shift parameter of a firm's demand schedule representing pure income effects;

ε : price elasticity of a firm's demand;

1/N : inverse number of N (identical) firms in the market (market concentration);

λ : degree of appropriability of a firm's own knowledge;

β : shift parameter of the knowledge production function (cost and quality shifts);

---

3 See e.g. Dasgupta (1986), Dosi (1988), Cohen and Levin (1989) and Cohen (1995) for reviews of theoretical and empirical literature in general. There are relatively few studies dealing theoretically with the specific features of the innovative activity in the service sector, thus innovation in the service sector does not (yet) build a distinctive branch of theoretical reasoning (see e.g. Barras, 1986, Quinn, 1987, Sunbo, 1997 (strong oriented to the management literature), Galloux and Weinstein, 1997, Gallouj, 1999 and Tether et al., 2001).

4 For a detailed description of the model see Arvanitis (1999), ch. 2. A preliminary version of this model was used in an earlier investigation of the innovative behaviour of manufacturing firms (see Arvanitis and Hollenstein 1994).

α : elasticity of costs (demand) with respect to innovation output;

φ : unit costs of innovation;

c‘ : unit costs of production (relevant only for product innovation).

The signs of the determinants reported in (1) seem to be economically plausible, so we refrain here from a more detailed discussion; wherever two signs are given, the first one refers to process, the second one to product innovations. The most important feature of innovative activity which the model tries to capture is the interdependency of the generation processes of innovative knowledge among firms. According to the approach adopted here, at least two dimensions of such spillover effects have to be taken in consideration: first, the degree of absorption of extramural knowledge of a given firm (which is equal to one minus the degree of appropriability λ of a firm’s own, i.e. intramural knowledge) and, second, the amount of knowledge which is available or is anticipated to be available to the firm. The shift parameter β measures the influence of the *amount* of available knowledge on innovation output, whereas the elasticity α reflects the *productivity* of used knowledge. The model is derived under the assumption of identical firms, therefore the influence of firm size is not considered explicitly. The number of firms on the market N is fixed in the short-term (no entry) and the effect of N on the short-term equilibrium (thus on innovative output k) is derived as a result of an exercise of comparative statics with respect to the equilibrium parameter N.

Empirical Model

Table 1 gives information on the measures of innovation output used as dependent variables in this study. These are, first, the two binary variables INNOPD ("introduction of *product* innovations yes/no") and INNOPC ("introduction of process innovations yes/no"), second, two ordinal variables assessing the *technological* importance of introduced product and process innovations (intensity measures ININTPD and ININTPC resp.) and, third, a variable measuring the sales share of innovative products (SP). Alternatively to these variables, we also analyzed qualitative input-oriented innovation measures (RD: "R&D activities yes/no"; DEVINT: intensity of development expenses).

Table 1. Specification of the Innovation Variables

Variable	Definition	Measurement Scale
RD	R&D activities yes/no	nominal (1, 0)
DEVINT	Development expenditures	ordinal; measured on a five point Likert scale
INNOPD	Product innovations yes/no	nominal (1, 0)
INNOPC	Process innovations yes/no	nominal (1, 0)
ININTPD	Assessment of the importance of the introduced product innovations from a technical point of view	ordinal; measured on a five point Likert scale
ININTPC	Assessment of the importance of the introduced process innovations from a technical point of view	ordinal; measured on a five point Likert scale
SP	Sales share of "highly improved" products or "entirely new" products	metric (%)

Table 2 summarizes the relevant information with respect to the independent variables of the empirical model. This contains proxies for six out of eight theoretical variables in equation (1). We could not find in our data proxies for φ (unit innovation costs) and c' (unit production costs) in (1), so we were obliged to omit these variables in the specification of the empirical model; we assume that part of their impact can be captured by the control variables inserted in the estimation equations (firm size and industry dummies). We also included in the empirical model a proxy for *financing conditions* not explicitly considered in the theoretical model and a measure of the *follow-up investment expenses related to the innovation* (expenses for equipment needed for introducing new products and production techniques, acquisition of external knowledge, personnel training related to the introduced innovations and marketing for new products) which account for most of total innovation costs in service sector. These two variables capture the effects of the magnitude and type of financial resources dedicated to innovative activities. Finally, we also used some additional variables referring to *information technology aspects* of the innovations, which is a particular feature of innovative activities in the service sector.

We used assessments of firm-specific demand expectations (D) as a measure of (expected) demand shifts. As proxy for the price elasticity of demand we used assessments of the intensity of the price competition on a firm's (global) product market (IPC); in addition, we also included in our empirical equation a measure of the intensity of non-price competition (INPC). Market concentration CONC is represented by four dummy variables related to the firm-specific number of principal competitors on the (world) market; this specification of CONC seems to be more appropriate for an open economy than traditional measures such as the concentration ratio or the Herfindahl index referring exclusively to the home market.

The relevance of the appropriability of the benefits of introduced innovations as an incentive for a firm's innovative activity (dummy variable: "relevance of appropriability for innovative activity yes/no") is a proxy for the theoretical variable degree of appropriability λ.

Table 2. Specification of the Explanatory Variables of the Innovation Equation

Variable	Description / Economic Interpretation	Sign
1. Demand		
D	Medium-term expected change of demand 2000-02	+
2. Market Conditions		
IPC	Intensity of price competition in the product market; (negative (positive) sign expected for product (process innovation)	-/+
INPC	Intensity of non-price competition in the product market	+
CONC	Concentration measure based on the number of competitors in the product market (four dummies: 16-50 competitors, 11-15 competitors, 6-10 competitors, less than 5 competitors; reference group: more than 50 competitors)	+
3. Appropriability		
APPR	Relevance of property rights protection for a firm's innovative activities (dummy variable)	+
4. Technological Opportunities		
TPOT	Technological potential, i.e. scientific, technological and organizational knowledge relevant to the firm's innovative activity	+
External knowledge sources: (Factor values of five factors extracted through a principal component factor analysis of ordinal data (measured on a five-point Likert scale) referring to the importance of fourteen external knowledge sources)		
KS1	Universities, technical schools, research laboratories, patent disclosures	+
KS2	Software suppliers, consultancy, computer-based information networks	+
KS3	Suppliers of materials, components, equipment	?
KS4	Users of a firm's products, firms of the same conglomerate, competitors	(+)
KS5	Fairs, exhibitions, professional associations and journals	?
5. Other variables		
FIN	Extent to which introduced innovations have been financed through a firm's internal resources (cash flow, reserves, etc.)	+
INTF	Follow-up investment related to product and process innovations (equipment, acquisition of external knowledge, training, marketing)	+
INFORM	Extent to which introduced innovations are applications of information technology	+
INFOEXP	Expenditures for information technology	+
Firm Size:		
L	Number of employees	?
L	Square of number of employees	?
(8 industry dummies; reference group: personal services)		

Note: Unless otherwise specified, the variables reflect assessments of the surveyed firms measured on a five-point Likert scale for the period 1997-1999.

The technological opportunities (represented through the parameters α and β in equation (1)) are proxied by two (sets) of variables: The first one (TPOT) reflects the general technological potential characterizing the fields of activity relevant to the firm and leading to substantial quality and cost shifts. The second variable measures more specifically the contribution of external knowledge to the firm’s own innovative activity (elasticity of innovation output with respect to external knowledge). This specific impact is proxied by the factor values of five factors extracted from information on a set of 14 single external knowledge sources (assessments of the importance of these sources for a firm’s own innovative activity) by means of a principal component factor analysis. The first factor (KS1) refers to science-related external knowledge from universities, private and public research institutions as well as information from patent disclosures, the second one (KS2) to knowledge on information technology from software suppliers, computer-based networks and consultancy, the third one (KS3) to know-how stemming from suppliers of materials, components and equipment, the fourth one (KS4) to knowledge from users of products, firms of the same firm conglomerate and competitors and the fifth one (KS5) to information from fairs, exhibitions as well as professional associations and professional journals.

As a measure of total follow-up investment expenses we used the ordinal variable FOLINT ("intensity of total follow-up costs"); the qualitative variable FIN ("extent to which introduced innovations have been financed through internal funds (cash flow, reserves, etc.)") is considered to cover the effects of financing conditions.

Firm size was proxied by the number of employees; we included a linear and a quadratic term in order to investigate the type of quantitative relation between innovation variables and firm size.

The variables INFORM ("extent to which introduced innovations are applications of information technology"; this output-oriented variable was used in the estimates for ININPD, ININTPC and SP) and INFOEXP (qualitative measure for the expenditures for information technology; this input-oriented variable was used in the estimate with DEVINT) were added to the model to capture the specific influence of information technology on the innovative activity.

Most variables shown in Table 2 reflect assessments of the surveyed firms on a five-point Likert scale, which were introduced into the model either directly in their original form or indirectly as the outcome of a factor analysis; these assessments are assumed to be measurements on an interval scale. The sign expectations as given in the last column are derived partly from the theoretical model (for the variables D, IPC, INPC, CONC, APPR, TPOT), partly from the theoretical and empirical literature (for the variables FIN, INTF, INFORM, INFOEXP). The effects of each specific knowledge source and particularly of each single variable used in the model (variables KS1 to KS5) is not a priori given. However, there is a restriction stemming from the theoretical model that the overall effect of all five variables should be non-negative. For some variables (KS1, KS2 and KS1 for product innovations) we postulate a sign expectation based on evidence from earlier empirical work (for example Cohen and Levinthal, 1989, 1994 for the USA; Arvanitis and Hollenstein, 1994 for Switzerland). The impact of firm size is also not a priori given.

In sum, we used following three (slightly) different model specifications depending on the type of innovation indicator considered as the dependent variable:

$$INNOPD, INNOPC, RD = f(D, IPC, INPC, CONC, APPR, TPOT, KS1, KS2, KS3, KS4, KS5, L, L2, \text{industry dummies}) \quad (2)$$

$$ININTPD, ININTPC, SP = f(D, IPC, INPC, CONC, APPR, TPOT, KS1, KS2, KS3, KS4, KS5, FIN, FOLINT, INFORM, L, L2, \text{industry dummies}) \quad (3)$$

$$DEVINT = f(D, IPC, INPC, CONC, APPR, TPOT, KS1, KS2, KS3, KS4, KS5, FIN, INFOEXP, L, L2, \text{industry dummies}) \quad (4)$$

Equation (2) can be interpreted as containing the determinants of the basic firm decision "to innovate or not" (variables INNOPD; INNOPC) or "conduct R&D or not" (in the case of the variable RD). In the equations (3) and (4) are then found the explanatory factors for the follow-up firm decision "to choose a certain level of innovation (R&D) activity" (ININTPD, ININTPC with respect to innovation output; DEVINT with respect to innovation input). Finally, the variable SP, which is actually more related to the *outcomes* of the innovation process with respect to the marketplace, is also explained by equation (3).

3 Innovative Activities in Swiss Service Industries: Some Descriptive Information

The level of innovation activity varies considerably across service industries. Table 3 gives a picture of the range of innovative activities in nine industries of the Swiss service sector for the period 1997-1999. The frequency of product *versus* process innovations is almost the same across industries with the notable exceptions of hotels and catering (as a "traditional" industry) and computer and research services (as a "modern" industry) both of them having a significantly higher propensity to product than to process innovations (columns 1 and 2 in the upper part of Table 3). Intensity indicators referring only to innovating firms tend to reduce the differences among the industries (columns 3 and 4; once more is the industry of computer services an exception with a considerably above-average share of firms with high intensity of product innovation). Input-oriented activities demonstrate clearly the dominant position with respect to innovative activities of the three "modern" industries computer and research services, business services, banking and insurance (columns 5 and 6). These industries are responsible for practically all of R&D expenditures of the service sector which amounted to about 22% of R&D expenditures of the Swiss business sector in 2000 (see SFSO, 2001).[5] Finally, we find the highest sales

[5] The R&D share of services in Switzerland is higher than that in Germany (5.1%), France (12.3%), Italy (16.2%), Japan (5.5%) and the United Kingdom (19.6%), but lower than the corresponding share for Canada (36.7%) and the United States (28.8%); the figures for the G-7 countries are cited in Jankowski, 2001 and refer to the period 1996-1998.

shares of innovative products in computer and research services as well as in business services (column 7; in this case banking and insurance show only an average performance).

Table 3. Innovative Activities by Industry and Firm Size 1997-1999

	Innovations		Innovation Intensity		R&D activities	Intensity development activities	Sales share new products
	product	process	product	process			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				percentage			
Wholesale trade	36.2	32.4	50.7	58.3	19.6	23.7	17.5
Retail trade	33.3	30.3	38.3	30.3	11.5	8.3	11.1
Hotels, catering	51.2	40.5	36.8	44.8	15.7	8.9	15.8
Transport, communications	25.6	29.3	41.9	36.4	15.2	10.4	11.3
Finance, insurance	71.7	63.6	41.8	55.2	41.8	16.0	11.5
Real estate, leasing	14.3	35.7	0.0	0.0	7.1	0.0	4.0
Computer services, R&D	61.1	41.7	81.8	35.7	47.2	56.0	31.8
Business services	38.1	46.5	59.3	64.3	36.8	24.7	18.1
Personal services	35.0	35.0	42.9	66.7	15.0	0.0	6.6
Firm Size (number of employees)							
6-19	31.4	28.8	47.1	52.6	14.9	10.7	17.6
20-49	35.9	33.6	50.0	53.7	18.1	14.3	16.0
50-99	44.5	37.3	46.7	38.9	28.4	18.0	13.7
100-199	37.9	43.7	50.0	36.6	28.2	10.5	16.0
200-499	66.7	56.9	50.0	56.4	39.4	18.2	14.6
500-999	69.4	77.8	45.5	65.4	55.6	14.8	12.9
over 1000	65.5	79.3	52.6	55.0	44.8	31.8	11.2

Note: (1): firms with product innovations; percentage of all firms; (2): firms with process innovations; percentage of all firms; (3): firms with high intensity of technically important product innovations (intensity higher than 3 on a Likert scale of 1 to 5); percentage of innovating firms; (4): firms with high intensity of technically important process innovations (intensity higher than 3 on a Likert scale of 1 to 5); percentage of innovating firms); (5): firms with R&D activities; percentage of all firms; (6) firms with high expenditures for the development of new products and new processes (intensity higher than 3 on a Likert scale of 1 to 5); percentage of innovating firms; (7): sales share of "new" and "highly improved products" (%).

Table 3 also demonstrates that there is a tendency for a higher frequency of product as well as process innovation in larger firms (columns 1 and 2 in the lower part of Table 3). This size effect almost disappears in the case of the intensity indicators (only innovating firms; columns 3 and 4) and also for the sales share of new products (column 7). Nevertheless, size differences seem to be of considerable importance for R&D activities (columns 5 and 6). The missing firm size effects with respect to the output-oriented variables and to the input-oriented variable DEVINT are at odds with the strong size dependence of innovative activity usually observed in manufacturing (see Arvanitis, 1997).

4 Data and Method

The data used in this study came from the Swiss Innovation Survey 1999 which in its core questions was quite comparable with the "Community Innovation Survey" (CIS II) conducted in most European countries between 1996 and 1997. The survey was based on a (with respect to firm size) disproportionately stratified random sample (28 industries, 9 of them belonging to the service sector, and, within each industry, three industry-specific firm size classes with fill coverage of the upper class of large firms). The firms were asked to fill in a questionnaire on several aspects of innovative activity and economic performance during the period 1997-1999.[6]

Table 4. Composition of the Data Set

	N	%
Industry		
Wholesale trade	140	23.6
Retail trade	90	15.1
Hotel, catering	39	6.6
Transport, communications	85	14.3
Finance, insurance	77	12.9
Real estate, leasing	9	1.5
Computer services, R&D	31	5.2
Business services	115	19.3
Personal services	9	1.5
Firm Size (number of employees)		
6-19	185	31.2
20-49	136	23.0
50-99	84	14.1
100-199	81	13.6
200-499	57	9.6
500-999	27	4.5
over 1000	24	4.0
Total	595	

The present analysis is confined to the subsample of firms in service industries (2731 firms; nine industries). We received valid answers from 880 firms, i.e. 32.2% of the firms in the underlying sample. The response rates do not vary much across industries and size classes with a few exceptions (overrepresentation of business services and large firms in general, underrepresentation of hotels, catering). The non-response analysis (based on a follow-up survey of a sample of the non-respondents) did not indicate any serious

[6] German, French and Italian version of the questionnaire are available at request.

selectivity bias with respect to the structure of the original sample. In the study we used two final data sets, one containing data for both innovating and non-innovating firms (N=595) and another with only data for innovating firms (N= 303). Table 4 shows the structure of the used data set by industry and firm size class.

For binary independent variables (INNOPD, INNOPC, RD) a standard Probit model was applied as estimation method. For polychotomous ordered variables (ININTPD, ININTPC, DEVINT) we used an ordered Probit model. A Tobit estimation procedure was the appropriate method in the case of the censored variable SP.

5 Empirical Results

Input-oriented Innovation Variables

Table 5 (columns 1 and 2) contains the model estimates for the input-oriented variables RD and DEVINT. According to the results reported in this table the probability to conduct R&D in the service sector is positively correlated to supply-sided explanatory factors such as the degree of appropriability of the gains of the innovative activity (APPR) and the technological potential (TPOT; i.e. the amount of available knowledge which can be utilized in new products and processes); in particular, know-how stemming from users and competitors (KS4) seems to be important in this case. For firms operating in highly concentrated market environments (up to 10 competitors) is significantly more probable to conduct R&D than for firms in less concentrated markets; we could not find any effects of the market structure at other concentration levels. Demand expectations (D) and competition intensity (IPC for price competition, INPC for non-price competition) do not exert any significant influence on the probability of conducting R&D.

There are two possible explanations for the result with respect to demand expectations: first, for firms which intend to conduct *permanently* R&D demand expectations are relevant presumably only for the determination of the level of the R&D activities, but not for the basic decision to conduct R&D. From earlier innovation surveys we know the firms conducting R&D *occasionally* are reluctant to report such activities, so we can reasonably assume that firms reporting R&D activities belong to the first category which opt for *permanent* R&D activities; such a decision we consider to be independent of demand level. Secondly, the demand expectations referring to the years 2000-2002, which are generally expected to be a boom period, may look almost equally favorable for most firms, thus generating too little variance in the data to show up in the estimates. The data do not allow to discriminate between these two explanations.

There is also a discernible effect of firm size on R&D activity: the coefficient of the linear term is positive and the coefficient of the quadratic one negative which means that there is an inverted U-shaped relation between the firm size and the probability to conduct R&D. This result can be interpreted as a hint that there are no positive scale effects with respect to R&D.

The probability to choose a certain level of R&D activity, particularly development activity (variable DEVINT for development expenses) depends partially on the same factors as the probability to conduct R&D, namely appropriability conditions, technological potential and, to some extent, market concentration (6-10 competitors). Competition conditions and, at first glance rather astonishingly, demand expectations seem to exert no influence on the level of development expenses; in this case only the second of the two possible explanations mentioned above, quite similar expectations across industries, looks plausible enough to interpret this result.

There are also some important differences compared to the results for RD: science-related knowledge from universities, technical colleges, etc. (KS4) is now the relevant external knowledge source which looks quite plausible. Contrary to the estimates for RD, we found no effect of firm size on development expenses. We also do not find any influence of financing conditions (FIN), a rather unexpected result (but see next paragraph on the output-oriented indicators). There was no significant correlation of the variable DEVINT to the expenses for information technology (INFOEXP), which we interpret as a hint that, for most firms in our sample not belonging to the computer services industry, R&D activities are not closely related to information technology, even if the resulting innovations need much information technology for their implementation, which takes the form of follow-up investments (see also next paragraph on the output-oriented indicators).

Output-oriented Innovation Variables

There are separate estimates for product and process innovations for both types of variables (probability of introducing new products and new processes respectively: INNOPD, INNOPC (Table 5; columns 3 and 4) and probability of choosing a certain level of innovative activity for products and processes respectively: ININTPD, ININTPC (Table 5; columns 5 and 6)). There are some differences between the estimates for product and process innovations for the first type of variables (INNOPD, INNOPC). The probability to introduce product innovations correlated also in this case positively with supply-sided factors such as the appropriability conditions, the technological potential and especially knowledge from suppliers of materials, components and equipment. There was a linear relation to firm size and a non-linear relation to market concentration (positive coefficients for high levels (less than 5 competitors, 6 to 10 competitors) as well as for rather low concentration levels (16-50 competitors)). As in the estimates for input-oriented indicators we could not find any influence of demand expectations and competition conditions (IPC, INPC) on the probability to introduce product innovations. The estimates for process innovations showed a rather weak positive effect of demand expectations and no impact at all for the variables related to the market conditions including the dummies for market concentration (IPC, INPC, CONC). Appropriability was also not important, whereas the technological potential remained a considerable explanatory factor also in the case of process innovations. We found positive effects for two categories of external knowledge, know-how related to information technology stemming from software suppliers, computer-based networks, etc. (KS2) and, rather unexpectedly, generally accessible information from fairs, exhibitions, professional associations and journals (KS5); for one type of external knowledge (KS3: information from suppliers) we obtained, contrary to our result for product innovations, a negative effect. Firm size showed a parabolic relation to the probability to introduce process innovations similar to that found for R&D activities.

Table 5. Probit, Ordered Probit and Tobit Estimates with Input- and Output-oriented Innovation Measures

Explanatory Variable	RD	DEVINT	INNOPD	INNOPC	ININTPD	ININTPC	SP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
D	0.005	-0.065	0.017	0.173*	0.029	-0.057	-0.469
	(0.072)	(0.080)	(0.067)	(0.068)	(0.085)	(0.096)	(1.270)
IPC	-0.011	-0.054	0.039	0.008	-0.154**	-0.028	-2.118**
	(0.053)	(0.063)	(0.050)	(0.049)	(0.074)	(0.070)	(1.017)
INPC	-0.006	-0.005	0.060	0.037	0.004	-0.092	2.102*
	(0.060)	(0.070)	(0.056)	(0.055)	(0.079)	(0.083)	(1.168)
CONC							
16-50 firms	0.297	0.246	0.522**	0.063	-0.275	-0.206	-5.147
	(0.191)	(0.219)	(0.183)	(0.180)	(0.239)	(0.242)	(3.701)
11-15 firms	-0.005	0.146	0.044	-0.232	-0.184	-0.335	-0.618
	(0.257)	(0.329)	(0.238)	(0.240)	(0.362)	(0.374)	(5.547)
6-10 firms	0.287*	0.439**	0.282*	0.091	0.107	0.135	0.082
	(0.172)	(0.201)	(0.164)	(0.163)	(0.231)	(0.226)	(3.359)
< 5 firms	0.331**	0.085	0.364**	0.102	0.086	0.228	-3.403
	(0.166)	(0.199)	(0.158)	(0.154)	(0.222)	(0.219)	(3.247)
APPR	0.475**	0.499**	0.396**	-0.033	0.389**	-0.115	2.592
	(0.141)	(0.156)	(0.141)	(0.140)	(0.176)	(0.190)	(2.752)
TPOT	0.220**	0.115*	0.192**	0.138**	0.074	0.074	0.892
	(0.057)	(0.066)	(0.054)	(0.053)	(0.073)	(0.075)	(1.074)
KS1	0.044	0.189**	-0.048	-0.103	0.099	-0.141	1.380
	(0.061)	(0.069)	(0.059)	(0.069)	(0.076)	(0.095)	(1.205)
KS2	0.026	-0.119	-0.107	0.188**	0.023	0.018	-0.751
	(0.064)	(0.078)	(0.072)	(0.062)	(0.087)	(0.095)	(1.295)
KS3	0.035	0.117	0.112*	-0.120**	0.043	0.134	-0.697
	(0.063)	(0.073)	(0.062)	(0.061)	(0.082)	(0.087)	(1.231)
KS4	0.106*	0.028	0.068	0.057	-0.088	0.205**	2.003*
	(0.064)	(0.075)	(0.060)	(0.059)	(0.085)	(0.086)	(1.238)
KS5	0.018	-0.048	0.006	0.107*	-0.059	0.113	-2.216*
	(0.062)	(0.072)	(0.059)	(0.059)	(0.080)	(0.083)	(1.235)
INFOEXP		0.054					
		(0.057)					
FOLINT					-0.002	0.227**	0.449
					(0.075)	(0.078)	(1.078)
INFORM					0.210**	0.240**	0.156
					(0.061)	(0.068)	(0.910)
FIN		0.077			0.127**	-0.010	4.908**
		(0.053)			(0.060)	(0.060)	(2.052)

Note: Each column contains the estimated parameters and the standard errors in brackets; the statistical significance is indicated with ** and * representing the 5% and 10%-level respectively (Wald Chi2 test). Intercepts have been throughout omitted.

Table 5. continued

Explanatory Variable	RD	DEVINT	INNOPD	INNOPC	ININTPD	ININTPC	SP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
L	2.7E-4**	1.3E-4	3.1E-4**	8.0E-4**	-3.0E-5	4.0E-5	-2.2E-3
	(1.2E-4)	(0.9E-4)	(1.5E-4)	(2.1E-4)	(9.4E-5)	(9.7E-5)	(1.5E-3)
L^2	-7.2E-9**	-3.1E-9	-5.5E-9	-1.7E-8**	9.2E-10	-5.8E-10	6.1E-8
	(3.5E-9)	(2.5E-9)	(4.7E-9)	(0.6E-8)	(2.5E-9)	(2.5E-9)	(4.0E-8)
N	595	303	595	595	216	216	250
Left censored							22
McFadden R^2	0.139	0.110	0.151	0.131	0.081	0.121	
df	24	25	24	24	26	26	
LR statistic	100**	89**	124**	197**	50**	70**	66**
ESA test (df)		98**(75)			114**(78)	98*(78)	
%-concordant	75	72	75	73	70	71	

Note: Each column contains the estimated parameters and the standard errors in brackets; the statistical significance is indicated with ** and * representing the 5% and 10%-level respectively (Wald Chi2 test). Intercepts have been throughout omitted.
LR statistic: Likelihood Ratio Test (Chi2 test); ESA test: Score Test for Equal Slopes Assumption (Chi2 test); 8 industry dummies.

Some of the effects found for the estimates for the probability to innovate (variables INNOPD, INNOPC) became insignificant with respect to the probability to choose a certain level of innovative activity (variables ININTPD, ININTPC): market concentration showed no effect on product innovation and also the variable for the intensity of nonprice competition (INPC) had a statistically insignificant coefficient.

In accordance to the theoretical model we obtained a negative effect for the variable measuring the intensity of price competition (IPC; a similar effect was also found in the estimates for the variable SP; see below). In the estimates for the variables ININTPD and ININTPC the variables measuring technological opportunities (technological potential, specific knowledge sources with the exception of KS4 for process innovations) lost their explanatory power. Firm size also became statistically insignificant. On the other hand there was a discernible positive effect of the financing conditions (FIN) in the case of product innovations and, as expected, a strong positive influence of the variable INTF (measuring the follow-up investment expenses) on process innovation. Further, INFORM (measuring the extent to which introduced innovations are applications of information technology) was strong positively correlated to the probability to choose a certain activity level for both types of innovation (product innovations and process innovations). This result demonstrates clearly that most innovation activity in the service sector is closely related to applications of information technology.

For the outcome-oriented indicator SP we found a negative effect of the variable of price competition (IPC) and a positive one for the variable for non-price competition (INPC); both effects are in accordance with the theoretical predictions. Financing conditions seem to be also important for the sales share of new products. The influence of supply-sided factors such as appropriability and technological opportunities were only weakly correlated to SP (positive effect of users knowledge SP4, negative effect of know-how from general accessible information sources KS5). The variables for follow-up investments FOLINT and for information technology content of the introduced innovations (INFORM) were of no importance for this innovation variable. Further, the sales share of new products is not size-dependent. Rather unexpectedly, we could not find a significant demand effect. On the whole, it looks quite plausible that demand-sided factors such as market competition conditions together with financial resources exerted a stronger influence on this outcome-oriented variable than supply-sided factors such as appropriability and technological opportunities.

Relation between Input-oriented and Output-oriented Innovation Measures

In order to investigate the influence of various kinds of innovation *inputs* on innovation *output* we inserted seven variables measuring different types of innovation costs alternatively as regressors in the equations for the output-oriented innovation indicators ININTPD, ININTPC and SP. These additional estimates are reported in Table 6. To make the table more comprehensible, we omitted all other variables and kept only the coefficients for the seven cost variables: expenses for research, development and total follow-up investment as well as four components of follow-up costs: expenses for equipment, acquisition of external knowledge, personnel training and marketing of new products. R&D expenses have no significant impact on innovation output in the service sector on the whole (which is does not exclude that a positive effect does exist in more R&D-intensive industries such as computer and business services). Follow-up investment expenses seem quite relevant for innovations in the service sector, especially in the case of process innovations: all four cost components showed positive effects on the output-oriented variable ININTPC, the strongest effect stemming from human capital investment (training costs). In the case of product innovations only expenses for equipment (presumably for computers, etc.) were of relevance.

6 Conclusions

We obtained a pattern of explanation of the innovative activity which looked quite plausible across the different types of innovation measures used (input-oriented and output-oriented innovation variables); it was also consistent to that found earlier for manufacturing. In general, the theoretical model captured rather the characteristics of the basic decision to innovate rather than those of the decision to choose some level of innovative activity. Supply-sided factors such as appropriability and technological opportunities seem to be more important for the decision to introduce innovations than demand-sided variables like demand-perspectives and intensity of price and non-price competition. These results are similar to those found in earlier work for manufacturing. There is a stronger influence of market structure in the service sector than in manufacturing. However, we also find some differences from our previous results. For example, contrary to manufacturing sector firm size seems to be less important in explaining the intensity of innovative activity in the service sector. Follow-up costs are important for the level of innovative activity, especially for process innovations. The high information technology content is a particular feature of the innovative activity in the service sector.

Table 6. Relations Between Input-oriented and Output-oriented Innovation Measures (Model Estimates of ININTPD, INNINTPC and SP with *Alternative* Input-oriented Measures)

Input Measure	ININTPD	ININTPC	SP
	(1)	(2)	(3)
R&D Expenses:			
RESINT	-0.163	-0.004	2.165
	(0.102)	(0.114)	(1.587)
DEVINT	-0.019	0.112	0.787
	(0.069)	(0.070)	(1.035)
Follow-up Investment:			
FOLINT	-0.002	0.227**	0.449
	(0.075)	(0.078)	(1.078)
EQUIP	0.114*	0.137**	1.348
	(0.065)	(0.066)	(0.959)
KNOW	0.000	0.130*	-1.576
	(0.075)	(0.079)	(1.105)
TRAIN	0.107	0.171**	0.658
	(0.071)	(0.071)	(1.044)
MARKET	0.030	0.108*	1.860**
	(0.063)	(0.063)	(0.915)

Note: RESINT: research expenses; DEVINT: development expenses; FOLINT: total follow-up investment linked to product and process innovations; EQUIP: follow-up expenses for equipment; KNOW: follow-up expense for acquisition of external know-how; TRAIN: follow-up expenses for personnel training; MARKET: follow-up expenses for marketing of new products (all variables are measured on a five-point Likert scale). This table contains the coefficients of the input measures being used as regressors in estimates of the full model with the output-oriented measures ININTPD, ININTPC and SP (see Table 5); the standard errors are in brackets; the statistical significance is indicated with ** and * representing the 5% and 10%-level respectively.

References

Arvanitis, S., 1997, "The Impact of Firm Size on Innovative Activity - An Empirical Analysis Based on Swiss Firm Data", Small Business Economics, 9, 473-490.

Arvanitis, S., 1999, Generierung von neuem technischen Wissen, Produktivität und Arbeitsqualifikation in der schweizerischen Industrie. Eine Querschnittsanalyse auf der Basis von Unternehmens-daten, Dissertation Universität Zürich, Zurich.

Arvanitis, S., Donzé, L., Hollenstein, H. und S. Lenz, 1998, Innovationsaktivitäten in der Schweizer Wirtschaft, Teil II: Bauwirtschaft und Dienstleistungssektor, Berne: Bundesamt für Wirtschaft und Arbeit.

Arvanitis, S. and H. Hollenstein, 1994, "Demand and Supply Factors in Explaining the Innovative Activity of Swiss Manufacturing Firms", Economics of Innovation and New Technology 3, 15-30.

Arvanitis, S. and H. Hollenstein, 1996, "Industrial Innovation in Switzerland: A Model-Based Analysis with Survey Data", in A. Kleinknecht (ed.), Determinants of Innovation and Diffusion, London: Macmillan, pp. 13-62.

Barras, R., 1986, "Towards a Theory of Innovation in Services", Research Policy, 15, 161-173.

Brouwer, E. and A. Kleinknecht, 1996, "Determinants of Innovation: A Microeconometric Analysis of Three Alternative Innovation Output Measures", in A. Kleinknecht (ed.), Determinants of Innovation and Diffusion, London: Macmillan, pp. 99-124.

Brouwer, E. and A. Kleinknecht, 1997, "Measuring the Unmeasurable: A Country's Non-R&D Expenditure on Product and Service Innovation", Research Policy, 25, 1235-1242.

Cohen, W.M., 1995, "Empirical Studies of Innovative Activity", in P. Stoneman (ed.), Handbook of Innovation and Technological Change, Oxford: Blackwell, pp. 182-264.

Cohen, W.M. and R.C. Levin, 1989, "Empirical Studies of Innovation and Market Structure", in R. Schmalensee and R.D. Willig (eds.), Handbook of Industrial Organization, London: North-Holland, pp. 1059-1107.

Cohen, W.M. and D.A. Levinthal, 1989, "Innovation and Learning: The Two Faces of R&D", Economic Journal, 99, 569-596.

Cohen, W.M. and D.A. Levinthal, 1994, "Fortune Favours the Prepared Firm", Management Science, 40, 227-251.

Dasgupta, P., 1986, "The Theory of Technological Competition", in J.E. Stiglitz and G.F. Mathewson (eds.), New Developments in the Analysis of Market Structure, Cambridge, Mass.: MIT Press, , pp. 519-547.

Donzé, L. and S. Lenz, 1999, "Indicators and Determinants of Innovative Activity in the Service Sector: A First Empirical Analysis With Survey Data", in K.H. Oppenländer, G. Poser and S. Waller (eds.), Selected Papers Submitted to the 23rd CIRET Conference Held in Helsinki, July 30 - August 2, 1997, München: CIRET Studien, pp. 151-167.

Dosi, G., 1988, "Sources, Procedures, and Microeconomic Effects of Innovation", Journal of Economic Literature 26, 1120-1171.

Gallouj, F., 1999, "Les trajectoires de l'innovation dans les services: vers un enrichement des taxonomies évolutionnistes", Economies et Sociétés, 1, 143-169.

Gallouj, F. and O. Weinstein, 1997, "Innovation in Services", Research Policy, 26, 537-556.

Gellatly, G., 1999, "Differences in Innovator and Non-innovator Profiles: Small Establishments in Business Services", Working Paper No. 143, Micro-Economic Analysis Division, Ottawa: Statistics Canada.

Gellatly, G. and V. Peters, 1999, "Understanding the Innovation Process: Innovation in Dynamic Service Industries", Working Paper No. 127, Micro-Economic Analysis Division, Ottawa: Statistics Canada.

Griliches, Z., 1995, "R&D and Productivity: Econometric Results and Measurement Issues", in P. Stoneman (ed.), Handbook of Innovation and Technological Change, Oxford: Blackwell, pp. 52-89.

Janz, N. und G. Licht (Hrsg.), 1999, Innovationsaktivitäten in der deutschen Wirtschaft. Analyse des Mannheimer Innovationspanels im Verarbeitungsgewerbe und im Dienstleistungssektor, Schriftenreihe des ZEW, Band 41, Baden-Baden: Nomos Verlagsgesellschaft.

Licht, G., Hipp, C., Kukuk, M. und G. Münt, 1997, Innovationen im Dienstleistungssektor. Empirischer Befund und wirtschaftspolitische Konsequenzen, Schriftenreihe des ZEW, Band 24, Baden-baden: Nomos Verlagsgesellschaft.

Sirili, G. and R. Evangelista, 1998, "Technological Innovation in Services and Manufacturing: Results From Italian Surveys", Research Policy, 27, 881-899.

Tether, B.S., Hipp, C. and I. Miles, 2001, "Standardization and Particularization in Services: Evidence from Germany", Research Policy, 30, 1115-1138.

Innovation Modes in the Swiss Service Sector: A Cluster Analysis Based on Firm-Level Data

Heinz Hollenstein

Abstract

The aim of this paper is, firstly, to contribute to the understanding of innovation patterns in services. To this end, firms which are similar in terms of a large set of innovation indicators were grouped into clusters. For the Swiss case, it was possible to identify five clusters which exhibit clearly different configurations of a large number of innovation-related factors (appropriability, etc.) and several structural properties of firms (size, etc.). The clusters may thus be interpreted as specific "innovation modes". Secondly, we investigated whether these modes are "economically equivalent". In such a case, the unordered classifying of similar firms would be more appropriate than the ranking of industries according to their innovativeness. The evidence supports the classification approach quite well; however, the ranking procedure cannot be completely refuted. Finally, this paper yields some insights into the differences between the innovation patterns prevailing in services and in manufacturing.

1 Introduction

The innovation process is a complex phenomenon. It typically takes place over the course of several stages, ranging from basic research to market penetration by means of new products, and the introduction of new production techniques within firms. Therefore, an entire series of indicators is necessary, if we are to adequately describe and measure a firm's innovative activities. Each indicator stresses certain specific aspects of the innovative activities. Using the information provided by this system of indicators, we developed in earlier work an aggregate measure of innovation intensity and used it to rank firms and industries in the manufacturing (Hollenstein, 1996), as well as in the service sector (Arvanitis et al., 1998).

The ranking of industries implicitly assumes that these are (more or less) homogeneous entities with respect to the innovativeness of firms. This assumption is at variance with the evolutionary view of technological change (Gallouj, 1999). According to this approach, one would rather look for groups of firms characterised by similar patterns of innovation (irrespective of industry affiliation), which are conceptualised as unordered categories ("innovation modes"). The classificatory procedure reflects the heterogeneity of an industry with respect to the innovation strategies pursued by its firms. Moreover, it allows for the co-existence of different innovation modes which, at a certain point in time, are equivalent in terms of economic performance. Which of these innovation strategies is sustainable over the long run is decided during the course of a market-driven selection process, the outcome of which is not predictable (Metcalfe, 1995).

We explored the classificatory approach in a previous study of the Swiss manufacturing sector (Arvanitis and Hollenstein, 2001). The analysis yielded five innovation modes, which exhibited similarities to those identified by other authors (Pavitt, 1984; Cesaratto and Mangano, 1993; Arundel et al., 1995). The differences were explained to a large extent by several structural characteristics typical of the Swiss economy (e.g. limited scope for scale-intensive production).

In the present paper, we search for innovation patterns in the Swiss service sector, applying the same methodology used in our study devoted to manufacturing. The aim of the analysis is twofold: Our first goal is to contribute to a better understanding of innovative activities in this heterogeneous sector. Secondly, we explore the relative merits of the classification and the ranking approach as a means of analysing the innovation process. In this way, we hope to define the extent to which the two ways of looking at the matter can be applied. Our initial decision was to concentrate on services, although there are good arguments in favour of including manufacturing as well (see below). Proceeding this way, we are in a better position for comparing our results with other work.

Empirical research devoted to innovation in services is still quite scarce, despite the high and growing importance of this sector.[1] This unsatisfactory situation is attributable to conceptual problems ("what makes service innovations different?") as well as to a lack of (suitable) data. At the conceptual level, the literature, according to Evangelista and Sirilli (1995), identifies four main features that are specific to production and innovation in services: a) the close interaction between production and consumption (co-terminality), b) the increasing information content of services, c) the large and growing role played by human resources in service production, and d) the great importance of organisational change as a means of producing and delivering (new) services. This characterisation implies that non-technological innovations are an important feature of the service sector (Miles, 1995; Djellal and Gallouj, 1999). However, as a consequence of the tendency towards industrialisation in services and customisation in manufacturing, the distinctions between the two sectors have been blurred (see, for example, Coombs and Miles, 2000); it is perhaps co-terminality which is the most specific characteristic of production and innovation in services.

The second factor hampering the empirical investigation of innovative activities in services is data deficiency. This problem has been alleviated to a certain extent thanks to the second wave of the "Community Innovation Survey", carried out in EU member countries in 1996/97 (CIS II) and similar surveys conducted in other countries (Canada, Switzerland, etc.), which covered the service sector as well. However, some researchers (e.g. Djellal and Gallouj, 1999) are of the opinion that the questionnaires used in these surveys did not properly take into account the specific features of innovation in services

[1] It is very revealing that the "Handbook of the Economics of Innovation and Technological Change" edited by Stoneman (1995) does not treat the service sector at all.

(concentration on technological innovation, neglect of organisational innovation). Other researchers, such as Hughes and Wood (2000) or Evangelista and Sirilli (1995), maintain that this critique, though not without substance, is overrated. They refer, for example, to the growing importance of technology in service production, or to the specific definition of service innovations used in CIS II, which also includes new ways of delivering services. We shall address this problem again in Section 3.

In view of these difficulties, it is not surprising that there are only a few studies which examine innovation patterns in the service sector. The primary source of reference for most contributions is the well-known taxonomy developed by Pavitt (1984), which deals mainly with manufacturing. Pavitt distinguishes between four sectors: "Science-based industries", "Specialised producers", "Scale-intensive producers" and "Supplier dominated industries", with service industries included in the last category. In later work, Pavitt (and his co-authors) modified this taxonomy; a new category called "Information-intensive firms" was introduced, covering industries such as the financial sector or retailing, while other service industries (e.g. software) were identified as "Specialised producers" (Pavitt et al., 1989). Soete and Miozzo (1989) took Pavitt's original taxonomy as their starting point and added the category "Network-based industries", which covers two subgroups, namely "Scale-intensive industries based on physical networks" (transport, wholesaling) and "Industries relying on information networks" (finance, insurance, communications). These industries draw heavily on information technologies (IT). Evangelista (2000) recently developed a taxonomy generated by applying a formal statistical procedure (cluster analysis of industry data). He identified nine clusters which, in turn, were reduced into three main groups. The first and the second group are quite similar to Pavitt's "Supplier dominated industries" and "Science-based industries", while the third ("Interactive and IT based industries") reflects the importance of producer-user links and the widespread use of IT.

The taxonomies mentioned thus far are based on an industry level analysis; it is therefore (implicitly) assumed that innovation patterns at this level of aggregation are homogenous. Moreover, they are static in nature, since nothing is implied with respect to their evolution over time. In addition, they almost exclusively take technological innovations into consideration. Against this background, Gallouj (1999) developed an analytical framework based on a functional approach to service provision, which yields a multiplicity of service trajectories. The latter build on four specific technology-based methods of transforming basic competencies into (new) services (handling/transformation of tangibles, handling/treatment of codified information, interactive service production, method-based knowledge transformation), as well as on one mode of service provision, which does not rely on technology ("direct" transformation of competencies into services). At each point in time, innovation patterns can be characterised by the way new services are developed in terms of combinations of the five basic modes of service production. Over time, such trajectories evolve through changes in the proportions of the five constituent elements (ascribing different weights to the various elements, adding new ones, dropping old ones, etc.).

In principle, this evolutionary concept overcomes the limitations of the taxonomies mentioned above. It allows for different innovation modes within sectors and firms, it is dynamic in nature, and non-technological innovations are taken into account as well. However, in view of the multiplicity of innovation modes and the extensive data (highly differentiated panel data) required to identify them empirically, the detection of generalised regularities might prove difficult.

Taking the previous approaches to developing a taxonomy of service sector innovation as our point of reference, our method can be characterised as follows: a) it is based on an analysis of firm-level data; it is therefore not assumed that industries are homogenous in terms of innovation modes; in contrast, we explicitly investigate whether there is a correlation between innovation modes and industry affiliation (see Section 4); b) we use a formal statistical procedure in the identification of innovation modes (cluster analysis); of the taxonomies described above, only that developed by Evangelista (2000) is based on such an approach; c) at this stage, our analysis remains static (as is the case in most other taxonomies), although expanding our work towards a more dynamic view would be feasible (see Section 6); d) the database we use is better suited than the CIS data for taking into account the non-technological character of many innovations in services (see Section 3).

The rest of the paper is structured as follows: In the next section, we describe the data base and the methodology used in searching for innovation modes. The empirical results of our attempt to identify innovation modes are presented in Section 3. Next, we analyse the sectoral distribution of innovation modes and the variation of economic performance across these categories of firms; this enables us to assess the relative merits of the classification and ranking procedures. In Section 5, we investigate whether the differences between innovations in services and those in manufacturing, as identified in the literature, are confirmed by our findings. Finally, we discuss the main results and draw some conclusions.

2 Data and Procedure

2.1 Data

The data used in this study were collected as part of the Swiss Innovation Survey 1999, which was based on a stratified random sample (28 industries and three industry-specific classifications of firm size, with full coverage of large companies). The firms were asked to fill in a questionnaire (downloadable from www.kof.ethz.ch) on their innovative activities, which yielded a large number of innovation indicators (in this paper we use seventeen). In addition, the survey provided information on many variables, which can be used to describe the innovation process in some detail, and to explain the level and intensity of innovative activity. Most variables refer to the period 1997/99.

The analysis is confined to the subsample "services” (2731 firms; nine industries). We received valid answers from 880 firms, i.e. 32.2% of the underlying sample. The response rates do not much diverge across industries and firm sizes (with a few exceptions), as illustrated by Table A1 in the appendix. In view of the rather low (overall) response rate, it was necessary to conduct a survey among a sample of non-respondents using a few core questions related to innovative activity (response rate 90%). The non-response analysis did not indicate any signs of a serious selectivity bias with respect to the structure of the basic sample. By imputing missing values in the case of item non-response ("multiple imputation"; see Rubin, 1987 and Donzé, 2001), we were able to avoid a loss of observations that would have introduced a bias to the final data set.[2] As a whole, the data set may be considered representative of the underlying sample. For obvious reasons, the search for innovation modes is based on the subsample of innovative firms only, which amounts to 54% of the respondents (see Table A1).

2.2 Procedure

The present analysis searches for innovation modes using firm-level information. We assume that a firm pursues only one type of innovation strategy. Since our data refer to the firm's main activity, this assumption should not be too much of a simplification.

In the first step, a cluster analysis is performed in order to group firms into homogeneous categories with respect to seventeen indicators of innovation. These cover the input and the output sides of the generation of innovation, as well as the introduction of new products to the market or new processes in the firm (for details, see Section 3). Cluster analysis, however, was not directly applied to these variables. Instead, we started by synthesising the information contained in these measures by means of a factor analysis into a small number of variables ("factors"). The latter are uncorrelated (standardised) variables containing the information common to the original variables.[3] Then, we performed a (non-hierarchical) cluster analysis of these factors, in order to group the firms into a number of categories which are, with respect to the variables under investigation, as homogenous as possible (small within-cluster variance) and at the same time as different as possible (large between-cluster variance).[4]

In a second step, the analysis examined whether the clusters previously identified can really be interpreted as different modes of innovation. To this end, the clusters are characterised and interpreted in terms of a) the innovation indicators used in cluster analysis itself, b) a series of important determinants of innovation activity, such as innovation opportunities, appropriability, human resources (supply side), as well as market prospects and intensity of competition (demand side), c) a set of variables capturing the firms' position in knowledge networks (use of various external knowledge sources, formal

---

2 As shown by Donzé (2001), "multiple imputation" is a method that yields robust estimates for missing values. It is, for example, clearly superior to "simple imputation" which is used by EUROSTAT in producing the CIS data base.

3 The choice of the number of factors depends not only on statistical criteria, but also on the plausibility and interpretability of the resulting factor pattern in terms of innovative activity; see Manly (1986) for an introductory treatment of this method.

4 This procedure involves partitioning the sample, allowing observations to move in and out of groups at different stages of the analysis. At the beginning, more or less arbitrary group centres ("cluster seeds") were chosen and individual observations were allocated to the nearest one. An observation was later moved to another group, if it proved to be closer to that group's centre than to the centre of the initial group. This process, during which close groups were merged and distant ones split, was continued until stability was achieved with a predetermined number of clusters (see Manly, 1986).

co-operation), d) structural characteristics of firms (size, age, industry affiliation, export orientation), and e) two measures of performance. The distinction between innovation indicators (used in clustering) and determinants of innovative activities (used in characterising the clusters) is based on a microeconomic model of innovation behaviour, which was developed and empirically confirmed in earlier work on manufacturing (Arvanitis and Hollenstein, 1994, 1996), as well as on services (Arvanitis, 2000).[5]

As mentioned in the introductory section, a second aim of this paper is to evaluate the relative merits of ranking firms according to innovativeness and classifying them by modes of innovation. To this end, we investigate the relationship between innovation modes and industry affiliation. If the industrial composition of the clusters and that of the service sector as a whole are similar, industries are heterogeneous in terms of innovation patterns. In this case, classification is a more sensible procedure than sectoral ranking. The same holds true if economic performance does not significantly differ between clusters.

3 Empirical Results I

3.1 Identifying modes of innovation

The identification of innovation modes is based on the seventeen indicators of innovation listed in Table 1. On the input side, in addition to R&D expenditures, we take into account expenditures on IT (hardware and software) as well as the level and composition of innovation-related follow-up investments (the purchase of innovative machinery; the acquisition of external knowledge, such as licences, trademarks, etc.; human capital investments; and marketing outlays).

The output side of innovative activity is captured firstly by the firms' assessments of the technical and economic significance of the innovations. In addition, we include a variable which measures the IT content of innovation output; moreover, we also use two indicators which represent the innovation output of firms in terms of patent applications and licences granted. Finally, the sales share of innovative products and cost reductions induced by process innovations, both representing innovation-related improvements in a firm's market position, are used as market-oriented indicators of innovation. Most of these variables are qualitative, either binary (yes/no) or ordinal with five response levels ranging from "very low" (value 1) to "very high" (value 5). As has been shown in earlier econometric work (e.g. Arvanitis and Hollenstein, 1996), the information content of these subjective (assessment) measures is high.

As pointed out in Section 1, it is desirable that the indicators we use also capture some elements of non-technological innovations. To a certain extent, this is the case in the study at hand. The definition of "innovation" used in our questionnaire does not make any direct reference to technology. Moreover, several indicators (for example, the economic significance of innovations, the outlays for innovation-related training, expenditures on the

[5] See Cohen (1995) for a detailed survey of the empirical literature dealing with the explanation of innovative activities.

introduction of innovations to the market or the sales share of new or improved products) also capture non-technological aspects of innovation. As a consequence, we consider our database to be more appropriate for a comprehensive analysis of innovation in services than the information collected by the CIS II, in which innovation is explicitly defined as "technological innovation".

Table 1. Innovation Indicators Used in Cluster Analysis

Innovation Indicator	Measurement Scale	Value Range
1. Input-oriented measures		
Expenditures for		
- Research	ordinal	1, 5
- Development	ordinal	1, 5
- IT (hardware, software)	ordinal	1, 5
Follow-up investments		
- Total	ordinal	1, 5
- By type		
- Machinery and equipment	ordinal	1, 5
- Acquisition of external knowledge (licences, trademarks, etc.)	ordinal	1, 5
- Training	ordinal	1, 5
- Market introduction of innovations	ordinal	1, 5
2. Output-oriented measures		
Significance of the innovations in technical terms		
- Product	ordinal	1, 5
- Process	ordinal	1, 5
Significance of the innovations in economic terms		
- Product	ordinal	1, 5
- Process	ordinal	1, 5
IT-content of innovations	ordinal	1, 5
Patent applications (yes/no)	nominal	1, 0
Licences granted to other firms (yes/no)	nominal	1, 0
3. Market-oriented measures		
Sales share of new or highly improved services (%)	metric	0, 100
Cost reduction generated by process innovations (yes/no)	nominal	1, 0

The results of the preliminary step in the identification of innovation modes, i.e. of the factor analysis used to synthesise the information contained in the seventeen innovation indicators, are satisfactory (see Table 2). The five factors extracted account for 56% of the total variance. The first factor, which captures 20% of the total variance, reflects the various components (and their sum) of innovation-related follow-up investments. The second factor, which accounts for 11% of the variance, represents primarily the R&D input and the science-oriented innovation output (patent applications, licences granted). The third factor (10% of total variance) refers to the technological and IT dimension of service innovations. While the first three factors do not differentiate between product and process innovations, the last two factors do. Both factors stress the economic side of innovation, with the fourth one capturing product market-orientation and the fifth one reflecting cost reductions based on process innovations. We conclude that the factor pattern convincingly reflects the most important dimensions of the underlying innovation indicators.

Next we performed a non-hierarchical cluster analysis based on the scores of the factor analysis. Solutions with four, five or six clusters were of comparable quality according to the usual statistical criteria (approximate expected overall R2, cubic clustering criterion, etc.). In order to determine the final number of clusters, we took three criteria into account, namely a) the statistical properties in terms of the relationship between within-cluster and between-cluster variance, b) the plausibility of the clusters identified ("can the clusters convincingly be interpreted as innovation modes"?), and c) the number of firms per cluster. Based on the last criterion, the version with six clusters was dropped (one cluster contained very few observations). The result containing four clusters was inferior to that with five in terms of criteria a) and b). Therefore, we ultimately arrived at a five cluster solution, which is satisfactory in statistical terms; the approximate expected overall R2 of 0.45 suggests an acceptable fit of the data to the underlying cluster model. More importantly, the five clusters can convincingly be interpreted as innovation modes, as will be shown in the next section.

3.2 Basic characteristics of the innovation modes

Step 2 characterises the five innovation modes, firstly, in terms of the innovation indicators used in clustering (see Table 1 above). Secondly, we implement the variables listed in Table 3, which pertain to the demand and supply variables determining innovation intensity, the knowledge networks of firms, the structural characteristics of firms and selected measures of performance.

On the demand side, we take into account demand prospects, as well as the intensity of price and non-price competition in the relevant product markets. On the supply side, we include as a proxy for innovation opportunities, a variable which represents a firm’s assessment of its (overall) potential to generate novelties in (or around) its field of activity; moreover, we consider a measure of the appropriability of knowledge. A proxy for human capital is added to this group of variables, not only because firms that are well-endowed with highly skilled personnel are in a good position to absorb knowledge from other sources (Cohen and Levinthal, 1989), but also because this variable might prove to be particularly important in services (see Section 1).

Table 2. Factor Analysis of the Innovation Indicators Used in Cluster Analysis

Innovation indicator	Rotated Factor Pattern (equamax)				
Follow-up investments: total	.77	.09	.08	.09	.07
Follow-up investments: training	.75	-.03	.21	-.00	.05
Follow-up investments: machinery and equipment	.67	.01	-.02	.03	-.01
Follow-up investments: market introduction of innovations	.66	.13	-.01	.14	.17
Follow-up investments: acquisition of external knowledge	.54	.08	.30	-.09	-.20
Development expenditures	.19	.74	.08	.17	.01
Patent application	-.08	.73	-.02	.06	.07
Research expenditures	.14	.68	.10	.13	-.01
Granting of licences	-.02	.61	.14	-.09	-.02
IT-content of innovations	.06	-.07	.80	-.01	-.25
IT expenditures (hardware, software)	.46	.03	.68	.01	-.09
Significance of product innovations in technical terms	-.04	.15	.57	.18	.17
Significance of process innovations in technical terms	.03	-.01	.57	.13	.39
Sales share of new or highly improved products	-.02	.13	.01	.76	-.20
Significance of product innovations in economic terms	.07	-.06	.12	.73	.22
Cost reduction related to process innovations	.03	.06	-.08	-.11	.78
Significance of process innovations in economic terms	.07	-.05	.14	.47	.59
Number of observations					475
Kaiser's overall measure of sampling adequacy (MSA)					.720
Variance accounted for by the first five factors					.557
Root mean square off-diagonal residuals (RMSE)					.079
Variance accounted for by each factor	3.32	1.94	1.68	1.46	1.06
Final communality estimate (total)					9.46

There are numerous methods to rotate a factor loading matrix in order to facilitate the interpretation of the factors. Common to them is the attempt to simplify the factor matrix. The equamax criterion is a combination of varimax and quartimax which seek to simplify the columns and rows respectively of the (unrotated) matrix of factor loadings; see, for example, Ost (1984).

Table 3. Indicators Used to Characterise Innovation Modes

	Measurement Scale	Value Range
1. Innovative activities		
Innovation indicators as shown in Table 1	see Table 1	see Table 1
2. Determinants of innovative activity		
Demand side		
- medium-run demand prospects in the product market	ordinal	1, 5
- intensity of price competition in the product market	ordinal	1, 5
- intensity of non-price competition in the product market	ordinal	1, 5
Supply side		
- opportunities for innovation in the fields relevant to the firm's activities	ordinal	1, 5
- appropriability of knowledge	ordinal	1, 5
- employment share (%) of highly qualified labour (tertiary level)	metric	0, 100
3. Knowledge networks		
Use of fourteen types of external knowledge sources (see text)	ordinal	1, 5
Out-contracting of R&D		
- in Switzerland (yes/no)	nominal	1, 0
- abroad (yes/no)	nominal	1, 0
R&D co-operation:		
- number of domestic partners (3 and more vs. 0-2 partners)	nominal	1, 0
- number of foreign partners (3 and more vs. 0-2 partners)	nominal	1, 0
4. Structural characteristics of the firm		
Industry affiliation: share of firms (%) in 9 industries (see appendix, Table A1)	metric	0, 100
Firm size: share of firms (%) by 5 size classes (5-19, 20-49, 50-199, 200-499, 500 and more employees)	metric	0, 100
Share of firms (%) by start-up year: 3 classes (up to1988, 1989/94, 1995/99)	metric	0, 100
Share of firms (%) by export to sales ratio: 3 classes (up to 1, 2-19, 20+)	metric	0, 100
5. Economic performance		
Nominal value added per employee in 1998 (1000 SFr.)	metric	> 0
Share of firms (%) with increasing nominal sales in the period 1996/98	metric	0, 100

The ordinally scaled variables reflect the firms' assessments on a five-point Likert scale; the response levels range from "very low" (value 1) to "very high" (value 5).

Under the heading "position of the firm in knowledge networks", we take into account the intensity of use of fourteen external sources of knowledge: customers; suppliers of components, of equipment and of software; competitors; firms in the same enterprise group; universities; other research institutions; consultancy firms; institutions of technology transfer; patent disclosures; professional conferences and journals; fairs and exhibitions; and computer-based networks. Moreover, we include variables representing R&D out-contracting, as well as institutionalised R&D co-operation; for both types of arrangements, we distinguish between domestic and foreign relationships. There are good reasons to draw on such detailed information pertaining to external knowledge relations. Firstly, the importance of co-operation and networking has, over time, become significantly more important to the generation of innovations (Haagedoorn, 1996; for Switzerland, see Arvanitis et al., 1998, 2001a). Secondly, the pattern of use of external knowledge is one of the most important features of the modes of innovations in manufacturing, as identified in the studies mentioned in Section 1.

Moreover, we also include several structural characteristics of firms, such as industry affiliation, size, age and export orientation. Finally, our description of clusters draws on two measures of firm performance: namely, nominal value added per employee and the development of nominal sales over time.[6]

The Tables 4a (indicators of innovation), 4b (factors determining innovation), 4c (the knowledge network) and 4d (the structural characteristics of firms) show the means of these variables for each of the five clusters, as well as for the service sector as a whole. The corresponding information regarding industry affiliation and performance is presented in Section 4 (Tables 5 and 6 respectively). In the summary that follows, we shortly characterise the five clusters in terms of these categories. For more detail, we ask the reader to study the corresponding tables.

Mode 1: "Science-based high-tech firms with full network integration"

This cluster consists of 21 firms (4.4% of the firms, 18.1% of employment), which are endowed with an excellently qualified staff, and are engaged intensively in R&D within a highly favourable environment in terms of innovation opportunities and market perspectives. Internal R&D is supported by an intensive use of science-related external sources of knowledge, as well as many institutionalised co-operative R&D projects (and research contracts), with domestic and foreign universities serving as the primary partners. Innovation output consists in many instances of products and processes which are new to the industry and are protected by patents (accompanied by the granting of licences). The sales share of new products is high. This cluster contains an above-average proportion of export-oriented, medium-sized firms, in addition to some very large firms, which are heavily concentrated in IT/R&D-services, business services (70% of the firms), as well as in banking, insurance and other financial services (15% of the firms). Nominal labour

[6] In this study, we are not confronted with the well-known problem of measuring productivity in services, since we only take nominal productivity at a certain point in time (cross section-analysis) into consideration. Difficulties arise solely in the measurement of the change in real output over time, which requires adequate price deflators.

productivity is distinctly below-average; although sales growth is more favourable, i.e. about average.

Mode 2: "IT-oriented network-integrated developers"

This cluster contains 19 firms (4.0% of the firms, 1.7% of employment) whose innovative activities are supported by very favourable market perspectives and a highly qualified labour force. Based on high investments in development and IT (but not in research), the firms in this cluster generate product and process innovations which are of high technical standards and, in many instances, new to the industry. The innovations, often patented and licensed to other firms, are technology-oriented and characterised by a high IT-content and a great potential for cost reduction. These firms are intensive users of manifold sources of external knowledge (suppliers of software and investment goods, universities, competitors, firms in the same enterprise group). Among the more formal knowledge links, the out-contracting of R&D (at home, as well as abroad) and the use of licences are of higher importance than more far-reaching R&D co-operation. Medium-sized and export-oriented firms are clearly more frequent in this cluster than in services as a whole. Compared to the sector average, IT/R&D-services as well as banking/insurance/-other financial services are overrepresented, whereas the opposite is true for retail trade, hotels/restaurants and real estate, which are characterised, on average, by a rather low innovation intensity. Value added per employee is distinctly higher than in the other four categories of firms, whereas growth in sales is lower than in services as a whole.

Mode 3: "Market-oriented incremental innovators with weak external links"

The innovative activities of the 99 firms belonging to this cluster (20.9% of the firms, 9.1% of employment) are fostered by very favourable market prospects, whereas the supply-side conditions for the generation of novelties are just average. The firms in this cluster generate product and process innovations with high IT-content, which are primarily incremental in nature (this is not surprising in view of the rather low level of innovation input). Nevertheless, innovation output is of a high value in economic and technological terms, and is successfully brought into the market place (high sales share of innovative products). In general, networking is rather weak; only market-oriented sources of knowledge (users, software suppliers) and easily accessible knowledge sources (fairs/exhibitions, computer-based networks) are of any importance. Compared to the sector mean, this cluster consists of a high proportion of (very) small firms with an average export orientation. The firms are distributed across industries quite similarly to services as a whole, with a slight overrepresentation of business services and wholesale trade and only few firms in transport/telecommunication. In this innovation mode, labour productivity is high, whereas growth in sales is not above average.

Mode 4: "Cost-oriented process innovators with strong external links along the value chain"

This cluster of 229 firms is by far the largest (48.2% of firms, 62.9% of employment). In view of the strong price competition and only slightly above-average market growth, it is no surprise that (incremental) process innovations aimed at cost reductions are the most prominent feature of innovative activity in this cluster. Innovation input concentrates on IT-expenditures and innovation-related follow-up investments, whereby all components are highly relevant (machinery, external knowledge, training, marketing). The technological and economic significance of innovation output is high. The firms' own innovative activity strongly benefits from a wide (primarily informal) network that spans the entire value chain, from suppliers (in particular, of software) to users, with strong links to different partners between each end of the chain (consultancy firms, competitors, fairs and exhibitions, computer-based networks, conferences). Institutionalised co-operation (R&D contracts and R&D co-operation) is only of average importance. Large firms are somewhat overrepresented, very small ones distinctly underrepresented, and export orientation is rather low. In view of the large number of firms in this cluster, it is not surprising that the industry structure is close to the sector average. The same holds for labour productivity. Sales growth is, however, higher than in any other cluster.

Mode 5: "Low-profile innovators with hardly any external links"

The (process) innovations of the 107 firms belonging to this cluster (22.5% of firms, 8.2% of employment) appear to be quite marginal. This is not surprising in view of the unfavourable demand- and supply-side factors determining innovative activity: weak demand prospects, strong price competition, low appropriability and innovation opportunities, and relatively poor human capital endowment. This cluster performs most weakly with respect to the majority of variables used to characterise modes of innovation. The adoption of novelties generated elsewhere is the most important form of innovation. Correspondingly, innovation input consists primarily of machinery and equipment supplied by manufacturing firms. The use of external knowledge, which is below average for almost all sources, is concentrated (in relative terms) around suppliers and competitors. This cluster comprises an above-average share of small firms, which produce for domestic markets and belong to less innovative industries, such as personal services, real estate, hotels and restaurants, retail trade and transport. Not surprisingly, the economic performance of the firms in this cluster is weak.

We conclude that the five clusters differ in terms of the specific configurations of the variables we use to characterise innovative activities and the innovation-related environment. It is particularly important that this also holds true with respect to the "external criteria", that is those variables not implemented in the clustering process (i.e. the variables shown in Tables 4b-4d). Against this background, the five clusters can be safely interpreted as specific innovation modes.

Table 4a. Innovative Activity by Cluster

	Cluster					
Innovation Indicators	1	2	3	4	5	Total
1. Input-oriented measures						
Expenditures for						
- Research	50	0	2	1	2	9
- Development	90	39	8	14	3	14
- IT (hardware, software)	30	56	32	55	20	42
Follow-up investments						
- Total	30	17	6	29	11	20
- By type						
- Machinery and equipment	30	17	5	30	21	23
- Acquisition of external knowledge	5	17	3	14	7	10
- Training	25	22	13	52	25	36
- Market introduction of innovations	50	28	9	43	16	30
2. Output-oriented measures						
Significance of the innovations in technical terms						
- Product	75	67	71	59	29	56
- Process	55	67	69	69	26	58
Significance of the innovations in economic terms						
- Product	65	17	76	60	27	55
- Process	55	44	70	70	19	57
IT-content of innovations	15	67	65	46	26	45
Patent application (% yes)	75	44	0	1	1	6
Licences granted (% yes)	45	61	5	3	1	7
3. Market-oriented measures						
Sales share of new or highly improved services (%)	28	9	29	13	10	16
Cost reduction generated by process innovations (yes/no)	35	50	23	49	12	35

If not otherwise specified, the table shows for each cluster and the total sample the percentage share of firms with scores 4 or 5 on a 5-point ordinal scale (for definition, see Table 1); for example, 50% of the firms in Cluster 1 spend (very) much on research.

Table 4b. Determinants of Innovation Performance by Cluster

	Cluster					
Innovation Determinants	1	2	3	4	5	Total
1. Demand side						
Demand prospects	75	67	70	65	46	62
Intensity of price competition	70	61	62	69	65	66
Intensity on non-price competition	50	50	53	61	50	56
2. Supply side						
Innovation opportunities	75	22	35	41	31	38
Appropriability of knowledge	35	17	9	7	2	8
Highly qualified labour (%)	24	13	10	7	7	9

If not otherwise specified, the table shows for each cluster and the total sample the percentage share of firms with scores of 4 or 5 on a 5-point ordinal scale (for definitions, see Table 3); for example, 75% of the firms in Cluster 1 have (very) good demand prospects in their product market.

Table 4c. Characteristics of the Firms' Knowledge Network by Cluster

	Cluster					
Knowledge Sources / R&D Networks	1	2	3	4	5	Total
1. Sources of external knowledge						
Users	35	44	38	51	32	43
Suppliers of materials/components	40	28	29	36	26	32
Suppliers of software	25	72	32	44	28	38
Suppliers of machinery/equipment	25	28	10	19	9	15
Competitors	20	44	30	47	39	41
Other firms in the same group	35	33	12	21	9	18
Universities	45	33	18	14	18	18
Other research institutions	40	0	9	7	5	8
Consultants	25	17	11	24	7	17
Technology transfer organisations	10	11	3	9	7	8
Patent documents	20	0	2	0	3	2
Fairs/exhibitions	25	22	26	27	13	23
Scientific/trade journals, conferences	55	39	34	44	36	41
Computer networks	20	33	33	34	26	32

Table 4c. continued

	Cluster					
Knowledge Sources / R&D Networks	1	2	3	4	5	Total
2. Out-contracting of R&D						
At home (% yes)	60	50	11	20	13	19
Abroad (% yes)	40	17	4	6	2	7
3. R&D co-operation						
% with more than 2 domestic partners	45	19	10	10	13	14
% with more than 2 foreign partners	45	25	13	8	3	12

If not otherwise specified, point ordinal scale (for definition, see Table 3); for example, in Cluster 1, users are a (very) important source of knowledge the table shows for each cluster and the total sample the percentage share of firms with scores of 4 or 5 on a 5- for 35% of firms.

Table 4d. Selected Structural Characteristics of Firms by Cluster

	Cluster					
Structural Characteristics	1	2	3	4	5	Total
1. Share of firms (%) by size class (number of employees)						
5-19	25	22	34	24	34	28
20-49	20	17	25	23	21	23
50-199	30	28	21	27	27	26
200-499	10	28	12	13	8	12
500+	15	5	8	13	10	11
2. Share of young firms (%)						
Start-up in 1989 or later	21	18	15	8	11	11
3. Share of firms by export/sales ratio (%)						
Up to 1	20	50	68	67	67	65
2-19	15	17	11	14	17	14
20+	65	33	21	19	16	21

4 Empirical Results II: Are Innovation Modes Equivalent in Economic Terms?

It is general practice to rank industries according to their innovativeness, in order to assess their competitiveness or to predict their opportunities and risks with respect to structural change (see e.g. European Commission, 1997). As mentioned in Section 1, sectoral ranking is sensible if industries are sufficiently homogeneous in terms of innovation intensity. This assumption holds only if innovation modes (which according to their construction are homogeneous groups) and industries closely correspond in terms of innovativeness. In addition, inferences regarding competitiveness (and the like), which are based on innovation rankings by industry, require, as a necessary though not sufficient condition, systematic differences between innovation modes in terms of economic performance. It is precisely this hypothesis, which is denied by advocates of the classificatory approach to innovation. In the following section, we shall discuss the first of these conditions, whereas Section 4.2 will be devoted to the second.

4.1 The relationship between innovation modes and industries

Table 5 shows the industrial composition of the five innovation modes and the service sector as a whole. Industries are ordered according to decreasing levels of innovation intensity, as measured by an indicator which aggregates the information contained in the seventeen innovation measures listed in Table 1 by means of a factor analysis; for details of the procedure, see Hollenstein (1996). Although the five innovation modes are primarily regarded as unordered classes, we can also rank them according to innovation intensity, with Mode 1 ("Science-based high-tech firms with full network integration") at the top and Mode 5 ("Low-profile innovators with hardly any external links") at the bottom of the ladder.

A first look at Table 5 reveals that the firms in four of the five innovation modes (the exception is Mode 1) are distributed across several industries, thereby contradicting the "homogeneity assumption". However, at least three innovation modes are quite strongly concentrated in a few industries. This holds most clearly for Mode 1 ("Science-based high-tech firms with full network integration"), which primarily contains firms from the two most innovative industries. To a lesser extent, the industry composition of Mode 2 ("IT-oriented network-integrated developers") is also biased towards the most innovative industries. The opposite is true for Mode 5 ("Low-profile innovators with hardly any external links") which, in comparison to the sector average, is mainly present in industries with low or intermediate levels of innovation intensity. Concentration is less pronounced in Mode 3 ("Market-oriented incremental innovators with weak external links"), whereas firms in Mode 4 ("Cost-oriented process innovators with strong external links along the value chain") are distributed across industries in almost the same way as the service sector as a whole (which is not very surprising, since nearly half of the firms in our sample belong to this group).

Thus we find, on the one hand, a clear and positive correspondence between industries and innovation modes (when ranked according to innovation intensity). This result is supported by a statistically significant value of the Goodman-Kruskal γ, which is an appropriate measure of association in the case of ordinally-scaled variables. On the other hand, four out of five innovation modes are distributed quite widely across industries, which, however, might partly reflect the high sectoral aggregation (nine industries only). We conclude that the evidence does not clearly favour any one of the competing hypotheses (classification vs. ranking).

Table 5. Industrial Structure by Innovation Mode

Industry	Innovation Mode (Cluster)					
	1	2	3	4	5	Total
	Distribution of firms by industry (%)					
Innovativeness above average	*95.3*	*83.2*	*73.8*	*72.6*	*67.5*	*73.0*
IT and R&D services	33.3	16.7	5.1	3.9	2.8	5.5
(Other) business services	38.3	16.7	23.2	18.3	15.7	19.6
Banking/insurance/financial services	14.3	22.1	16.2	17.0	15.7	16.4
Wholesale trade	4.7	22.1	24.2	22.5	15.7	20.5
Transport/telecommunication	4.7	5.6	5.1	10.9	17.6	11.0
Innovativeness below average	*4.7*	*16.8*	*26.2*	*27.4*	*32.5*	*27.0*
Retail trade	0.0	5.6	14.1	13.0	16.7	13.3
Hotels, restaurants	4.7	5.6	9.1	12.2	11.1	10.7
Real estate	0.0	0.0	1.0	0.9	1.9	1.1
Personal services	0.0	5.6	2.0	1.3	2.8	1.9
Total	*100*	*100*	*100*	*100*	*100*	*100*

The innovation modes are: (1) "Science-based high-tech firms with full network integration", (2) "IT-oriented network-integrated developers", (3) "Market-oriented incremental innovators with weak external links", (4) "Cost-oriented process innovators with strong external links along the value chain", (5) "Low-profile innovators with hardly any external links"; see description in the text.

4.2 Innovation modes and economic performance

To assess the validity of the "homogeneity hypothesis", we must also investigate whether there are significant differences between innovation modes with respect to average firm performance. A negative result would support the "heterogeneity hypothesis", according to which more than one innovation strategy is, at least temporarily, economically feasible. To evaluate the two conflicting propositions, we consider nominal labour productivity (value added per employee) as a measure of firm performance, and nominal sales growth as an indicator of change in performance over time.

As is illustrated by Table 6, levels of labour productivity vary strongly between the five innovation modes; it is 44% higher in Mode 2 (the cluster with the highest level of productivity) than in Mode 1 (the cluster with the lowest level of productivity). We also find substantial differences with respect to the growth of sales. These data seem to be at variance with the "heterogeneity hypothesis". However, this view of the matter is too simple, since firm performance is determined not only by the variable "innovation mode" but also (and probably to an even larger extent) by other factors, such as those listed in the lower part of Table 6. For example, it is obvious that a firm which intensively uses physical capital, exhibits (ceteris paribus) a higher level of labour productivity than one in which production is more labour intensive; neglecting the contributions of other determinants of productivity would clearly bias a comparison of the performance of Mode 3 (high value added per employee, high capital intensity) and Mode 1 (low value added per employee, low capital intensity), and so on.

Table 6. Economic Performance by Innovation Mode

	Innovation Mode (Cluster)					
Indicator	1	2	3	4	5	Total
	Cluster means					
Performance indicators						
Nominal value added per employee in 1998 (1000 SFr.)	153	221	196	171	172	178
Share of firms (%) with increasing nominal sales 1996/98	50	28	34	43	32	38
Factors determining firm performance (in addition to cluster affiliation)						
Gross capital income per employee (1000 SFr.)	56	67	87	86	72	81
Employment share of highly qualified labour (%)	50	32	27	22	19	24
Employment share of R&D personnel (%)	34	6	1	2	1	3

The innovation modes are: (1) "Science-based high-tech firms with full network integration", (2) "IT-oriented network-integrated developers", (3) "Market-oriented incremental innovators with weak external links", (4) "Cost-oriented process innovators with strong external links along the value chain", (5) "Low-profile innovators with hardly any external links"; see description in the text.

As a consequence, instead of comparing cluster means, we regress labour productivity at the firm level with the following independent variables: a) four dummies representing affiliation with Clusters 2, 3, 4 and 5 respectively (with Cluster 1 as the reference group), b) variables which control for physical capital intensity (gross capital income per employee), human capital intensity (the employment share of personnel holding tertiary level degrees), and knowledge capital (the employment share of R&D personnel), and c) industry dummies. Sales growth, as a measure of the development of firm performance over time, was used as a dependent variable in a regression containing not only the explanatory variables used in the labour productivity estimates, but also two dummies which control for changes of the firms' boundaries (the selling-off or closure of parts of the firm, as well as mergers and acquisitions).

The results of these estimates can be summarised as follows: In the case of labour productivity, only an affiliation with Cluster 2 ("IT-oriented network-integrated developers") exerts a statistically significant (positive) impact. The other four modes cannot be distinguished from each other in terms of labour productivity. The productivity differences, as shown in Table 6, are mainly explained by variations in physical, as well as human and/or knowledge capital intensity (positive signs). In addition, we obtain statistically significant signs for some industry dummies, which are negative in the cases of two less innovative industries (retail trade, hotel and restaurants) and positive for the highly innovative banking and insurance industry.[7] When "sales growth" is the dependent variable, two of the dummies which measure affiliation with innovation modes exert a statistically significant influence: the sign of Mode 2 ("IT-oriented network-integrated developers") is negative, that of Mode 4 ("Cost-oriented process innovators with strong external links along the value chain") is positive. In addition, we find a significantly positive influence of knowledge capital intensity on sales growth, whereas the human and physical capital variables yield no statistically significant results.

In sum, these estimates of the relationship between performance and innovation mode (controlling for other important determinants of performance) reveal that labour productivity and sales growth differ with respect to innovation modes in only a few "extreme cases" (see Table 4). This result is more or less in line with the "heterogeneity hypothesis", which states that firms exercise a certain degree of freedom in choosing economically viable innovation strategies. However, an assessment of this conclusion should also take into consideration the fact that performance is additionally influenced by the intensity with which human and/or knowledge capital is used, as well as by several industry dummies systematically related to innovativeness. Therefore, we can conclude that the selection of an innovation strategy also depends quite strongly on structural characteristics that are closely related to the ranking of industries in terms of innovation intensity.

5 Do Innovations in Services Differ from those in Manufacturing?

As mentioned in the introductory section, many authors maintain that innovations in services differ significantly from those in manufacturing, whereas others postulate that such differences have been reduced in recent years due, for example, to the strong diffusion of IT in the service sector or the growing importance of customisation in manufacturing. What are the implications of the present analysis and a similar study pertaining to the Swiss manufacturing sector (Arvanitis and Hollenstein, 2001)?

Firstly, our results are in line with the proposition of lower R&D in services, as compared to manufacturing. Only innovation Modes 1 and 2 (less than 10% of firms) are strongly R&D-based (see Table 4a), whereas in Swiss manufacturing, the same holds true for three out of five clusters that cover more than 50% of firms. This result is confirmed by data from the OECD (2001), even if it is assumed that in service industries R&D is

---

[7] Dropping the industry dummies to avoid multicollinearity (as shown above, there is some correlation between industry affiliation and innovation mode) did not affect the results with respect to the cluster variables.

generally underestimated in official statistics (measurement problems; only partial coverage of the service sector).[8] Hence, a relatively low level of R&D remains an important characteristic of service innovations, although the share of this sector in total business R&D increased significantly during the nineties.

Secondly, it is argued that human resources play a particularly important role in the generation of service innovations. We observed that human capital input is high in three out of five innovation modes. Since the use of highly qualified manpower is almost as intensive as in manufacturing, we conclude that human resources indeed play a very important role in the innovation process within the service sector. However, this is not a peculiarity of service innovations.

Thirdly, some of the literature mentions the high information content of services and the widespread use of IT as an additional characteristic of service innovations. Our results confirm this proposition. However, the use of IT in services is not more intensive than in manufacturing, at least in Switzerland (Arvanitis et al., 2001b). Hence, IT is a characteristic of innovative activities in general, rather than a specific feature of service innovations.

Finally, the high relevance of non-technological innovation is often seen as the most distinct feature of innovation in services. The cluster-specific pattern of innovation found in this study allows a tentative assessment of the role played by this type of innovative activity. We therefore hypothesise that non-technological innovations are an important element of innovation patterns, when the following two conditions are fulfilled: a) high values of innovation variables which are related to non-technological factors (i.e. the economic significance of innovation, innovation-related outlays for training and marketing respectively, the sales share of innovative products, and human capital intensity), and b) low values of innovation variables closely related to technological factors. In contrast to this proposition, our results indicated that in three of the five innovation modes, both sets of innovation variables are either high (Clusters 1 and 2) or low (Cluster 5); these three clusters contain about one third of the firms. Cluster 4 shows no clear pattern with respect to the relative importance of technological and non-technological aspects of innovation (nearly 50% of the sample firms). Only in Cluster 3 (about 20% of the firms), did we find some confirmation of the hypothesis that non-technological factors are more important than technological ones. We thus tentatively conclude that innovations based primarily on non-technological factors play a dominant role in only some segments of the service sector. In contrast, in Swiss manufacturing, technological factors shape the pattern of all innovation modes we identified in earlier work (Arvanitis and Hollenstein, 2001).[9] However, in most parts of the service sector, a strong technology base seems to be a prerequisite to good performance in innovation as well (which does not exclude the possibility that such firms also generate non-technological innovations). Recent survey data on the importance of

---

[8] The problem of measurement of R&D in services is discussed in detail in Jankowski (2001) and in various papers of the OECD devoted to the revision of the Frascati Manual, is due to be completed at the end of 2002 (OECD, 2002). Even if this revision will lead to a (further) correction of the underestimation of R&D in services, manufacturing, on average, will certainly remain more R&D intensive than services.

[9] This holds true although non-technological innovations (e.g. customisation of new products; fundamental organisational change) have become more important also in manufacturing.

organisational innovations[10] in the Swiss business sector, which are not covered by the data base used in this study, indicate that in this respect, there are only minor differences between manufacturing and services (Arvanitis et al., 2001b). This result is consistent with our assessment that the differences between manufacturing and services with respect to non-technological innovations are not very large.

All in all, we can conclude that innovations in services do differ from those in manufacturing. However, it seems that these are less accentuated than hypothesised by many authors. Therefore, we are rather inclined to support the hypothesis put forward by, among others, Coombs and Miles (2000), who postulate that differences between the two sectors have been blurred in recent years and are now one of degree rather than of substance.

6 Summary, Discussion and Conclusions

By applying cluster analysis to a large set of innovation indicators (which, to some extent, also capture non-technological aspects of innovation), we identified five clusters. These were characterised by the use of several groups of variables: a) innovation indicators, b) demand- and supply-side determinants of innovative activity, c) the firms' position in knowledge networks, d) several structural characteristics of firms, and e) measures of firm performance.

In view of the distinct patterns exhibited by these variables, the clusters can be interpreted as specific "innovation modes":

- "Science-based high-tech firms with full network integration"
- "IT-oriented network-integrated developers"
- "Market-oriented incremental innovators with weak external links"
- "Cost-oriented process innovators with strong external links along the value chain"
- "Low-profile innovators with hardly any external links"

According to the evolutionary view of technical change, such a classificatory procedure ("innovation modes") is more appropriate than an approach which ranks industries according to innovativeness. Whereas the starting point of classification is the heterogeneity of firms with respect to innovation strategies, the ranking approach assumes that an industry is more or less homogeneous in terms of the innovativeness of its firms. In order to evaluate the relative merits of the two approaches, we investigated the sectoral distribution of innovation modes, as well as the differences across innovation modes with respect to economic performance. We found, firstly, that the firms in most innovation modes are distributed across several industries; however, taking the service sector average as our benchmark, three of five innovation modes are (heavily) concentrated in specific industries. Secondly, we found that economic performance is related to the

[10] Change in the number of management layers, redistribution of competencies among hierarchical levels, introduction of team-based work, job-rotation programmes, etc.

affiliation with a specific innovation mode in only one or two of the five modes, depending on the performance measure used. However, there is strong evidence that in addition to cluster affiliation, variables related to innovativeness (such as human and knowledge capital intensity), as well as dummies for industries with an above-average (below-average) innovation performance, exert a positive (negative) influence on firm performance.

These results imply that neither the "classical" ranking of industries according to innovativeness nor the classification of firms into unordered categories representing innovation modes of equal "economic value" capture the whole reality. This ambiguous result can be interpreted as follows: In accordance with the heterogeneity hypothesis (classification approach), firms exercise a certain degree of freedom in selecting an economically viable innovation strategy, even in similar economic and technological environments; however, their room for manoeuvre is restricted by structural characteristics closely related to the hierarchy of industries in terms of innovation intensity (ranking approach).

Therefore, we conclude that the widespread practice of ranking industries according to their innovation intensity in order to assess, for example, their competitiveness, (still) makes sense. However, such a procedure requires broadly-based measures of innovation intensity (use of many and different types of indicators); otherwise the diversity of innovation modes within an industry cannot be adequately taken into account. This aspect rarely receives the attention it deserves. Most sectoral rankings are based on a single, easily available indicator (e.g. R&D or patent intensity). In services, however, these two indicators are hardly relevant in three of the five innovation modes we identified. The aggregate measure of innovation, which we developed in earlier work as a means of ranking industries according to innovation intensity, is a useful instrument for taking into account the heterogeneity aspect, since it contains information from a large number of indicators covering quite different (and also non-technological) aspects of "innovativeness".

A comparison of our classification with the taxonomies characterised in Section 1 cannot be performed in a straightforward way, since there are important differences with respect to the level of aggregation (firm vs. industry), the method used (formal statistical procedures at the one extreme vs. (purely) conceptual work at the other) and the time dimension of the analysis (static vs. dynamic view).

Evangelista (2000), who also applied cluster analysis (although at the industry-level), basically identified three categories of industries; two of them ("S&T-based" and "Technology users") are very similar to our Innovation Modes 1 and 5 ("Science-based high-tech firms with full network integration", and "Low-profile innovators with hardly any external links"). His third category ("Interactive and IT-based”) exhibits some similarities to our Innovation Modes 4 and 2 ("Cost-oriented process innovators with strong external links along the value chain" and "IT-oriented network-integrated developers"), although the correspondence is far from perfect. However, there is no cluster in Evangelista’s taxonomy which corresponds to our Mode 3 ("Market-oriented incremental innovators with weak external links"). This result might reflect, to some extent, the specific procedure used in the Italian case where, at a certain stage in the analysis, market-oriented innovation indicators were dropped for technical reasons (Evangelista, 2000, p. 211). If this interpretation holds true, the two taxonomies would be quite similar. We also find similarities to the taxonomy proposed by Soete and Miozzo (1989). "Specialised producers/Science-based industries" and "Supplier-dominated industries" correspond strongly to our Innovation Modes 1 and 5. These two modes have thus been identified in all the studies compared up to this point. However, both "Network-based sectors" proposed by the two authors can only be loosely related to our classification (subsector "information networks" vs. our Mode 2, i.e. "IT-oriented network-integrated developers").

A comparison with Gallouj (1999) is difficult, since he primarily strives to define a suitable analytical framework for developing a taxonomy of innovation trajectories. A basic difference between his and our work and the taxonomies mentioned so far, is the dynamic nature of his approach; it yields a multiplicity of innovation trajectories which evolve over time in terms of a changing mix of the basic elements constituting a trajectory. Nevertheless, it would be possible to develop a dynamic version of our approach. By using information from different waves of the Swiss innovation survey, we could pool cross-sectional and longitudinal data. In this way, we would be able to identify innovation modes based on a time-dependent database (panel); moreover, we could analyse to what extent firms are changing their innovation mode over time and whether there are typical patterns of transition from one innovation mode to another.

Another aspect stressed by many authors is the greater importance of non-technological innovations in services as compared to manufacturing. In this respect, we conclude that such a difference exists, although it seems to be one of degree rather than of substance. Against this background, it would be sensible to look for innovation modes using data covering both sectors. This would enable us to gain greater insight into the differences between innovation patterns prevailing in the two sectors.

Furthermore, it would be valuable to make cross-country comparisons of innovation modes based on the same type of data and method. This might be feasible, since the innovation surveys conducted in the EU member states (CIS) and other countries are (more or less) harmonised. Such work would enable the identification of the innovation modes that are common to most countries, as well as country-specific innovation patterns. The identification of country-specific features would contribute to the characterisation of "National Innovation Systems".

As far as policy conclusions are concerned, one must exert caution at this stage of the research. Nevertheless, the results do have some implications which should be considered by policy makers. Firstly, in assessing and shaping policy measures, one should take into account the variety of innovation patterns, since firms belonging to specific innovation modes have different needs with respect to public policy. For example, firms belonging to the innovation mode "IT-oriented network-integrated developers" would profit most from measures contributing to the enlargement and improvement of the supply of IT-professionals and from programmes facilitating the diffusion of IT. On the other hand, "Science-based high-tech firms with full network integration" may be supported, in the first place, by strengthening the production of (basic) scientific knowledge, as well as by measures to improve its transfer to the business sector. Considering the poor economic performance of "Low-profile innovators with hardly any external links", suitable measures for strengthening outside links could perhaps contribute to improving the innovativeness of this type of firm. Secondly, to the extent that innovation modes are country-specific, there are limitations to designing best policy practices at the international level.

References

Arundel, A., van der Paal, G., and L. Soete, 1995, Innovation Strategies of Europe's Largest Industrial Firms, Report Prepared for the SPRINT Programme (European Commission, DG XIII, Brussels).

Arvanitis, S., 2000, Explaining Innovative Activity and its Impact on Firm Performance in Service Industries: Micro Data Evidence for Switzerland, Paper Presented at the "25th CIRET Conference", Paris, October 10-14.

Arvanitis, S., Bezzola, M., Donzé, L., Hollenstein, H. und D. Marmet, 2001a, Innovationsaktivitäten in der Schweizer Wirtschaft. Eine Analyse der Ergebnisse der Innovationserhebung 1999, Studienreihe Strukturberichterstattung Nr. 5 (Staatssekretariat für Wirtschaft, Bern).

Arvanitis, S., Donzé, L., Hollenstein, H. und S. Lenz, 1998, Innovationstätigkeit in der Schweizer Wirtschaft, Teil I: Industrie, Teil II: Bauwirtschaft und Dienstleistungssektor, Studienreihe Strukturberichterstattung (Bundesamt für Wirtschaft und Arbeit, Bern).

Arvanitis, S. and H. Hollenstein, 1994, Demand and Supply Factors in Explaining the Innovative Activity of Swiss Manufacturing Firms, An Analysis Based on Input-, Output- and Market-oriented Innovation Indicators, Economics of Innovation and New Technology 3, 15-30.

Arvanitis, S. and H. Hollenstein, 1996, Industrial Innovation in Switzerland: A Model-based Analysis with Survey Data, in: A. Kleinknecht (Editor), Determinants of Innovation and Diffusion (Macmillan, London) pp. 13-62.

Arvanitis, S. and H. Hollenstein, 2001, Innovative Activity and Firm Characteristics. A Cluster Analysis of Swiss Manufacturing Using Firm-level Data, in: OECD (Editor), Innovative Networks: Co-operation in National Innovation Systems (OECD, Paris) pp. 48-76.

Arvanitis, S., Marmet, D. und D. Staib, 2001b, Einsatz von Informations- und Kommunikationstechnologien in den schweizerischen Unternehmen, Bericht z.Hd. des Bundesamtes für Statistik (Konjunkturforschungsstelle der ETH, Zürich).

Cesaratto, S. and S. Mangano, 1993, Technological Profiles and Economic Performance in the Italian Manufacturing Sector, Economics of Innovation and New Technology 2, 237-256.

Cohen, W.M., 1995, Empirical Studies of Innovative Activity, in: P. Stoneman (Editor), Handbook of the Economics of Innovation and Technological Change (Basil Blackwell, Oxford and Cambridge (Mass.)) pp. 182-264.

Cohen, W.M. and D.A. Levinthal, 1989, Innovation and Learning: The Two Faces of R&D, Economic Journal 99, 569-596.

Coombs, R. and I. Miles, 2000, Innovation, Measurement and Services: The New Problematique, in: S.J. Metcalfe and I. Miles (Editors), Innovation Systems in the Service Sectors. Measurement and Case Study Analysis (Kluwer, Boston-Dordrecht-London) pp. 85-104.

Djellal, F. and F. Gallouj, 1999, Services and the Search for Relevant Innovation Indicators: A Review of National and International Surveys, Science and Public Policy 26, 218-232.

Donzé, L., 2001, L'imputation des données manquantes, la technique de l'imputation multiple, les conséquences sur l'analyse des données: l'enquête 1999 KOF/ETHZ sur l'innovation, Revue suisse d'économie et de statistique 137, 301-318.

European Commission, 1997, Second European Report on S&T Indicators 1997, 2 Vols., Brussels.

Evangelista, R., 2000, Sectoral Patterns of Technological Change in Services, Economics of Innovation and New Technology 9, 183-221.

Evangelista, R. and G. Sirilli, 1995, Measuring Innovation in Services, Research Evaluation 5, 207-215.

Ost, F., 1984, Faktorenanalyse, in: L. Fahrmeir und A. Hamerle (Editors), Multivariate statistische Verfahren (de Gruyter, Berlin-New York) pp. 575-662,

Gallouj, F., 1999, Les trajectoires de l'innovation dans les services: vers un enrichissement des taxonomies évolutionnistes, Économies et Sociétés, Série Économie et Gestion des Services (EGS) 1, 143-169.

Haagedoorn, J., 1996, Trends and Patterns in Strategic Technology Partnering Since the Early Seventies, Review of Industrial Organization 11, 601-616.

Hollenstein, H., 1996, A Composite Indicator of a Firm's Innovativeness. An Empirical Analysis Based on Survey Data for Swiss Manufacturing, Research Policy 25, 633-645.

Hughes, A. and E. Wood, 2000, Rethinking Innovation Comparisons between Manufacturing and Services: The Experience of the CBR SME Survey in the UK, in: S.J. Metcalfe and I. Miles (Editors), Innovation Systems in the Service Sectors. Measurement and Case Study Analysis (Kluwer, Boston-Dordrecht-London) pp. 105-124.

Jankowski, J.E., 2001, Measurement and Growth of R&D Within the Service Economy, Journal of Technology Transfer 26, 323-336.

Manly, B.F.J., 1986, Multivariate Statistical Methods. A Primer (Chapman and Hall, London).

Metcalfe, S., 1995, The Economic Foundations of Technology Policy: Equilibrium and Evolutionary Perspectives, in: P. Stoneman (Editor), Handbook of the Economics of Innovation and Technological Change (Basil Blackwell, Oxford and Cambridge (Mass.)) pp. 409-512.

Miles, I., 1995, Services Innovation, Statistical and Conceptual Issues, Working Group on Innovation and Technology, Doc. DSTI/EAS/STP/NESTI/(95)12 (OECD, Paris).

OECD, 2001, OECD Science, Technology and Industry Scoreboard 2001. Towards a Knowledge-based Economy (OECD, Paris).

OECD, 2002, Revision of the Frascati Manual: Chapter 2. Basic Definitions and Conventions, Working Party of National Experts on Science and Technology Indicators, Doc. DSTI/EAS/STP/NESTI(2002)9/CHAP2 (OECD, Paris).

Pavitt, K., 1984, Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory, Research Policy 13, 343-373.

Pavitt, K., Robson, M. and J. Townsend, 1989, Accumulation, Diversification and Organisation of Technological Activities in UK Companies, 1945-83, Management Science 35, 81-99.

Rubin, D.B., 1987, Multiple Imputation for Nonresponse in Surveys (John Wiley & Sons, New York).

Soete, L. and M. Miozzo, 1989, Trade and Development in Services: A Technological Perspective, Research Memorandum, 89-031 (MERIT, Maastricht).

Stoneman, P. (Editor), 1995, Handbook of the Economics of Innovation and Technological Change (Basil Blackwell, Oxford and Cambridge (Mass.)).

Appendix

Table A1. Structure of the Sample and the Final Data Set

	Sample		Respondents			Innovators		
	(1) N	(2) %	(3) N	(4) %	(5) (3)/(1)	(6) N	(7) %	(8) (6)/(3)
Industry								
Wholesale trade	596	21.8	207	23.5	34.7	101	21.2	48.8
Retail trade	516	18.9	132	15.0	25.6	63	13.3	47.7
Hotels, restaurants	403	14.8	84	9.6	20.8	50	10.6	59.5
Transport/communication	378	13.8	133	15.1	35.2	52	11.0	39.1
Banking/insurance	266	9.7	99	11.2	37.2	77	16.2	77.8
Real estate	38	1.4	14	1.6	36.8	5	1.1	35.7
IT and R&D services	100	3.7	36	4.1	36.0	26	5.4	72.2
Business services	384	14.1	155	17.6	40.4	92	19.3	59.4
Personal services	50	1.8	20	2.3	32.2	9	1.9	45.0
Total	*2731*	*100*	*880*	*100*	*32.2*	*475*	*100*	*54.0*
Firm size (number of employees)								
Small	1487	54.4	465	52.8	31.3	218	46.0	46.9
Medium	1021	37.4	330	37.5	32.3	194	40.7	58.8
Large	223	8.2	85	9.7	38.1	63	13.3	74.1
Total	*2731*	*100*	*880*	*100*	*32.2*	*475*	*100*	*54.0*

Column 5 shows the response rate by industry and size class, column 8 the share of innovating firms.

Spillover Effects and R&D-Cooperations: The Influence of Market Structure

Anita Wölfl

Abstract

This paper[1] examines empirically the role of market structure for the influence of spillover effects on R&D-cooperations. The results of a microeconometric analysis, based on firm data on innovation, let in general presume that with intensified competition also the influence of spillovers on R&D-cooperation increases. However, competition seems to induce firms to search for effective firm-specific appropriation facilities first. Spillovers that are sufficiently high such that the internalisation effect from R&D-cooperation more than outweighs the competitive effect from research, only arise whenever firms are not able to protect their research results through any appropriation facility. Additionally, there is some evidence that spillover effects may even hinder firms from cooperating in R&D when there is intensive competition on the research stage.

1 Introduction

This paper examines empirically the incentive of firms to cooperate in R&D whenever they are not able to fully appropriate their research results. In contrast to previous theoretical and empirical papers, it mainly analyses which role market structure, essentially the intensity of competition, plays for the influence of spillover effects on R&D-cooperations.

Spillover effects arise whenever know-how or research results of one firm are used by other firms without the latter having to bear any expenses. Having in mind these external economies due to research, several lines in economic theory see spillover effects as purely advantageous. However, when the researching firm has to fear that its internal research may indirectly spur competitors' profits, it will retain from investing into R&D. On the aggregate, investment into R&D may be suboptimal. Economists are proposing cooperations in R&D as one means for solving this market-failure due to spillover-effects. The idea is that by coordinating R&D, the external effect is exploited and by exchanging research results, know-how is disseminated. However, firms actually have to have an incentive to bind themselves in such contractual arrangements in order to internalise

[1] Particular thanks go to my former head of division at the Insitute for Economic Research in Halle, IWH, Joachim Ragnitz, and to my former collegues, especially Jacqueline Rothfels, Ralf Müller and Gerald Müller. For comments or information please contact: anita.woelfl@oecd.org.

spillover effects. According to the basic line of theoretical results the incentive for cooperations in R&D exists whenever spillover-effects are sufficiently high such that the positive internalisation effect of cooperation more than outweighs the negative competitive effect from investment into research.

By and large, the incentive to cooperate depends on the structure of the market where the firms are operating. On the one hand, the type and the intensity of competition on the product market and during the innovation process determine the extent of the competitive effect from research. Additionally, spillover effects themselves may intensify (potential) competition. As a consequence, the threat of potential competition induces firms to invest heavily into R&D; there may be a high incentive to internalise spillover effects within R&D-cooperations. On the other hand, there are some doubts whether firms, especially those that have been successful in building up high market power, actually have an incentive to cooperate in R&D. One reason is that high market power reflects efficient research and production. Therefore, this firm is not easily affected. Or, it may even fear to increase the probability of their competitors winning the race by passing on its internal research results via coordination.

This paper analyses empirically the role of market structure for the incentive to cooperate in R&D in order to internalise spillover effects. This is done using a logistic regression model based on firm level data on innovation. The analysis starts with a short discussion of the main theoretical arguments in chapter 2. After elaborating on some crucial methodological points, the presentation of the empirical results will follow in chapter 3. The paper will close with a final assessment of the role of market structure for the internalisation aspect of R&D cooperations.

2 Theoretical Background

2.1 Spillover Effects in Theory

The term "spillover effects" has reached a central position within the theoretical and empirical literature on the economics of innovation and technological change. Spillover effects[2] arise whenever firms use know-how of another firm without the researching firm being able to control or influence the degree of this unintended knowledge transfer. R&D thus produces a positive external effect in favour of other firms. From the viewpoint of social welfare, there is reason enough to promote spillovers, since they spur the dissemination of new knowledge available for the whole society. From the viewpoint of the single firm, however, spillovers may be judged negatively. The individual firm fears that competitors use its internal research results and thus probably increase their profits without having to bear the expenses. Therefore, the researching firm will only have limited incentive to invest into R&D.

2 This notion is equivalent to the notion "technological spillovers" that Katz and Ordover (1990) use in order to distinguish the external effects due to lack of appropriability from the pecuniary external effect from research itself, which they call the "competitive spillovers".

Spillover effects may arise in the production as well as in the diffusion phase of the innovation process.[3] According to Greeks (1994:102), the appropriability of a firm, i.e., its capacity to protect its research results, depends on various factors. These are for example the technology itself, the barriers to entry which exist on the market where the technology is used, and the capability of other firms to absorb external knowledge within their internal innovation process. According to Tirole (1993:403-4) there is only limited incentive for this firm to develop and launch a new product whenever one firm has to fear quick imitation. Instead of investing into R&D it will rather wait for a competitor to develop and launch the product first.

Spillover effects may thus lead to suboptimal R&D investment. However, R&D is essential for the production of innovations itself, both for the internal innovating process of the individual firm as well as for its capacity to absorb external knowledge. Additionally, R&D is essential as a foundation for innovations and thus for the overall performance of an economy. On the one hand, process innovations may be growth-inducing via increased productivity. On the other hand, product innovations may lead to higher consumer welfare via increased product variety.[4] Standard means to correct for the market failure due to spillover effects include R&D-subsidies and a stronger patenting system. However, subsidies have to be financed by taxes and may - according to Katz and Ordover (1990:140) - distort the incentive structure of firms with respect to their R&D investment decision. Strengthening the property rights through patents guarantees (temporary) monopoly power for one firm. This may restrict competition and may -according to Klodt (1995:32-38) - lead to welfare losses. More recent theoretical models focus on the question whether spillover effects can be internalised by the firms through binding contractual arrangements in research and development (R&D-cooperations). Fundamental for the functioning of such R&D-cooperations as an internalisation device is that firms actually have an incentive to coordinate their R&D-activities with other firms when spillover effects arise.

2.2 R&D-Cooperations in Theory

With the help of a simple duopoly-model, d’Aspremont and Jacquemin (1988) made clear how spillover effects are actually internalised by R&D-cooperations.[5] In their model two individual firms engage in Cournot competition, i.e., they are maximising their profits

[3] With regard to this, Katz and Ordover (1990:154) distinguish between "intermediate" and "final tech-nological spillovers".

[4] Concerning the relevance of spilloves for economic welfare and growth see for instance, Krugman (1992), Leahy and Neary (1997), Grossman and Helpman (1991), and Coe and Helpman (1995).

[5] This analysis will concentrate on the empirical test of the general theoretical relationship between spillover effects and R&D-cooperation. Therefore, the basic model of the duopoly itself should only illustrate the decision-making process of the individual firms. A direct implementation would have to start from an oligopoly and distinguish between product and process innovations. However, according to Kamien et al. (1992), Suzumura (1992), Morasch (1994) and the model for product-innovations of Motta in Konrad (1997) the central results will not change significantly by these extensions. R&D-cooperations and spillover effects in an open economy are analysed by Leahy and Neary (1997).

through the choice of the output level. Since they are using R&D as strategic investment, the firms additionally have to decide how much they want to invest into R&D. In order to separate the maximisation process with respect to these two parameters R&D-input and level of production it can be thought of a two-stage-decision process: On the first stage the firms decide upon their R&D-investments. On the second stage, they maximise their respective profits through the choice of the production level given a specific level of R&D.

D'Aspremont and Jacquemin concentrate on a model with process innovation. There are three channels through which internal R&D influences a firm's profit: Firstly, R&D directly causes (fixed) investment costs. Secondly, in the case of successful process innovation, R&D reduces production costs. Thus, the direct costs of producing innovations counterweight the indirect cost reduction in the production of the final good. Thirdly, there is an indirect channel the sign of which depends on the reaction of the competitors. In this respect, Katz and Ordover (1990:150) speak of "competitive spillovers".[6] With the firms acting independently from each other, each firm has an incentive to invest into R&D. R&D reduces marginal costs and thus increases profits from extended production. In the case of strategic substitutes, i.e., if the other firm reacts with a reduction of its own production, the researching firm can increase its market share and thus its profit. There is no incentive to cooperate. However, with both firms investing into R&D simultaneously, more overall output is produced, leading to lower prices and thus lower profits for each firm. In this case, R&D intensifies competition on the product market.

Whenever spillover effects prevail, there is a fourth channel through which R&D influences the profits of the firms. In models of process innovations, spillover effects occur when additionally to internal R&D some part of external R&D influences the production costs of the individual firm. In that case, the reduction of marginal costs through R&D is already achieved at a lower level of internal R&D. However, internal R&D reduces marginal costs of the competitor, too. Thus it increases the competitors' profits indirectly by influencing its market share. When both firms act independently from each other (R&D competition), each firm can profit from the know-how of the competitor. However, the incentive to invest into R&D diminishes due to the external effect of spillovers in favour of the competitor. In contrast, when the firms are coordinating their R&D activities, they can exploit this external effect. Morasch (1994:52) names it the positive "indirect efficiency effects": Synergies between the research projects are exploited and double research is prevented. According to the theoretical model, there is an incentive to cooperate whenever

[6] Sometimes, the terminus 'business-stealing effect' is used, for instance in Konrad (1997) and König et al. (1994).

high spillover effects prevail.[7] In this case, the positive internalisation effect from cooperation more than outweighs the negative competitive effect from research.[8]

R&D-cooperation in general can take two forms. First, the firms form an R&D-cartel on the first stage of the decision process. Thereby, they coordinate the amount of R&D they want to invest in order to maximise the common profit. Second, they form an R&D joint venture on the second stage by exchanging their research results without any agreement on the actual R&D-investment. According to Kamien et al. (1992), the optimal form of cooperation is the RJV-cartel. There, the firms cooperate with respect to both, the amount of R&D to be invested as well as the exchange of knowledge. On the one hand, exchanging the research results enables both firms to make use of all knowledge. On the other hand, while R&D still influences the profit of the competitor, this external effect is internalised by maximisation of the common profit. Both firms therefore have again an incentive to invest into R&D. Furthermore, according to Kamien et al. (1992:1302) the binding agreements on both stages reduce the incentive for freerider-behaviour. Thus, internalising the spillover effects through an RJV-cartel counteracts exactly the two channels through which suboptimal R&D would result in the case of competitive research: the coordination of the R&D-inputs encourages R&D and thus innovation. The exchange of research results provides the dissemination of knowledge throughout the whole economy.

2.3 The Influence of Market Structure

The incentive to cooperate depends on the structure of the market where the firms are operating. On the one hand, the type and the intensity of competition on the product market and during the innovation process determine the extent of the competitive effect from research. Additionally, spillover effects themselves may intensify (potential) competition: They reduce production costs of the competitors, which again may alleviate market entry. Or, immediate technological spillovers may increase the probability with which the competitor succeeds with his innovation. As a consequence, the efficiency effect of R&D-investment comes into place. According to Tirole (1993:393), firms have a high interest to use R&D investments as an instrument to strategically influence the market

---

7 In d'Aspremont and Jacquemin (1998) the critical value is around 0,5. This means that half of the re-search results of one firm can be used by the other firm. Simulations by Morasch (1994) support this result.

8 In this case, additional effects preventing cooperation like asymmetric information and coordination problems are counteracted. For a survey of theoretical models that handle these topics, see Bihn (1997) and Scherer (1997). Morasch (1994) shows that also in oligopoly models the incentive for R&D cooperations increases with increasing spillover effects. The incentive is even higher when the spillovers between members are higher than spillovers to non-members.

when they constantly have to fear the loss of their market shares.[9] This scenario is most likely in situations with a patent race, i.e. in situations where the firms face severe competition in the research stage.[10] There, the innovation success is a priori uncertain; the probability to succeed depends upon recent and past investment into R&D of all competitors. In the end, only one firm can succeed in implementing and launching its innovations, while the others have to suffer losses due to the R&D-investment expenditures. As a consequence, in models of patent races, threat of potential competition induces also the market leader to invest heavily into R&D. As a consequence, there may be a high incentive to internalise spillover effects within R&D-cooperations. On the other hand, there are some doubts whether firms, especially those that have been successful in building up high market power, actually have an incentive to cooperate in R&D. One reason is that high market power reflects efficient research and production. Therefore, this firm is not easily affected. Or, it may even fear to increase the probability of their competitors winning the race by passing on its internal research results via coordination. In these cases, spillovers may even hinder firms from cooperating in R&D.

3 Empirical Analysis

3.1 Data Source and Definition of the Variables

We analyse the empirical relevance of spillover effects for R&D-cooperations with the help of a cross-sectional analysis, based on the Mannheim Innovation Panel (MIP), the German contribution to the Community Innovations Survey (CIS). In the set of 1993 all relevant data have been available simultaneously.[11] With respect to the choice of the variables, the analysis comes as close as possible to the theoretical considerations. First of all, the relevant data set comprises only firms for which there are data on R&D-cooperations with direct competitors.[12] Thereby, we estimated the internalisation aspect of R&D-cooperations by the help of a logistic regression model, i.e., the incentive for R&D-cooperations has been approximated by the probability with which firms cooperage in

9 According to Mazzucato (1998) and Symeonides (1996:55) increasing concentration will not occur because of diminishing returns, even without spillover effects. Mazzucato (1998:66) calls it the problem of dynamic diseconomies of scale. The potential for further cost reduction due to innovations may shrink with increasing market share. In this case, there may be leapfrogging through small firms with drastic innovations, especially if they are able to exploit economies of scale- and scope by networking and clustering. According to Mazzucato, spillover effects increase the probability for this leapfrogging.

10 See also Katz and Ordover (1990). For an overview of models of patent races see Reinganum (1989).

11 For a description of the MIP see ZEW (1998).The argument of a lack of up-to-date data can be op-posed by the fact that there are no big changes to be expected within short run with regard to the gen-eral attitudes of firms towards spillover effects and R&D cooperations.

12 In contrast, the indicator "cooperation variety" in the analysis of König et al. (1994) measures only the extend to which spillovers influence the degree of variety in cooperation partners. It also includes co-operations with universities and research institutes, and thus does not reflect the competitive element of spillovers.

R&D. This is due to the fact that the available data set does not contain information about the actual number of R&D-cooperations, but only whether firms have been engaged in R&D-cooperations at all.

Secondly, in our analysis the capability to appropriate R&D-results is translated into the empires by the help of firm data on the effectiveness of various appropriation facilities. By running a factor analysis the overall number of appropriation facilities could be reduced to three main groups: official measures like patents and registered designs, firm-specific measures like time lead, complex product design and secrecy, and finally, low fluctuation of qualified personnel.[13] In our view, the theoretical definition is best approximated when there is no means for knowledge protection that firms judge as really effective.[14] Within the MIP, the effectiveness of the various appropriation facilities has been measured on a scale ranging from 1 to 5 points. We then measured lack of appropriation of one group of appropriation facility as one minus the ratio between the total score actually achieved and the maximum score that is achievable within the respective group of appropriation facility. For instance, if firms judge patents as very effective (score =5) and registered design as medium effective (score=3) the spillovers due to the patenting system takes the value 1-((5+3)/10). Finally, the indicator for overall spillovers is a weighted average of the spillovers due to the three main groups of protective facilities.

Table 1 gives an overview of spillover effects in various industries.[15] For instance, in the data processing and electronics industry some 30 per cent of research results cannot be appropriated on average, whereas about 51 per cent can not be appropriated through the patenting system. According to these descriptive results in table 1, the critical value of spillovers that resulted from the theoretical model of d'Aspremont and Jacquemin fits to the assessment of firms as long as it concerns the patent system. In contrast, it looks like as if firm-specific appropriation facilities are much more capable to protect knowledge as compared to the patenting system. Therefore, firms presumably use firm-specific appropriation facilities as a reaction on insufficient patent protection – similar to the basic assumption in König et al. However, the role as well as the effectiveness of the various appropriation facilities differ between industries. For instance, the result of Mansfield prevails again, where patenting was found to be relevant for pharmaceuticals, chemicals and mineral-oil processing, while they had only limited relevance in industries like office machinery.[16]

---

13 The terminology is similar to Harhoff (1997:348). The results of the fatcor analysis are given in the appendix.

14 In contrast, König et al. (1994) see spillovers arising whenever firms judge "firm-specific appropria-tion facilities" as effective in protecting internal research results, since – in their view - the latter di-rectly reflects the reaction of the firms to an ineffective patenting system. Despite of the appeal of this view, no high and significant negative correlation could be found between high effectiveness of firm specific appropriation facilities on the one hand and ineffective patenting system on the other hand.

15 The classification of the industries is given in the appendix.

16 Cited from Geroski (1995). With regard to appropriation conditions, and thus with respect to the relevance of spillover effects see also Symeonides (1996).

Table 1. Knowledge spillovers in manufacturing industries

	General	Type of appropriation facility		
		Patents	Firm-specific	Personnel
Mining	0,49	0,68	0,51	0,28
Wood products	0,37	0,59	0,33	0,19
Food and Textiles	0,35	0,57	0,30	0,18
Ceramics	0,36	0,55	0,31	0,21
Steel and Metal products	0,36	0,53	0,35	0,20
Precision/optical instruments	0,32	0,52	0,27	0,15
Rubber products	0,30	0,52	0,25	0,16
Electronics	0,33	0,51	0,30	0,16
Automobiles and aircraft	0,31	0,49	0,29	0,15
Chemicals	0,29	0,46	0,25	0,16
Machinery	0,30	0,46	0,28	0,17

Source: Mannheim Innovation Panel 1993.

With respect to market structure, we have chosen indicators following the concept of the "optimal intensity of competition". According to Kantzenbach[17] this is given within a wide oligopoly, i.e., a market that is characterised by a relatively small number of firms, thus with relatively high concentration, but that is open for potential competition. As indicators we use therefore firstly, the level of the Herfindahl-Index of 1993 measuring the absolute level of concentration, and secondly, the average rate and the range with which concentration as well as the number of firms within an industry have been changing since the end of the seventies, indicating intensified and potential competition. These are given in Monopolkommission (1996). Due to differences in the industrial classification, the Herfindahl-index that fits to the classification in the MIP is a weighted average of the Herfindahl-indices of the industries at three-digit-level. The weights are the number of firms of the respective industry on the three-digit-level relative to the total number of firms of the aggregated industry on the MIP-level. An additional indicator for potential competition is directly given within the MIP. There, firms have been asked to assess the previous and the expected intensity of competition within the output market where they are operating.

According to table 2, the criteria for the optimal intensity of competition are mostly fulfilled by precision and optical instruments and data processing and electronics. Very competitive industries are further on steel and metal products, wood and paper products, and rubber products. With respect to rubber products and data processing and electronics, the enormous change and variation in the number of firms indicate a market that is open for potential competition. In contrast, chemical industry, and automobiles and aircraft are

[17] Cited from Berg (1993: 245).

industries characterised by stable concentration, i.e., that are not very affected from (potential) competition.

Table 2. Concentration and Competition in Manufacturing

	Herfindahl-Index				Change in firm number			Con-cen-tra-tion	Com-peti-tion
	Level	Std. Dev	Change	Std. Dev	Std. Dev	Change	Std. Dev		
Machinery	2,2	46,5	-1,4	42,8	7,3	1,9	4,3	-	+
Food/textiles	4,8	9,6	-1,9	11,9	4,2	-0,5	4,2	-	-
Steel/metal	6,2	41,3	-11,1	14,8	6,0	1,0	5,1	-	+
Rubber products	6,9	16,1	-6,5	7,5	10,7	4,1	4,9	-	+
Wood/paper	8,0	12,3	4,5	12,2	6,0	-1,0	5,0	-	+
Ceramics	12,10	4,6	-1,5	5,0	9,7	-2,0	4,0	-	-
Precision/ optical	15,0	15,8	4,0	18,1	5,5	-1,3	3,4	+	+
Chemicals	33,9	6,4	-1,7	5,7	2,5	0,2	2,1	+	-
DP/electronics	60,6	7,9	1,7	12,4	12,0	4,1	5,1	+	+
Auto/aircraft	99,8	5,1	0,2	5,2	6,0	-1,3	4,6	+	-
Median	12,1	12,3	3,5 [a)]	11,9	6,0	1,7 [a)]	4,6		

a) Median of the absolute values; light shading stands for values below, dark shading for those above the mean value.
Source: Monopolkommission (1996).

Table 2 only describes the market structure on the German product markets. However, additionally intensive international competition may be expected in some industries, especially in electronics and data processing, and automobiles and aircraft. With respect to the latter, Hammes (1994) and Graves (1996) explain the competitive pressure with the success of Japanese firms. They are gaining ground due to their capacity to keep on producing new goods while they are at the same time continuously lowering production costs. According to Hammes, intensive competition in automobiles additionally results from the extremely high fixed costs for R&D in times of shrinking product life cycles. Thus, immense investment into R&D has to pay for itself within increasingly shorter time periods. Table 2 can also only indirectly account for competition on the research stage of the innovation process. Competitive scenarios that are similar to models of patent races however, may prevail in R&D-intensive industries like machinery, chemicals and especially automobiles and aircraft, data processing and electronics, and the precision and optical instruments industry.

The role of market structure for the internalisation effect of R&D-cooperations then has been analysed by looking at the effect that spillovers have - given they are interacted with market structure. The idea behind is that high incentives to cooperate in R&D may be expected whenever firms are faced with spillover effects as well as intensive competition.

In detail, the role of market structure is first of all, indirectly analysed by testing for differences in the probability for R&D-cooperations due to spillovers arising in industries that are characterized by intensive competition. Alternatively, we directly interacted the spillover variable with dummies for market structure. Here, due to correlations among the individual indicators for market structure which are sufficiently high such that they bias parameter estimation, but too low to allow an interpretable factor analysis, we have selected the Herfindahl-Index, the change in the number of firms and the intensity of expected competition, since they can be seen as best indicators for the above mentioned characteristics of the optimal intensity of competition.

However, table 1 together with table 2 let presume that spillovers may be interrelated with market structure. This raises the question of exogeneity of the explanatory variables. On the one hand, spillovers may intensify (potential) competition. Thus, the variables for market structure may not be seen as exogenous anymore, but are rather determined by spillovers. However, one can presume that spillover effects do not dramatically influence market structure within one single period. On the other hand, just the other way round, spillover effects themselves may be determined by market structure and may not be exogenous anymore, since firms that are faced with intensive competition may try to prevent spillovers from arising. This problem has been solved by explicitly taking into account that these firms may first choose among the different appropriation facility, especially those that are different from the patenting system, before they decide about participating in R&D-cooperations. Thus, we asked whether the fact that the level of spillovers depends on the type of appropriation facility accounts for different influences of spillovers on R&D-cooperations.[18] As a consequence, in the case where firms are able to protect knowledge through some mechanism spillovers have no or only low effects on R&D-cooperations, despite intensive competition; except, these alternative appropriation facilities are as ineffective as the patenting system.

Finally, we included several controlling variables that may determine the innovation process and thus, the propensity to cooperate in R&D. These are industry and firm-size dummies as well as impediments for innovations. With respect to the latter, by running a factor analysis we were able to reduce the various impediments for innovations into four main groups: first, factors like costs and risks related with innovation projects, second, government regulation or slow administration, third, lack of internal capital or insufficient access to external financing, and finally, what we called "linkages"; these comprise problems that arise due to lack of willingness to innovate within the individual firm or at the side of suppliers and customers, or that arise due to insufficient access to relevant information.

3.2 Empirical Results

Whether there is empirical evidence for the hypotheses above can be seen from the following results. Thereby, the first part focuses on possibly different influences of

---

[18] With this, we implicitly ask for the "true" spillovers for which the theoretical relationship between spillovers and R&D-cooperations is valid.

spillovers on R&D-cooperations, depending on the specific industries or on the degree of concentration and competition. The second part then asks for different influences of spillovers on R&D-cooperations taking into account that the level of spillover depends on the type of appropriation facility. Here again, we analyse both, spillovers in various industries as well as interacted with the indicators for market structure.

Spillovers, Market Structure and R&D-Cooperations at First Glance

In general, one can presume from the results in table 3 that the incentive to cooperate in R&D increases with the degree of competition when spillover effects arise. R&D-cooperations are strongly encouraged by spillover effects in machinery, in automobiles and aircraft, data processing and electronics, and in precision and optical instruments. Furthermore, highly positive and significant coefficients of spillovers prevail in the competitive industries machinery and especially in the precision and optical instruments industry, as well as when spillover effects are interacted with the intensity of competition. However, noticeably significant coefficients prevail also in industries that are – according to table 2 - not at all characterised by intensive competition, i.e., in the food and textiles industry, and in the chemical industry. Instead, spillovers in very competitive industries like rubber products, data processing and electronics, and automobiles and aircraft do not show significant influences.

There are mainly two reasons that may account for these results: Firstly, intensified competition may lead to a scenario where the competitive effect is too high to be counter-weighted by the reduction of the research cost within R&D-cooperations. In that case, the individual firm facing intensive competition may be induced to search for effective means other than the patenting system to protect knowledge. As a consequence, spillovers are low, and when firms do not cooperate each firm has – like Shy (1996: 233) describes it – "a lot to gain from R&D since under small spillover effects, the R&D intensifies the cost advantage of the firm that undertakes a higher level of R&D". From table 1 we have seen that firms that are not able to protect their research effectively through patents and registered designs may search for more effective appropriation facilities. Accordingly, insignificant coefficients let presume that spillovers do not arise because of effective firm specific appropriation facilities. Therefore, although spillovers may have a strong influence on R&D-cooperations whenever they arise in competitive industries, the coefficients are not significant due to the low number of firms that are not able to appropriate at all.

Table 3. Spillovers, Market Structure and R&D-Cooperations

Number of observations	616	616
Deviation [a)]	40,39	251,13
Correct Classification (in per cent)	60,71	79,22
Influencing variables [b) c)]	Exp(ß) [d)]	Exp(ß) [d)]
Industry-specific spillovers		
Food/textile	3,77	
Machinery	1,61	
Chemicals	2,59	
Automobiles/aircraft	1,17	
Data Processing/Electronics	1,42	
Rubber products	0,84	
Ceramics	0,82	
Steel/Metal products	0,46	
Precision/optical instruments	1,56	
Wood products	1,11	
Interaction spillovers – market structure		
Herfindahl		0,77
Change in firm number		0,73
Expected intensity of competition		1,43
Indicators for market structure		
Herfindahl	1,21	
Change in firm number	0,95	
Expected intensity of competition	1,03	
Firm Size		
Medium		0,42
Large		0,84
Impediments for innovation		
Linkages	0,85	0,91
Costs and risks	0,92	1,27
Government	1,04	1,13
Finance	0,89	0,82

a) Difference between the Loglikelihood measure of the estimation without the exogenous variables and the estimations where the exogenous variables are included. The larger the value the better the fit of the model.
b) To include firm-size or industry dummies in the first regression was not possible due to high correlations with some firm-specific spillovers. Instead market structure indicators are included.
c) In the second regression, additionally industry dummies have been included. They showed significant influences except for steel and metal products.
d) Dark shading is equivalent with a significance level of 95%, light shading with 90% significance.

Source: MIP 1993, Monopolkommission (1996), own calculations.

Second, insignificant but highly positive coefficients in the case of automobiles and aircraft, and data processing and electronics may be due to the fact that firms operating in these industries have generally a high incentive to cooperate in R&D. According to Hammes (1994), industries like electronics, chemicals, automobiles, and machinery show a high degree of cooperation anyway.[19] He justifies this with three properties characterising these industries: They use technology intensively, they have a global orientation, and firms in these industries are producing complex and differentiated products. One of the most convincing arguments in favour of cooperations is the flexibility that allows firms to handle changing and country-specific needs. The significant coefficient when costs or risks are hampering innovation may additionally speak in favour of this presumption.

The drastic influence of spillovers in the chemical industry together with the significant coefficient of the Herfindahl-Index for R&D-cooperation – without spillovers - reflect a result in Gerybadze et al. (1997) that is similar to the one just mentioned: They point at the increasing tendency of firms within the chemical industry to specialise on specific technologies within R&D. Especially dominant firms within the chemical industry have been intensively outsourcing R&D-units to producers that are specialised in the specific competencies. R&D-cooperations with firms that are competent in the specific field may provide access to up-to-date knowledge. Or, these cooperations enable firms to build up a common pool of qualified staff.

The Internalisation Effect Reconsidered

According to the results in table 4, the incentive to cooperate in R&D depends crucially on the type of appropriation facility that firms use to protect knowledge. These results make clear that the true spillovers, i.e., spillovers that are sufficiently high such that R&D-cooperation is profitable, only arise whenever also firm-specific appropriation facilities are not effective. This can most of all be seen from the strong and significant coefficients when spillovers arise due to lack of appropriation from firm-specific means as compared to coefficients of spillovers due to an ineffective patenting system. It looks like as if firms trade-off between the returns from "internalising" research by keeping it secret etc., and the return from internalising it by coordinating R&D with competitors.

This latter presumption seems to be even more relevant when the intensity of competition is taken into account. R&D-cooperations are significantly influenced when spillovers arise in machinery, rubber products, precision and optical instruments and – to some degree – also in data processing and electronics, thus, in industries that are characterised by intensive competition. Additionally, it is the interaction between spillovers on the one hand and the expected intensity of competition or the change in firm number on the other hand, that influences significantly the incentive for R&D-cooperations.

[19] Hammes (1994:216-7) analyses strategic alliances in general. His study therefore, goes beyond the topic at hand. However, his results can be applied here, too, since R&D can be seen as one of the three basic motivations for building strategic alliances, next to production and marketing. See here also Hagedoorn (1997).

Table 4. Industry-specific spillovers and R&D-cooperations – various appropriation facilities

Number of observations	620	622	639
Deviation	26,79	53,66	20,54
Correct Classification (in per cent)	59,35	62,54	60,72
Influencing variables [a)]	Exp(ß)	Exp(ß)	Exp(ß)
Industry-specific spillovers	Patents	Firm-Specific	Personnel
Food/textile	1,18	4,70	1,67
Machinery	1,64	1,76	0,99
Chemicals	1,67	3,03	1,23
Automobiles/aircraft	1,02	1,10	1,23
Data Processing/Electronics	1,20	1,25	1,69
Rubber products	0,67	2,53	0,67
Ceramics	0,92	1,22	0,62
Steel/Metal products	0,59	1,00	0,57
Precision/optical instruments	1,22	2,11	1,37
Wood products	1,12	1,64	0,79
Indicators for market structure			
Herfindahl	1,24	1,13	1,20
Change in firm number	0,98	0,93	0,96
Expected intensity of competition	1,03	1,03	1,02
Number of observations	620	622	639
Deviation	251,08	258,29	259,61
Correct Classification (in per cent)	79,03	78,46	79,19
Interaction spillovers - market structure			
Herfindahl-index	0,81	0,80	0,94
Change in firm number	0,92	0,87	0,72
Expected intensity of competition	1,08	1,65	1,27
Firm-Size			
Medium	0,42	0,41	0,44
Large	0,84	0,87	0,86
Impediments for innovation			
Linkages	0,92	0,93	0,93
Costs and risks	1,26	1,26	1,28
Government	1,13	1,15	1,17
Finance	0,85	0,83	0,85

a) In the first three regressions impediments for innovation had the same values, but have not been significant. In the second three regressions additionally industry dummies have been included. Coefficients were similar to the ones from the regressions in table 3.

Source: MIP 1993, own calculations.

Finally, especially the result in data processing and electronics, precision and optical instruments, and in rubber products point out that the trade-off between R&D-cooperations and alternative appropriation facilities becomes more relevant with increasing competition. Only when firms are not able to protect their research through appropriation facilities other than the patenting system spillover effects are sufficiently high such that the internalisation in cooperations outweigh the competitive effect from research which has been increased due to intensified competition. Therefore, until now, the results speak in favour of the presumption from the theoretical considerations: the threat of potential competition may induce also the market leader to invest heavily into R&D. As a consequence, there may be a high incentive to internalise spillover effects within R&D-cooperations.

However, there is also some evidence for the hypothesis that in industries that are characterised by high competition in the research stage, the coordination of research results is rather seen as a danger of the own market share. Most of all this can be seen from the relatively weak and insignificant coefficients for automobiles and aircraft, and for data processing and electronics as compared to the strong and significant coefficients of the other industries – as long as firm-specific appropriation facilities are considered. From the classification of industries in the appendix, it becomes clear that these industries include highly R&D-intensive branches. This is aircraft in automobiles, and it is manufacturing data processing machines and television and communication equipment in electronics. In these branches, high pressure from competition in the research stage can be expected. Together with the theoretical considerations above it may therefore be presumed that firms from electronics and from automobiles and aircraft do not see R&D-cooperations as a method to internalise spillover effects. Rather they may judge R&D-cooperations as an additional channel through which internal knowledge can spill over to other firms.

4 Conclusion

The empirical results support in general the presumption that with intensified competition also the influence of spillovers on R&D-cooperation increases. Strong and significant influences of spillovers show up in competitive industries – most of all in precision and optical instruments, as well as when spillovers are interacted with indicators for the intensity of competition. However, competition seems to induce firms to trade off between alternative appropriation facilities and R&D-cooperations. True spillovers, i.e., spillovers that are sufficiently high such that the internalisation effect from cooperation more than outweighs the competitive effect from research, only arise whenever firms are not able to protect their research results through any appropriation facility. This can be seen from strong and significant coefficients whenever firms are not able to protect their knowledge through firm-specific appropriation facilities as compared to relatively weak and insignificant coefficients in the case of an ineffective patenting system.

There is also some evidence that spillover effects may even hinder firms from cooperating with other firms in R&D. Most of all this can be seen from the relatively weak and insignificant coefficients for automobiles and aircraft, and for data processing and electronics, as compared to the strong and significant coefficients of the other industries –

as long as firm-specific appropriation facilities are considered. What precisely supports this presumption is the fact that in both industries there are branches where high pressure from potential competition on the research stage may be expected. In the case of electronics this is manufacturing data processing machines and equipment. In the case of automobiles etc. this is aircraft and spacecraft. Taking this together with the positive overall influence of spillovers in these industries one conclusion may be drawn which is different from the general theoretical hypothesis: If there is the threat of existing or potential competition on both stages of the decision process, the relationship between spillover effects and R&D-cooperations can no longer be expected to be linear. Rather, it may be backward sloping when there is already high pressure from competition on the research stage.

These empirical results let presume that one crucial assumption within the basic theoretical models of the relationship between spillovers and R&D-cooperations may not be fulfilled: In contrast to these models, spillovers may not simply arise proportionally to the level of R&D produced within the economy; rather, they are endogenous; they are influenced by each firm's ability and decision to protect own and to use external knowledge; and the degree of (potential) competition seems to play a crucial role in this decision. Up to now, this results from a very simple empirical analysis. In future work, the relationship between spillovers and R&D-cooperations will have to take into account more explicitly – theoretically as well as empirically - the interdependencies between the level and the effects of spillovers and the intensity of competition or market structure in general.

References

Berg, H. (1993) Wettbewerbspolitik, in: Vahlens Kompendium der Wirtschaftstheorie und Wirtschaftspolitik, Munich:Vahlen.

Bihn, M. (1997) Forschungs- und Entwicklungskooperationen - eine Wettbewerbstheoretische und Wettbewerbspolitische Untersuchung; Frankfurt am Main: Europäische Hochschulschriften, Reihe V, Bd. 2078.

Coe, D. T., Helpman, E (1995) "International R&D Spillovers", European Economic Review 39: 859-87.

D'Aspremont, C., Jacquemin, A. (1988) "Cooperative and Noncooperative R&D in Duopoly with Spillovers", American Economic Review 78(5):1133-7.

Fahrmeir, L., Hamerle, A., Tutz, G. (eds) (1996) Multivariate Statistische Verfahren, Berlin: de Gruyter.

Geroski, P. (1995) "Markets for Technology: Knowledge, Innovation and Appropriability", in: Stoneman, P. (ed.) Handbook of the Economics of Innovation and Technological Change, Oxford, UK: Blackwell.

Gerybadze, A., Mayer-Krahmer, F., Reger, G. (1997) Globales Management von Forschung und Innovation, Stuttgart: Schäffer-Poeschel.

Graves, A. (1996) "Innovation in a Globalizing Industry: The Case of Automobiles", in: Dodgson, M. and Rothwell, R. (eds): The Handbook of Industrial Innovation, Cheltenham: Edward Elgar.

Grossman, G., M., Helpman, E. (1991) Innovation and Growth in the Global Economy, Cambridge, Massachusetts: MIT Press.

Grupp, H., Gehrke, B. (1994) Innovationspotential und Hochtechnologie: Technologische Position Deutschlands im Internationalen Wettbewerb; Heidelberg: Physika Verlag.

Gujarati, D. (1995) Basic Econometrics (3rd edition), New York: McGraw-Hill, International.

Hagedoorn, J. (1997) "The Economics of Cooperation among High-Tech Firms - Trends and Patterns in Strategic Technology-Partnering since the Early Seventies", in: Koopmann, G., Scharrer, H.-E. (eds) The Economics of High-Technology Com-petition and Cooperation in Global Markets, Baden-Baden: Nomos.

Hammes, W. (1994) Strategische Allianzen als Instrument der strategischen Unternehmensführung, Wiesbaden: Deutscher Universitätsverlag.

Harhoff, D. (1997) "Innovationsverhalten in einem strukturellen Oligopolmodell", Zeitschrift für Wirtschafts- und Sozialwissenschaften 117:333-64.

Kamien, M.I., Muller, E., Zang, I. (1992) "Research Joint Ventures and R&D Cartels", American Economic Review 82:1293-306.

Katz, M. Ordover, J. (1990) "R&D Cooperation and Competition", Brookings Papers of Economic Activity: Microeconomics, 137-203.

Klodt, H. (1995) Grundlagen der Forschungs- und Technologiepolitik, München: Vahlen.

König, H., Licht, G., Staat, M. (1994) "F&E-Kooperationen und Innovationstätigkeit", in: Gahlen, B., Hesse, H., Ramser, H-J. (eds): Europäische Integrationsprobleme aus wirtschaftswissenschaftlicher Sicht, Tübingen: Mohr.

Konrad, E. (1997) FuE-Kooperationen und Internationale Wettbewerbsfähigkeit, Wiesbaden: Gabler.

Krugman, P.R. (1992) "New Thinking about Trade Policy", in: Krugman (ed.): Strategic Trade Policy and the New International Economics, Cambridge, Massachusetts: MIT Press.

Leahy, D., Neary, J.P. (1997) R&D Spillovers and the Case for Industrial Policy in an Open Economy, Centre for Economic Performance Discussion Paper, London.

Mazzucato, M. (1998) "A Computational Model of Economies of Scale and Market Share Instability", Structural Change and Economic Dynamics 9:55-83.

Monopolkommission (1996) Wettbewerbspolitik in Zeiten des Umbruchs, Hauptgutachten 1994/1995 – Anlagenband, Heidelberg, Nomos.

Morasch, K (1994) Strategische Allianzen, Anreize-Gestaltung-Wirkungen, Heidelberg: Physica-Verlag.

Reinganum (1989) The Timing of Innovation: Research, Development, and Diffusion, in: Schmalensee, R., Willig, R.D. (eds): Handbook of Industrial Organization, Elsevier Publishers.

Scherer, N. (1997) Kooperationsentscheidungen in Forschung und Entwicklung, Frankfurt am Main: Europäische Hochschulschriften, Reihe 5, No. 1798.

Shy, O. (1996) Industrial Organization – Theory and Applications, Cambridge:MIT-Press.

Suzumura, K. (1992) "Cooperative and Noncooperative R&D in an Oligopoly with Spillovers", American Economic Review, 82(5):1307-20.

Symeonides, G. (1996) "Innovation, Firm Size and Market Structure: Schumpeterian Hpotheses and Some New Themes", OECD Economic Studies, No.27, 1996/II.

Tirole, J. (1993) The Theory of Industrial Organisation, Cambridge, Massachusetts: MIT Press.

United Nations Industrial Development Organization (UNIDO) (1997) International Yearbook of Industrial Statistics, Vienna: Edward Elgar.

Zentrum für Europäische Wirtschaftsforschung (ZEW) (1998): Das Mannheimer Innovationspanel (MIP) 1993 - data and description, corrected 1998.

Appendix

Table A1. Classification of Industries

Terminology in the text	Industry according to WZ 93 and NACE-Rev.1	2 digits accord. to WZ 93	ISIC-Classif [a)]
Mining	Mining, minerals, energy and water supply	10-14, 40, 41	0
Food Products/Textiles	Food manufacturing, Tobacco manufactures, textiles and wearing apparel	15, 16, 17, 18, 19	0
Wood Products	Manufacture of wood and paper products, printing and publishing; manufacture of furniture, jewellery, music instruments, sporting goods and manufacturing industries not classified elsewhere (n.e.c.)	20, 21, 22.2, 22.3, 36	0
Chemicals	Chemical industry, mineral oil processing manufacture of coal	24, 23	1 0
Rubber Products	Manufacture of rubber and plastics	25	0
Ceramics	Manufacture of glass, pottery and earthenware	26	0
Metal Products	Manufacture of fabricated metal products	27	0
Steel Processing	Iron and steel basic industries	28	0
Machinery	Manufacture of machinery, weapons; electrical appliances and houseware n.e.c.	29	1
Electronics	Manufacture of office, computing and accounting machinery, radio, television and communication equipment Manufacture of electricity distribution and control apparatus etc.	30, 31, 32	2 1
Precision/Optical Instruments	Manufacture of medical appliances, appliances for measuring, checking etc., Optical instruments, photographic equipment	33	2 1
Automobiles/Aircraft	Manufacture of aircraft and spacecraft, Manufacture of motor vehicles, parts and accessoires, Manufacture of transport equipment, n.e.c.,	34, 35	2 1 0
Construction	Construction	45.2	0
Services [b)]	Data processing a. database, research and development, technical, physical a. chemical services, Architecture- a. engineering, Recycling, metal waste and scrap	72, 73, 74.2, 74.3, 90	2 1 0

a) ISIC-Classification, "0" represents non R&D-intensive industries, "1": high-level technologies, "2": R&D-intensive industries.

b) Classification following the ISIC/SITC-Classification [Grupp and Gehrke (1994)]

Source: ZEW (1998), ISIC-Classification, UNIDO (1997).

Table A2. Results of the Factor Analysis "Appropriation Facilities"

	Main Factors [a]		
Facilities	Patents	Firm-Specific	Personnel
Registered product design	0,754		
Secrecy product innovation		0,735	
Complex product design		0,667	
Patents product innovation	0,819		
Qualified personnel product innovation			0,800
Time lead product innovation		0,598	
Other rights	0,801		
Secrecy process innovation		0,725	
Complex process design		0,699	
Patents process innovation	0,831		
Qualified personnel process innovation			0,886
Time lead process innovation		0,574	

a) Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalisation.
Source: Mannheim Innovation Panel (1993), own calculations.

Table A3. Results of the Factor Analysis "Impediments for Innovation"

	Main Factors [a]			
Impediments	Costs/Risks	Regulation	Finance	Linkages
High risks of innovation	0,538			
Impediments within the firm				0,446
Lack of info about technology level				0,696
Lack of info marketing				0,703
Lack of info external know-how				0,787
Lack of cooperation with firms				0,566
Lack of cooperation with universities				0,519
Restrictive regulations		0,820		
Bureaucracy		0,858		
Lack of tax incentives		0,485		
Lack of innovation at customers				0,330
Control of innovation costs	0,506			
Lack of innovation at suppliers				0,372
Market not yet capable				0,332
Lack of internal capital			0,972	
Lack of outside capital			0,721	
High costs of innovation	0,796			
Amortisation of costs	0,787			
Imitation	0,479			
Lack of qualified personnel				0,384

a) Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalisation.
Source: Mannheim Innovation Panel (1993), own calculations.

Chapter IV

Labour Market Analysis and Consumer Surveys

Exploring Unemployment in the European Union

Marcella Corsi

Abstract

The aim of this paper[1] is to explore EU unemployment by using survey data stemming from the ad hoc labour market survey carried out, among employers and employees, in 1999.

At first, particular attention is paid to the duration of unemployment and the occupational qualification of unemployed workers, in order to shed some light about the implications of labour mismatch and technological "skill-bias" in terms of structural unemployment. Secondly, the paper focuses on the issue of flexibility as a way to reduce unemployment, providing information on employees' readiness to work under non-standard contracts.

"Unemployment is the most serious and urgent problem facing the European Union. (..) It is degrading and demeaning for the unemployed and with damaging long run consequences, especially for the young that represent, in most countries, the bulk of the unemployment. And it is also a source of dangerous social tensions"

[Modigliani, F. et al. (1998), pp. 328-9]

1 Introduction

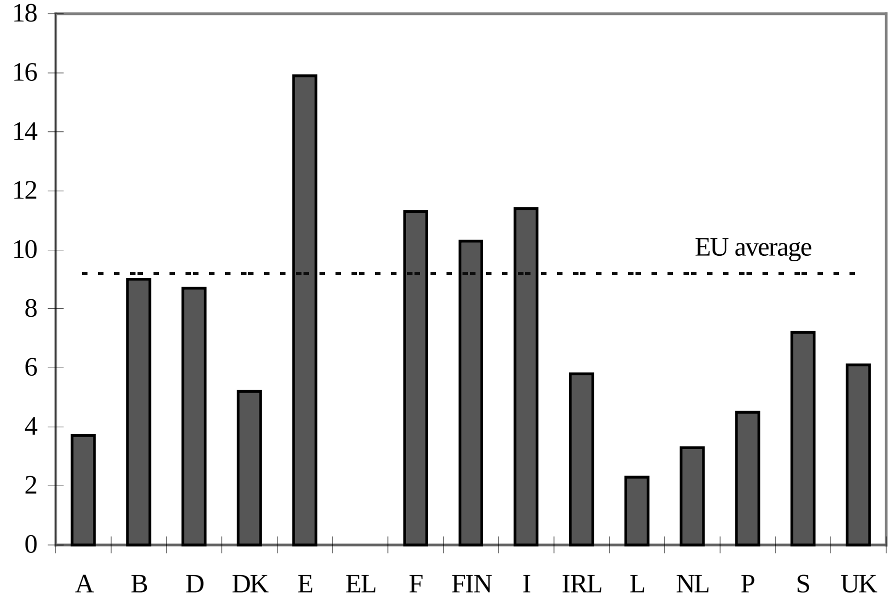
During the 1990s, unemployment in the European Union has been high, increasing and very unequal across member states and their regions. In 1999, the average unemployment rate was 9.2% (4.2% in the US, 4.7% in Japan); seven member states had an unemployment rate at or below 6% (Austria, Denmark, Ireland, Luxembourg, the Netherlands, Portugal and the United Kingdom) but three large member countries (France, Italy and Spain) had an unemployment rate much above the EU average (figure 1).

---

1 The author is grateful to Franz-Josef Klein (European Commission, DG ECFIN) for his help in dealing with the results of the 1999 ad hoc labour market survey. Financial support from MURST (project "Crescita e inflazione nell'UME: la politica monetaria, i mercati finanziari e il mercato del lavoro") is gratefully acknowledged.

The aim of this paper is to explore EU unemployment by using survey data stemming from the ad hoc labour market survey carried out, among employers and employees, in 1999. The survey partly overlaps with those carried out in 1985/86, in 1989 and in 1994.[2]

Figure 1. EU - Standardised unemployment rates (1999, % of total labour force)



Source: OECD

At first, survey results help to analyse the scale and duration of unemployment taking into account the level of training and qualification of the unemployed, their age and gender. Secondly, survey data give some hints about employers' employment plans (in industry, services and retail trade) and the kind of labour they look for. Thirdly, survey results allow to study employees' attitude in terms of external flexibility and to discuss the incidence of part-time contracts as a contribution to job creation in the European Union.

Such a precious information is analysed bearing in mind a simple idea: to achieve a lower unemployment rate (i.e., a higher employment rate) in the EU it is crucial to generate, over an extended period of time, economic growth well above the rate coming from increases in labour productivity in the overall economy.[3] In this sense particular attention must be drawn to technological change - as the main source of wealth and

[2] Cf. European Commission (1986, 1991 and 1995).

[3] Cf. Corsi – Roncaglia (2000), Sylos Labini (1999) and Modigliani et al. (1998) for detailed policy recommendations.

improvement in the quality of living standards over the long run - and the overall implications of the connected organisational changes and labour modifications.

2 Scale and Duration of Unemployment

The ad hoc labour market survey among employees provides useful information about unemployment in the European Union, especially in terms of duration. Among the employees who have been interviewed (either wage and salary earners or the unemployed), 30% have stated that they have been unemployed for at least four weeks since 1994 (27% of men and 34% of women). The percentage of young people (below 30 years) is much higher (42%) than that of other age groups, although, compared with the results of the 1994 survey, it has slightly decreased (it was 44% in 1994).

As table 1 shows, the scaling-up of unemployment – highlighted by the comparison with the 1989 and 1994 figures – concerns mainly Germany and Greece; the rise of unemployment in these two countries seems to have hit almost equally men and women, but older employees have been affected more significantly than other age groups. During the 1990s, female unemployment has dropped significantly in Portugal, Ireland, Italy and the Netherlands, but has increased, unexpectedly, in Denmark; youth unemployment has been steadily high all over the Union and has markedly decreased only in countries like Ireland and the Netherlands.

Half of the employees who have been unemployed for at least four weeks, between 1994 and 1999, were unemployed only once; such a percentage has decreased (it was 55% in 1994) especially for women (56% in 1994), who have been less affected by cyclical unemployment than in the past. By contrast, figures for multiple unemployment (three times or more) show that women and middle-aged workers have been worst and increasingly affected by such a phenomenon.

As figure 2 shows, multiple unemployment seems to affect mainly countries characterised by high rates of unemployment, like Finland, Greece, Italy and Spain. On the contrary, countries with low unemployment – i.e., Austria, Denmark, Ireland, the Netherlands, Portugal and the United Kingdom – register a high incidence of employees who have been unemployed only once.

Multiple unemployment is strictly connected with long-term unemployment that is more and more endemic in the European Union: the percentage of employees who have been unemployed for more than 12 months has risen from 44% in 1994 to 57% in 1999. As table 2 shows, long-term unemployment has a strong incidence among women and as far as the highest age group of workers is concerned.

Nevertheless, there is a great deal of variation among EU economies. Long-term unemployment is particularly a problem in Italy (79%) and Spain (59%), where it concerns mainly women and young employees also affected by multiple unemployment. But is also endemic in countries like Portugal (55%), Ireland (53%) and the United Kingdom (53%), which have managed to reduce significantly the overall incidence of unemployment in the last five years.

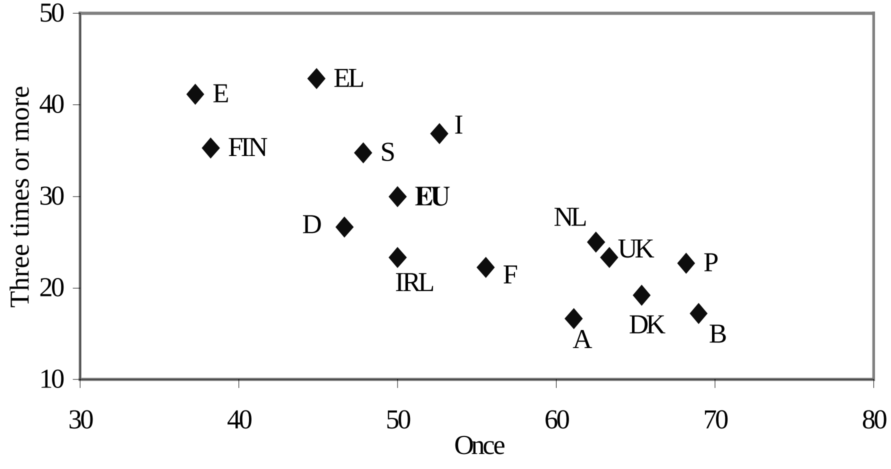
Table 1. Employees who have been unemployed for four weeks or more (% of replies)

	B	DK	D	EL	E	F	IRL	I	NL	P	UK
						Total					
1999	29	26	30	49	51	36	30	19	16	22	30
1994	29	28	22	39	52	38	43	25	13	36	37
1989	33	28	17	39	55	32	41	27	36	25	..
						Men					
1999	25	20	27	44	50	29	29	17	15	18	34
1994	23	26	19	37	49	37	41	21	14	30	39
1989	30	28	14	36	52	24	39	24	30	23	..
						Women					
1999	33	32	34	54	52	44	31	23	18	27	26
1994	36	30	26	43	56	39	46	30	13	44	33
1989	37	18	21	45	59	43	45	42	29	72	..
						Younger than 30					
1999	41	41	33	65	64	52	37	38	26	33	46
1994	41	42	26	58	69	59	50	42	21	46	47
1989	46	39	21	65	61	55	50	49	50	33	..
						30 - 49					
1999	26	26	28	38	45	32	25	15	14	17	22
1994	25	24	20	31	46	29	42	21	14	32	32
1989	28	27	17	32	44	21	37	24	28	17	..
						Older than 49					
1999	21	20	32	46	31	24	26	10	12	19	23
1994	13	17	20	28	25	34	36	20	3	19	30
1989	22	14	11	20	45	10	30	16	11	9	..

Source: EU ad hoc labour market survey

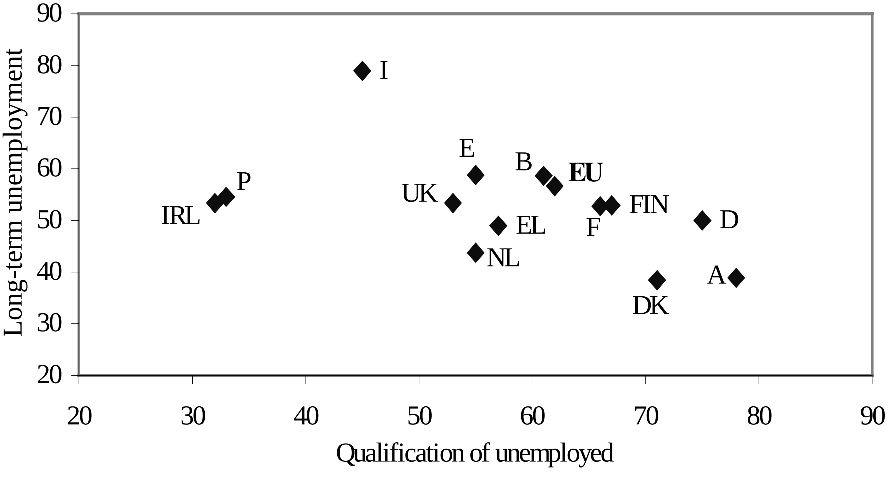
According to survey results, the common factor behind all countries characterised by long-term unemployment is a low level of qualification of the unemployed (figure 3). Indeed, a lack of qualification heightens the risk of unemployment. While 40% of the unemployed have no occupational qualification, this is true for only 29% of the employed workers. In particular, about two thirds of the unemployed in Portugal and Ireland declare not to have any qualification; in contrast, this share is below 30% in Austria, Germany and Denmark.

Figure 2. **Number of periods of unemployment (% of employees who have been unemployed for four weeks or more)**



Source: EU ad hoc labour market survey

Figure 3. **Long-term unemployment (% of employees who have been unemployed for four weeks or more)**



Source: EU ad hoc labour market survey.

Table 2. Duration of unemployment (% of employees who have been unemployed for four weeks or more)

	Total		Men		Women		Younger than 30		30 - 49		Older than 49	
	1999	1994	1999	1994	1999	1994	1999	1994	1999	1994	1999	1994
Less than 3 months	13	17	15	17	12	18	17	19	11	16	9	19
3-6 months	20	21	22	22	15	19	21	23	15	20	13	15
7-11 months	10	16	15	17	9	14	12	18	11	15	9	11
12-24 months	23	19	22	20	24	18	21	18	22	21	22	21
More than 24 months	33	25	26	21	41	28	29	19	33	27	48	32
Long-term unemployment	*57*	*44*	*48*	*41*	*65*	*46*	*50*	*37*	*56*	*48*	*70*	*53*

Source: EU ad hoc labour market survey

The disparity in qualification between men and women is one of the reasons why women are more threatened by unemployment. Survey results indicate that only 67% of women have received some form of training, the figure for men being 74% (see table 3). Taking the breakdown by age group, the figures show that the proportion of employees who have received training increases until they are in their 40s; more on-the-job training plays an important role in this sense.

Table 3. Occupational qualification (% of positive answers)

	Men		Women		Total	
	Unemployed	Employed	Unemployed	Employed	Unemployed	Employed
B	68	79	56	73	61	76
DK	90	83	63	82	71	82
D	74	89	75	85	75	87
EL	57	76	57	67	57	71
E	56	65	51	60	55	64
F	74	74	62	65	66	70
IRL	22	35	52	30	32	33
I	57	63	33	53	45	59
NL	44	65	62	66	55	65
A	48	91	92	77	78	86
P	35	42	30	42	33	41
FIN	64	82	69	79	67	81
S	..	98	..	99	..	98
UK	51	68	58	51	53	60
EU	63	74	59	67	62	71

Source: EU ad hoc labour market survey.

3 Technological Unemployment and Professional Mismatch

In recent years, economists have devoted much attention to technological unemployment, i.e. the type of unemployment that owes its origins to labour-saving technical change. However, such a form of technical change saves labour per unit of output, thus it creates unemployment only if output increases less than productivity. Thus, it is misleading to state that technical change necessarily generates unemployment.

At the outset, the relationship between technology, productivity and job creation, while controversial, appears straightforward at least from a macroeconomic perspective. Either the introduction of new technologies leads to more efficient production processes, reducing costs by saving on factors of production, or it leads directly to the development of new products that generate new demand. In either case, more welfare is created: in the first, through more efficient production combinations that liberate input resources; in the second, by satisfying new wants. The extent to which this higher welfare or increased productivity feeds back into job creation depends on the extent to which firms translate productivity gains into lower prices and new investment, and consumers respond to lower prices in terms of greater demand. The job losses that often follow the introduction of new labour-saving (rationalisation) processes, for example, may be compensated by the job creation associated with the output growth following the decline in prices, by additional job creation in other sectors, particularly the services sector, and by the possible substitution of labour to capital following the consequent downward labour costs adjustment.

In this context, the ad hoc labour market survey among employers provides information about the impact of technical change on job creation, testing both the importance of the introduction of new technologies as a reason for increasing the total number of employees in industry, services and retail trade, and the relevance of rationalisation as a reason for reducing employment in the same sectors.

As table 4 shows, the effect of technical change in terms of job creation differs according to the sectors considered. Labour-saving technologies are still diffused in industry while technical change mainly brings to a rise of employment in services and retail trade. As an overall effect - taking into account all sectors and both kinds of effect - technical change may have, according to interviewed employers, a positive effect in terms of job creation.

Although not very significant in statistical terms, a kind of negative (inconclusive) relationship between technical change and unemployment is depicted in figure 4. EU Member countries characterised by low unemployment rates (i.e., rates below EU average) normally attribute a more prominent role to technical change as a reason for increasing employment than countries with high rates of unemployment. This is particularly true when countries like Sweden and Spain or Ireland and Italy are compared on opposite sides.

As an overall effect, the ad hoc labour market survey results seem to support the idea that new technologies do not bring necessarily to a rise of unemployment. However, technological change may greatly alter the demand for worker skills. As new technologies become more widespread, certain skills may be less in demand - because many tasks once carried out manually are now performed by automated equipment - while the demand for workers able to maintain, program, and develop these sophisticated technologies rise.

Table 4. Technical change and job creation (coefficient of importance*)

	Industry			Services			Retail trade		
	1	2	(1-2)	1	2	(1-2)	1	2	(1-2)
B	20	28	-8	26	5	21	35	8	27
DK	9	35	-26	..	..	..	13	4	9
D	11	19	-8	25	6	19	2	12	-10
EL	23	32	-9	32	8	24	41	0	41
E	17	11	6	13	0	13	5	0	5
F	11	14	-3	16	6	10	9	5	4
IRL	15	70	-55	20	1	19	19	0	19
I	11	26	-15	6	0	6	6	4	2
L	14	22	-8	..	..	..	..	..	..
NL	29	7	22	..	..	..	15	4	11
A	13	14	-1	25	5	20	4	9	-5
P	26	32	-6	13	10	3	14	8	6
FIN	24	21	3	68	4	64	..	..	..
S	17	32	-15	43	9	34	24	7	17
UK	13	20	-7	18	2	16	15	6	9
EU	13	21	-8	19	4	15	10	6	4

Source: EU ad hoc labour market survey.
1) "Introduction of new technologies" as a reason for increasing the total number of employees.
2) "Rationalisation" as a reason for decreasing the total number of employees.
* The coefficient of importance ranks replies from 0 (all companies consider the reason to be "not so important") to 100 (all companies consider the reason to be "very important").

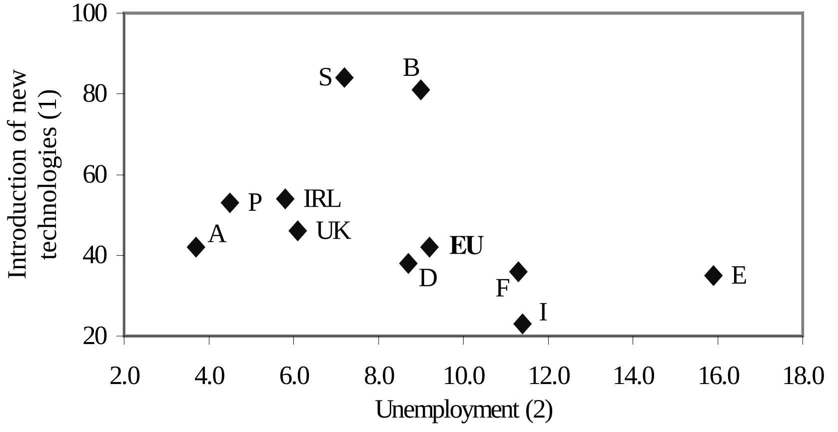
There is therefore concern that technological change may cause unemployment as the result of a mismatch between the demand for labour and the various skills of workers; in this way it may also increase the polarisation of society by widening the gap in income and employment opportunities between those whose skills have been displaced by new technology and those who create and use it.

As in 1994, the ad hoc labour market survey among employers seeks to gather Union-wide data on the skill structure of employment.[4] Taking the average for the European Union, skilled employees amount to 72% of total employment in industry, 78% in services and 70% in the retail trade sector.

In industry an 'upskilling' trend shows up when comparing the 1999 data for EUR-11 with past survey results: the percentage of skilled employees has increased 11 percentage points in the decade 1989-99, and 5 p.p. in the last five years (see figure 5).

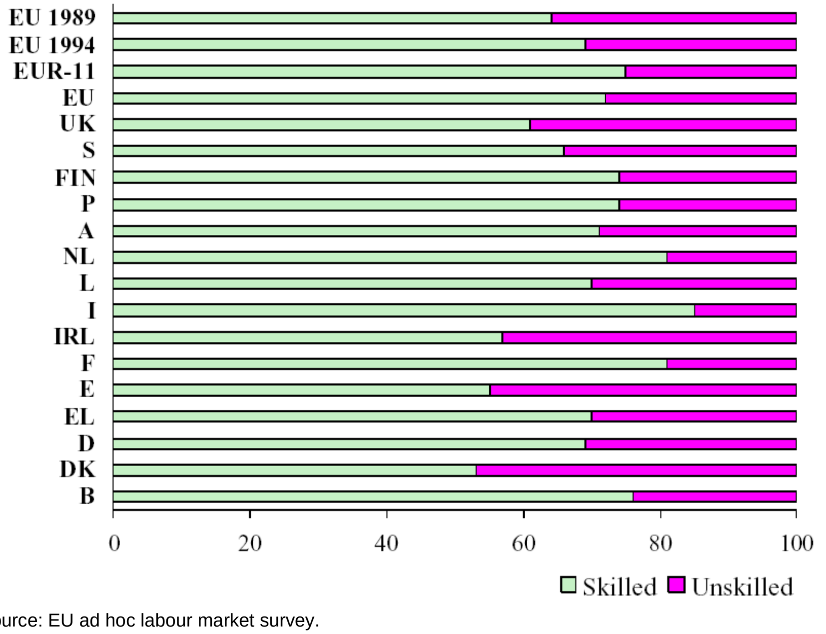
4 As in the past, these data must be interpreted with care since they rely in part on self-estimation.

Figure 4. New technologies and unemployment



Source: EU ad hoc labour market survey; OECD. 1) Global coefficient of importance taking account of all sectors under consideration. 2) Standardised unemployment rates.

Figure 5. Skill structure of industrial employment (%)



Source: EU ad hoc labour market survey.

Table 5. Prospective trend of employment (balances)

	Services		Retail trade		Industry	
	Skilled	Unskilled	Skilled	Unskilled	Skilled	Unskilled
B	50	-3	53	45	16	-40
DK	..	..	12	5	-12	-38
D	22	-9	-7	-20	15	-37
EL	32	9	71	55	-3	-9
E	26	6	6	-1	29	-14
F	19	5	21	12	11	-21
IRL	19	10	19	16	27	7
I	37	23	11	48	-30	-31
L	..	..	..	..	1	-45
NL	..	..	..	..	51	10
A	33	0	27	6	21	-19
P	15	8	4	-4	-4	-18
FIN	79	-36	10	9	24	-27
S	70	0	36	6	13	-26
UK	19	3	29	22	0	-26
EU	27	3	15	13	6	-27
EUR-11	27	3	11	11	8	-27
EU 1994	..	..	7	-2	-17	-33

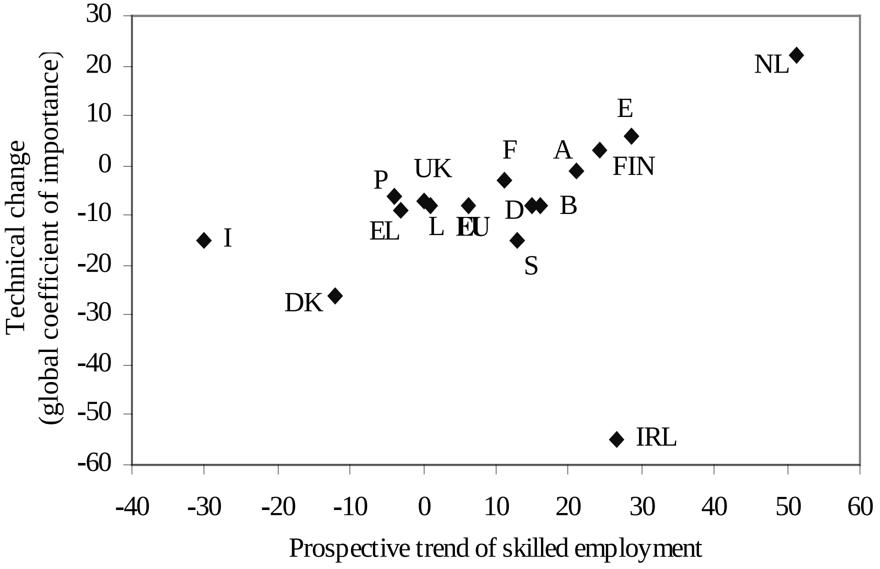
Source: EU ad hoc labour market survey.

Concerning the latest results, in Spain, Denmark, Ireland and the United Kingdom industry employs relatively few skilled workers compared with the EU average and with the average national skill level. On the other hand, the proportion of skilled workers in industry is particularly high in Italy (85%), France (81%) and the Netherlands (81%).

According to the survey results, entrepreneurs plan on balance to employ more skilled workers over the next two years, predominantly in the service sector. Moreover, a general improvement in employment plans can be observed, both in industry and in retail trade, comparing recent balances with the previous ones.

In all Member countries, except Ireland, the prospective trend of skilled employment in industry seems to be connected with the employers' view on the impact of technical change in terms of job creation, taking into account both rationalisation as a reason for decreasing the number of employees and the introduction of new technologies as a reason for increasing employment (see figure 6). The importance given to technical change by the industrial employers is positively related with the prospective trend of skilled employment as in countries characterised by 'booming' expectations, like the Netherlands, as in countries where skilled employment is deemed to decrease significantly in the next two years, like Italy and Denmark. Such a result underlines again that the lack of occupational qualification considerably increases the risk for workers of becoming or remaining unemployed, especially in view of the skill-bias induced by technical change.

Figure 6. Technical change and skilled employment in industry (%)



4 Flexibility as a Way to Reduce Unemployment

Technology provides new opportunities for expanding the range of goods and services, increasing productivity and increasing employment, but the organisation of firms and the institutional context for the introduction of organisational change determines the effectiveness and impact of the adoption of new technologies.

Indeed, a new work organisation is complementary to the adoption of new technologies in seeking greater functional and numerical flexibility (see Box 1). Employment flexibility may be achieved via two strategies: "internally" through multi-skilling, increased horizontal communication, use of better-trained and more responsive employees, etc., or "externally" by making employment contracts more flexible, for example by employing more workers on "non-standard" contracts. In what follows, the analysis focuses on "external" flexibility, not least because it has provoked debate about the social desirability of the increase of "precarious" jobs.

Non-standard forms of employment (e.g., part-time employment) are often considered to be an indicator of "external" labour market flexibility, since, compared to the norm of full-time employment, they imply weaker employment contracts, in terms of the greater ease of varying hours worked and terminating the contract.

BOX 1. Functional and numerical flexibility

Much of the literature on organisational change opposes two different kinds of flexibility (see OECD 1996).

- *Functional flexibility* usually involves high skill and collaborative approaches to work based on high quality labour inputs. Its most common features are: shifting job design and job boundaries away from traditional narrow ones, mobility across tasks, multi-skilling and wide-skilling, extensive training and retraining. Autonomous self-managed multi-functional teamwork is an indicator of this kind of flexibility.
- *Numerical flexibility* usually involves changing a quantity of labour input. These quantitative changes include, for example, numbers of employees, hours of work, use of part-time employees, use of temporary employees whose contracts can be terminated, making use of liberal provisions on hiring and firing.

There is also a spatial dimension to adjustment.

- *Internal flexibility* refers to operations carried out within the enterprise or performed within the existing contract structure of the enterprise.
- *External flexibility* involves interaction in markets, generally outside of the firm; it usually involves changing the nature and type of contracts.

There has been much overlap between internal and functional flexibility, and external and numerical flexibility.

Enterprises pursuing more adaptable organisation and production can shift the mix of functional and numerical flexibility in order to adjust labour use. Beyond that, they can also shift the locus of adjustment, moving it outside if they have traditionally relied on internal mechanisms for adjustment, and vice versa.

Firms can rely on (internal) variations in working hours instead of (external) hiring and firing to achieve numerical flexibility, for example, and on outsourcing of certain activities, instead of internal occupational restructuring, to acquire certain specialised competencies. Thus, countries that typically have greater recourse to external markets and strategies that depend largely on numerical flexibility have shown greater interest in making more functional and numerical adjustments within the firm. Interest in the "flexible" work practices in countries like the United Kingdom and US is largely about how to build functional flexibility and improve the quality of labour and other inputs. In the more protected European setting, the shift has been towards greater use of numerical flexibility and external functional adjustments.

Various demand-side motivations are commonly invoked for using these forms of employment. One is the lower overhead costs than those incurred for regular employees, because various non-wage labour costs are avoided. Second, greater flexibility reduces costs by providing a closer match between the required use and the workers employed.

Third, these forms of employment may also be used as mechanisms for screening employees.

On the supply side, engagement in non-standard employment contracts may be seen as either "voluntary", if the contract accords well with the employee's preferences, or "involuntary", if the employee has taken the contract owing to a lack of preferred alternatives.

Table 6. Incidence of part-time work (% of employed staff)

	Industry			Services			Retail trade		
	Total	M	F	Total	M	F	Total	M	F
B	4	1	17	13	4	9	35	8	62
DK	4	1	10	..	..	..	37	28	44
D	6	1	20	17	4	14	37	11	52
EL	1	0	3	9	3	6	33	18	43
E	2	1	5	11	4	7	14	10	20
F	4	3	7	26	6	19	35	15	48
IRL	5	2	11	11	4	7	42	27	51
I	2	0	8	21	3	17	29	7	45
L	1	0	8	..	..	..	..	..	..
NL	17	4	52	..	..	..	37	21	52
A	5	1	17	26	5	21	26	12	37
P	1	0	0	25	3	23	16	6	27
FIN	2	1	3	14	12	2	..	..	..
S	11	8	20	15	4	11	35	14	50
UK	5	3	14	10	4	6	66	50	73
EU	5	1	15	18	4	14	37	83	52
EUR11	5	1	15	20	4	15	32	88	47
EU1994	3	1	11	..	..	..	22	77	37

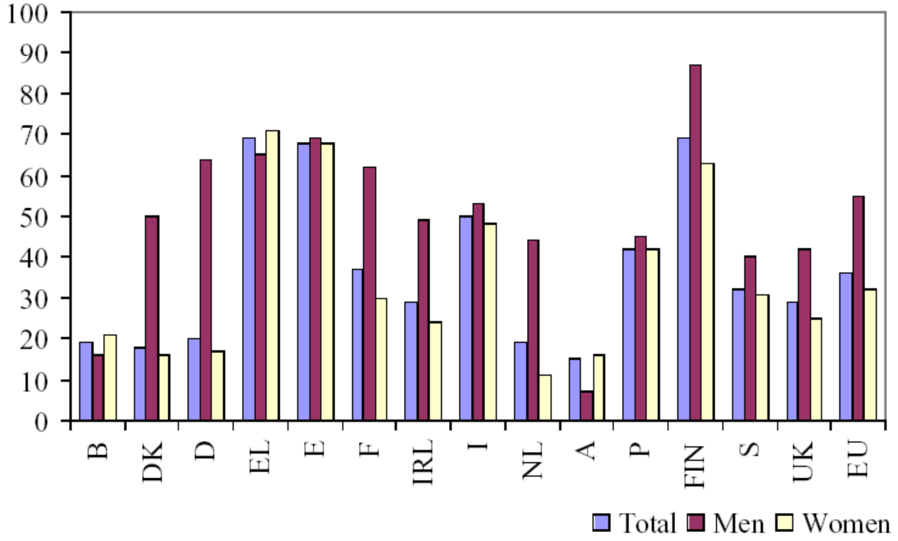
Source: EU ad hoc labour market survey.

As table 6 shows, part-time work has been increasing over the past five years, especially in services and retail trade. On average, only 5% (3% in 1994) of industrial employees in the Union work on a part-time basis; the percentage of part-timers rises to 18% in services and 37% in the retail trade.

In all sectors part-time work involves mainly women. The proportion of women employed part-time is remarkable in the Netherlands, both in industry and in retail trade (52% in both cases). The proportion of part-time female work is very high also in Germany concerning industry, and in Belgium and the United Kingdom as far as retail trade is concerned.

One of the distinguishing features of part-time employment, as opposed to other forms of non-standard employment (e.g., shift working), is the strongly favourable attitude of many employees towards it. This applies both to those who already work part-time and, in many cases, also to those who work full-time.

Figure 7. Readiness to work full-time (% of employees working part-time)



Source: EU ad hoc labour market survey.

Survey results measure the incidence of the so-called "involuntary part-time working" by asking part-time workers whether they would prefer working part-time or full-time. As figure 7 shows, the proportion of involuntary part-time workers - i.e. those who reply that they would prefer to work full-time - amounts to 36% for the Union as a whole, the figures been much above average in Greece, Finland, Italy, Portugal and Spain. Moreover, figures for men are usually higher than those for women (except in Belgium, Greece and Austria).

According to survey results, 33% of job seekers would prefer to get a part-time job than a full-time one, the figure rising to 50% for women. Percentages differ widely from country to country (85% in the Netherlands compared with 14% in Finland); this is probably due to the specific socio-economic structure in each of the Member States.

References

Corsi, M. - Roncaglia, A. (2000), The Employment Content of Growth in the EU, US and Japan, Study on behalf of the European Parliament, Bruxelles.

European Commission (1986), "Employment problems: views of businessmen and the workforce", European Economy, no. 27.

____ (1991), "Developments on the labour market in the Community", European Economy, no. 47.

____ (1995), "Performance of the European Union Labour Market", European Economy - Reports and Studies, no. 3.

____ (1999b), "Ad hoc Labour Market Survey in Services", European Economy - Suppl.B, no. 12.

____ (2000a), "Ad hoc Labour Market Survey in Industry", European Economy - Suppl.B, no. 1.

____ (2000b), "Ad hoc Labour Market Survey among Employees", European Economy - Suppl.B, no. 3.

____ (2000c), "Ad hoc Labour Market Survey in the Retail Trade Sector", European Economy - Suppl.B, no. 5.

Modigliani, F. et al. (1998), "An Economists' Manifesto on Unemployment in the European Union", BNL Quarterly Review, September.

OECD (1996), Employment and Growth in the Knowledge-based Economy, Paris.

____ (2000), Employment Outlook, Paris.

Sylos Labini, P. (1999), "The Employment Issue: Investment, Flexibility and the Competition of Developing Countries", BNL Quarterly Review, September.

Centralised CATI Method in the Finnish Consumer Survey: Testing the Effects of Nonresponse

Pertti Kangassalo and Veijo Notkola

1 Introduction

Statistics Finland introduced the Consumer Survey in November 1987. From October 1995, the data have been collected monthly in accordance with the harmonised EU data collection method (European Commission 1997). The main function of the Consumer Survey is to study the consumers' valuations and expectations concerning their own and the general economic situation, i.e. the economic sentiment.

Until December 1999, the data for the Consumer Survey were collected together with the data of the Labour Force Survey (Statistics Finland 1999). Six rotating panels were then used in the Consumer Survey. The same person was asked the same questions three separate times at six-month intervals, and each month a third of the target persons were new. Statistics Finland's experienced field interviewers conducted the telephone interviews around the country.

From January 2000 on, the Consumer Survey is based on a totally new individual sample each month. The same sample is used for the data collection of the Finnish Travel Survey and some other surveys, as well. All the interviews are now conducted centrally from Statistics Finland's computer-aided telephone interview centre in Helsinki (the CATI Centre). Interviews based on visits are not made any more due to costs and other practical reasons, even in case that the household does not have a telephone (Statistics Finland 2000).

During the period December 1999 to January 2000, the respondents' valuations concerning their own household's financial situation and the favourability of purchasing consumer durables and saving, in particular, have become more positive (Figure 1). In addition, some positive changes have taken place in some questions relating strictly to households instead of persons.[1]

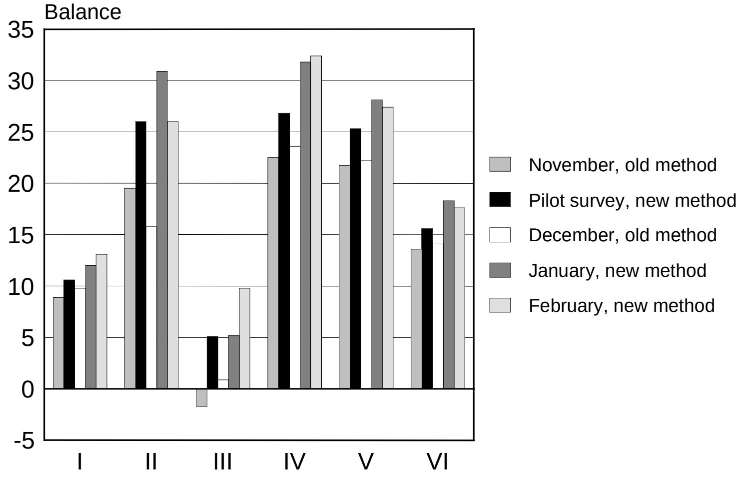
There are various possible reasons why these positive changes have occurred. Of course, the first possibility is that consumer attitudes have really changed to a more positive direction between December 1999 and January 2000. Usually during the turn of the year - and especially at the start of the new millennium - one might naturally expect consumers to become more optimistic about the future, as all kinds of sales start and the

[1] Quite similar experiences have also been gained from the Consumer Survey in Sweden (Statistics Sweden 2000)

spring comes closer. However, it is unlikely that this kind of explanation fully explains the changes, and looking back to a couple of earlier turns of the year shows that there were no such big changes. Furthermore, a limited pilot survey was carried out in November 1999 with the new method, and the comparison of results then pointed to exactly the same kind of differences as described above (Statistics Finland 2000).

The second explanation is that the change in the mode of the survey, from the Labour Force Survey to a centralised CATI survey, is the reason for the differences. As already mentioned above, in both cases interviews were conducted by telephone, but the Labour Force Survey is based on a panel design and field interviewers are used. The new Consumer Survey is based on the normal random sample of individuals monthly. The average age of field interviewers is about 50 years and in the CATI Centre the corresponding figure is about 28.

Figure 1. Effect of Method Change on Some Results of the Finnish Consumer Survey



I Own economy in the next 12 months
II Favourable time to make purchases
III Favourable time to save
IV Saving possibilities of household
V Household's financial situation
VI Consumer confidence indicator

It is clear that in this case there is not just one "mode effect", which can explain the change to a more positive direction. Normally, the mode effect is used to describe differences in coverage, nonresponse or measurement error between different methods (Nicholls II et al 1997). It is also worthwhile to notice that "it is important to make a distinction between the method or means of data collection (e.g. interview, self-administered questionnaire) and the technology used to capture the data" (de Leeuw and Collins 1997).

There are many earlier results about the central role of the survey mode as influencing the kind of answers the respondent gives. In particular, the differences can be great between a mail survey and a telephone interview or between a telephone survey and a face-to-face field interview (de Leeuw 1993). The topic of the survey is also relevant when the mode effect is analysed. It can be assumed that the data collection mode effect is quite small if the questions are mainly concentrated on factual questions, but the difference can be the greatest in those questions handling with more sensitive issues (Nicholls II et al 1997).

As a part of the mode effect, one possible explanation is that the differences in nonresponse between the surveys can explain the change towards a more positive direction. This is based on the fact that the nonresponse of the Finnish Consumer Survey, including mainly noncontacts and refusals, climbed from 21 to 29 per cent from November 1999 to February 2000. However, the nonresponse figures of November 1999, based on the Labour Force Survey, cannot be fully compared to the CATI Centre figures, due to the panel design of that survey.

The main aim of this paper is to analyse what kinds of differences there are in the structure of nonresponse between the consumer attitude studies based on the Labour Force Survey and the CATI Centre in Finland. The analysis of nonresponse will be made according to several background variables, such as the respondent's sex, age, education, income and region. The structure of nonresponse will be compared to the structure in the survey that was carried out earlier with the old method. In addition, we will compare the nonresponse to that of the Labour Force Survey representing the same month as the more recent Consumer Survey.

2 Data Sets and Analysis Method

We used three separate data sets for the analysis in this paper. The first one consists of the observations of the Consumer Survey carried out in February 2000 (called CS2000 below). The survey was based on a new sample with no (panel-type) connection to the samples of the previous months. The questions for the Finnish Travel Survey and the Finnish Fish Barometer (Finnish Game and Fisheries Research Institute) were asked at the same time as well. The interviewers of Statistics Finland's CATI Centre conducted all the interviews by telephone.

The second data set is based on the Consumer Survey that was carried out together with the Labour Force Survey in November 1999 (CS/LFS1999). The Labour Force Survey panel design was used, and Statistics Finland's field interviewers in provinces conducted the telephone interviews.

Finally, the third data set comes purely from the Labour Force Survey, and it represents the same month as the first data set, i.e. February 2000 (LFS2000). In that survey the common panel design was used, but only the first rotation group, the first-timers, were picked into the data set to allow the best possible comparisons with the Consumer Survey data of the same month.

The gross sample size of the Consumer Survey has generally been 2000-2200 persons monthly. For CS2000 there are 2192 observations in all, of which 1547 represent the persons having responded and the rest are cases that did not respond for different reasons (Table 1). For CS/LFS1999 the figures are 2011 and 1583 and for LFS2000 2398 and 2103, correspondingly. The over-coverage, e.g. some target persons, who appeared to have died or had moved permanently abroad, were excluded from the final analysis data set.[2]

In each of these surveys the target area was the whole Finland, and the respondents represented the 15 to 74-year-old population in the country, according to age, sex, province and native language. Systematic random sampling was used to extract the sample from the updated Central Population Register. The sorting system of the sampling frame was based on geographical population density. Our analyses indicate that the sampling method can be compared to the use of simple random sampling with the addition that the sample is geographically self-weighting. The extracting method itself has remained the same throughout the months in question, but for CS/LFS1999 and LFS2000 a panel design was used in addition (although the latter data set consists of the first-timers only, as mentioned above).[3]

The analysis of the nonresponse structure below is based on normal cross-tabulations and the testing of differences or interaction effects between the surveys on logistic regression models. Adjusted odds ratios for response by different background factors were also calculated by using logistic regression models (see for example Everitt and Dunn 1983). Logistic analysis was carried out by means of the GENMOD procedure.

In the analysis of the differences in nonresponse structure between surveys, we used the following socio-economic and demographic variables, which were all available in the Central Population Register, and they could be linked to the Consumer Survey data:

- sex;
- age: 15-24, 25-34, 35-44, 45-54, 55-64, 65-74 years
- language (mother tongue): Finnish, Swedish, other
- region: Greater Helsinki, rest of Southern Finland, Western Finland, Eastern Finland, Northern Finland
- education level: basic, secondary, higher education/lower, higher education/upper
- income: FIM 49,999 or less, FIM 50,000-99,999, FIM 100,000-149,999, FIM 150,000-199,999, FIM 200,000 or more.

[2] Both in CS/LFS1999 and in CS2000 the over-coverage rate was only 0.4 per cent. In LFS2000 this proportion was a bit higher, 1.1 per cent of the gross sample.

[3] The results of the Consumer Survey are afterwards weighted against the total population by means of sample weights. The weights are established by using the probability of each observation to be included in the sample and finally by a calibration method.

3 Results

3.1 Nonresponse and noncontacts

According to the results, the level of nonresponse differs quite drastically between the three data sets: in CS2000, the net nonresponse rate was clearly the highest, 29.4 per cent, unlike in LFS2000, which had nonresponse at only 12.3 per cent. In CS/LFS1999 the nonresponse rate was in between the two mentioned, i.e. 21.3 per cent (Table 1).

Table 1. Reasons for nonresponse by type of survey (%)

	CS2000	LFS2000	CS/LFS1999
Accepted interviews	70.6	87.7	78.7
Refusals	8.7	4.4	14.8
Noncontacts	19.2	7.2	5.8
Other reasons for nonresponse	1.5	0.7	0.7
All	100.0	100.0	100.0
	(N=2192)	(N=2398)	(N=2011)

CS2000 = Consumer Survey, February 2000

LFS2000 = Labour Force Survey, February 2000

CS/LFS1999 = Consumer Survey as part of Labour Force Survey, November 1999

Between the three surveys, there are great differences concerning, besides the level of nonresponse, also the distribution of nonresponse to refusals, noncontacts and other reasons more thoroughly discussed at the end of section 4. In CS/LFS1999, in about 70 per cent of the nonresponse cases the reason was refusing to answer and 27 per cent were not reached at all (i.e. noncontacts including cases without a telephone). On the contrary, in CS2000 about 65 per cent were noncontacts and only 30 per cent refusals. Rather conspicuously, the figures of LFS2000 resemble the latter ones, but that arises from the fact that in LFS data set only the first rotation group (first-timers) is included. Furthermore, in CS/LFS1999 the panel design clearly lowers the proportion of noncontacts.

3.2 Demographic and socio-economic differences in nonresponse by type of survey

Demographic differences

There is a clear difference between nonresponse rates by sex. In all surveys the proportion of nonresponse was higher among males and this difference was also statistically significant (Table 2). In addition, there was a tendency that the difference in nonresponse rates was highest in LFS2000, but this variation between surveys was not statistically significant.

Table 2. Nonresponse by sex and type of survey (%)

Sex	CS2000	LFS2000	CS/LFS1999
Male	30.4	14.1	23.4
Female	28.5	10.5	19.1
All	29.4	12.3	21.3
	(N=2192)	(N=2398)	(N=2011)

Sex effect: P=0.0011

Interaction sex * survey type: P=0.2716

Moreover, it seems obvious that no big differences exist between nonresponse rates by age, although in all three surveys the proportion of nonresponse in the youngest age group (15-24 years) belongs to the lowest end.[4] Also by type of survey, the differences between age groups are about the same, so the interaction effect remains statistically insignificant (Table 3).[5]

Table 3. Nonresponse by age and type of survey (%)

Age	CS2000	LFS2000	CS/LFS1999
15-24 years	25.6	9.7	19.4
25-34	30.2	12.9	22.1
35-44	30.4	15.3	23.5
45-54	30.0	13.2	20.3
55-64	30.4	11.6	18.9
65-74	30.0	8.3	25.3
All	29.4	12.3	21.3
	(N=2192)	(N=2398)	(N=2011)

Age effect: P=0.1424

Interaction age * survey type: P=0.4450

Contrary to what was discovered about a person's age, there appeared to be significant differences in nonresponse by the target person's mother tongue. In LFS2000 and CS/LFS1999, clearly the highest rates were observed among those who speak some

4 The modern nature of phone surveys and good mobile phone accessibility among younger people are probable explanations for that.

5 We also analysed the nonresponse effect by combining the variables sex and age (sex * age), but this only changed the results in a way that the effect of the combined factor reached a statistically significant level (as shown above, the variable age alone did not). The interaction effect was insignificant, and the results according to all the other factors remained about the same.

other language than Finnish or Swedish. It is quite remarkable that in CS2000 the biggest problems came out with the Swedish-speaking minority. Based on that, the interaction effect was also statistically significant in this case (Table 4).

Table 4. Nonresponse by language and type of survey (%)

Language	CS2000	LFS2000	CS/LFS1999
Finnish	27.3	12.1	20.2
Swedish	55.3	10.4	26.8
Other	51.0	34.3	49.0
All	29.4	12.3	21.3
	(N=2192)	(N=2398)	(N=2011)

Language effect: P=0.0001

Interaction language * survey type: P=0.0004

Regional differences

In all three surveys in question, the capital area seems to be the most problematic region in the sense of nonresponse, but the situation is similar throughout the survey field in Finland (Djerf et al 1996). Regarding the differences between the surveys, no significant difference was found at 95% level, although reaching that level was very close (Table 5).[6]

Table 5. Nonresponse by region and type of survey (%)

Region	CS2000	LFS2000	CS/LFS1999
Greater Helsinki	35.0	17.2	26.8
Rest of Southern Finland	27.6	14.6	18.5
Western Finland	28.5	10.4	22.9
Eastern Finland	24.5	8.2	17.2
Northern Finland	31.5	10.1	17.0
All	29.4	12.3	21.3
	(N=2192)	(N=2398)	(N=2011)

Region effect: P=0.0001

Interaction region * survey type: P=0.0540

[6] We found that by dropping LFS2000 out from the analysis and comparing only the two CS data sets we get somewhat stronger results, i.e. also the effect of region remains far from significant (the same holds for education as well; p-values 0.3274 and 0.1982, respectively).

Socio-economic differences

As regards the education level, the target persons with only basic schooling have the greatest tendency to nonrespond in all three surveys, which also represents a statistically significant difference. However, in LFS2000 the picture is somewhat different, but (at 95% level) no significant interaction effect was reported when comparing to the other two surveys (Table 6).

As was the case with the education level, also with respect to the person's income bracket, the highest nonresponse rates generally occurred among the target persons who had the lowest annual income. However, in LFS2000 and CS/LFS1999 about the same or even a higher rate was observed at the highest income level. The differences of the nonresponse rates were statistically significant, but there appeared to be no significant variation between the surveys (Table 7).

Table 6. Nonresponse by education and type of survey (%)

Education level	CS2000	LFS2000	CS/LFS1999
Basic	33.6	14.1	22.4
Secondary	28.9	10.3	21.3
Higher education, lower	24.1	12.0	20.7
Higher education, upper	20.0	14.1	15.7
All	29.4	12.3	21.3
	(N=2192)	(N=2398)	(N=2011)

Education effect: P=0.0001
Interaction education * survey type: P=0.0689

Table 7. Nonresponse by own income and type of survey (%)

Income level	CS2000	LFS2000	CS/LFS1999
FIM 49,999 or less/year	34.4	13.7	23.4
FIM 50,000-99,999	30.2	11.4	19.5
FIM 100,000-149,999	24.0	11.4	19.5
FIM 150,000-199,999	26.3	11.2	20.4
FIM 200,000 or more	25.4	13.6	24.3
All	29.4	12.3	21.3
	(N=2192)	(N=2398)	(N=2011)

Income effect: P=0.0001
Interaction income * survey type: P=0.6039

3.3 Adjusted differences in response by type of survey

Adjusting the different background factors does not have an effect on the basic result. There is still the same difference between the surveys. Based on the model where all seven background factors (sex, age, language, region, education, income and survey type) were adjusted at the same time, we can conclude that significantly the best response was attained in LFS2000 (Table 8). Also in CS/LFS1999 a clearly better response was reached than in CS2000 and this difference cannot be explained by differences in socio-economic and demographic factors.

In addition, on the basis of this analysis we can see that there are statistically significant differences in response rates by different variables, even if the effect of other variables is taken account. The response rates are better in the youngest age group (15-24 years), among females and among the Finnish speaking target persons. By region the lowest response rate was discovered in Greater Helsinki and there is no difference between other areas. There is also some difference by education and income level. The lowest response rate was among those having basic or secondary education and also by income, the lowest response rate was among those having the lowest income level.

4 Conclusions and Discussion

The results indicate that there are clear and statistically significant differences in nonresponse rates between surveys. However, - and what was our main aim to clarify - the structure of nonresponse seems to be about the same between surveys, although significant evidence was found by the language of the target person (and some minor effects by region and the respondent's education level). The person's mother tongue probably has no severe effect on survey results, because in any case more than ninety per cent of the target persons belong to the Finnish-speaking majority. Thus, our conclusion is that differences in nonresponse rates between surveys cannot explain the changes in the survey results.

After all this, what would then explain the changes in some survey results? Unfortunately, the change in the survey method at the turn of the year was quite comprehensive involving many kinds of different factors. The data collection mode changed (from the field to the CATI Centre), the type of interviewers changed (from middle-aged, experienced persons to young students) and the sample design was amended (from panel design to random sample design). Moreover, the whole context of the survey altered as well: from the Labour Force Survey to a more or less independent survey. After all these changes, it is quite logical that some kind of change must occur also in survey responses. Therefore, a pilot survey was conducted at the same time with the normal survey in November 1999, and extension of this study could well be partly based on that data in the future. One possible direction could be clarifying, in which respondent group(s) the changes in results are the greatest. On the other hand, there may also exist some differences between results based on the work of different interviewer types (young, old, new, regions etc.).

Table 8. Adjusted odds ratios for response by different background factors and by type of survey

		Odds ratio	Confidence interval
Sex	Male	0.80	0.70-0.91
	Female	1.00	
Age	15-24 years	1.51	1.15-1.96
	25-34	0.87	0.67-1.13
	35-44	0.72	0.55-0.93
	45-54	0.82	0.64-1.05
	55-64	0.96	0.74-1.24
	65-74	1.00	
Language	Finnish	2.64	1.83-3.80
	Swedish	1.49	0.98-2.28
	Other	1.00	
Region	Greater Helsinki	0.67	0.54-0.84
	Rest of Southern Finland	0.98	0.79-1.23
	Western Finland	0.99	0.81-1.22
	Eastern Finland	1.24	0.95-1.61
	Northern Finland	1.00	
Education level	Basic	0.57	0.41-0.79
	Secondary	0.70	0.51-0.96
	Higher education, lower	0.77	0.56-1.07
	Higher education, upper	1.00	
Income level	FIM 49,999 or less/year	0.69	0.52-0.91
	FIM 50,000-99,999	0.96	0.74-1.26
	FIM 100,000-149,999	1.17	0.89-1.53
	FIM 150,000-199,999	1.13	0.85-1.51
	FIM 200,000 or more	1.00	
Type of survey	CS/LFS1999	1.56	1.35-1.80
	LFS2000	3.04	2.61-3.55
	CS2000	1.00	

(Based on model: sex + age + language + region + education + income + survey type)

Due to the fact that there were so many changes at the same time in the mode of the survey, it is very difficult to differentiate the effects of these independent factors on survey results, and so far we can only conclude that the answer is not the different response rates of the surveys. As such, the change from the field to the CATI Centre might not have caused any differences, because in both methods the same basic technique is used; i.e.

interviewers, telephones and computer-assisted recording. Still, one could suspect that some kind of panel effect might have something to do with the phenomenon. Namely, earlier two thirds of the respondents each month were old cases, who had already replied once or twice in about one year's time.

On the other hand, taking into account the known differences in the interviewers' age and experience between the two modes, we could in future work concentrate on finding out whether the younger interviewers always achieve somewhat more positive results concerning the questions about the household's financial situation and some other subjects linked to it. In principle, that could be rather easily studied by separating the data gathered by the younger interviewers from the older consumer surveys and analysing only that against the data of the CATI Centre surveys. Furthermore, one interesting aspect is the local knowledge, dialect etc. of the field interviewers, which the CATI Centre interviewers less seldom have. However, even a field interviewer may very often not really know the interviewee after all.

We can still imagine, for example, a situation when the interviewer calls the interviewee and asks about his or her financial situation and the economic outlook of the person's household. If the respondent knows the interviewer (by name) beforehand or the respondent is conscious that the interviewer lives in the same small town or municipality – as most often is the case with Statistics Finland's field interviewers – in "Finnish-type-of-society", the respondent certainly tends not to give the interviewer too bright a picture of his or her finances. Therefore, in case of bigger cities and, what is important for this paper, especially in case of centralised telephone interviewing it might be more usual that households are ready to reveal also the good financial situation as it is, when there is no fear of losing one's privacy locally. That could be one possible explanation for the differences.

Also, the current independence or the association with a survey such as the Travel Survey, instead of the traditional and "serious-type" Labour Force Survey that concentrates on employment and unemployment, might have caused some similar effects as above.[7] Furthermore, now the Consumer Survey questions are always asked first, earlier they never came before the questions of the Labour Force Survey.

Finally, if nonresponse ever might have some effect on the results of the Consumer Survey, the critical point for the centralised CATI survey clearly is how the target persons are contacted at the beginning and how well they are then reached. (This is, of course, important for the level of the overall nonresponse rate of the survey as well.) The procedure of tracing the sample persons has changed to some extent along with the new CATI Centre method. Still, Statistics Finland sends out a target letter to each target person, telling him or her briefly the contents and the meaning of the survey. The phone numbers are mainly searched by means of a special paid service provided by the Helsinki Phone Company, but the CATI Centre interviewers complete it finally as far as possible. Also, the mobile phone numbers are looked up by this method (as is well known, Finland is the leading country in the world concerning the mobile phone coverage).

---

7 In the literature these effects are known as "context effects" (see for example Groves 1989 pp. 478-481).

Even though there are nearly 200 people every month who do not have a telephone or whose numbers remain unknown or secret, visits are not made, as already mentioned above. Those households are kindly asked by a target letter to contact Statistics Finland themselves. Unfortunately, very few do so. Before, Statistics Finland's field interviewers conducted the telephone interviews around the country, and whenever they did not have the target person's phone number, they could pay a visit on him or her. Often they might also have some better information than the CATI Centre interviewers about the target person's or his or her household's phone numbers and movements. Eventually, sending out the cases without any phone number to be handled by the field interviewers could possibly be one way of improving the present situation, but that would cause some extra costs for the survey. The crucial question is, how worthwhile, regarding the costs, it would be for the survey to get, in practice, only a small proportion of the numberless persons into the survey, when that may have only a minor effect or no effect at all on the results of the survey.

References

De Leeuw E. (1993): Mode Effects in Survey Research: A Comparison of Mail, Telephone and Face to Face Surveys. BMS 41, pp. 3-15.

De Leeuw E. and Collins M. (1997): Data Collection Methods and Survey Quality: An Overview. In "Survey Measurement and Process Quality", Lyberg L. et al (eds.). Wiley, New York.

Djerf K. and Takala K. (1997): Macroeconomy and Consumer Sentiment: Performance of the Finnish Consumer Barometer after Ten Years. Bank of Finland Discussion Papers 20/1997. Helsinki. (Paper presented in the 23rd CIRET Conference in Helsinki)

Djerf K. et al (1996): Trends of Nonresponse in the Surveys Conducted by Statistics Finland. In "Proceedings of the Sixth International Workshop on Household Survey Nonresponse", Laaksonen S. (ed.). Statistics Finland - Research Reports 219. Helsinki.

European Commission (1997): European Economy, The Joint Harmonised EU Programme of Business and Consumer Surveys. Reports and Studies No 6. Brussels.

Everitt B.S. and Dunn G. (1983): Advanced Methods of Data Exploration and Modelling. Heinemann Educational Books, London.

Groves R.M. (1989): Survey Errors and Survey Costs. Wiley, New York.

Groves R.M. and Couper M.P. (1998): Nonresponse in Household Interview Surveys. Wiley, New York.

Nicholls II W.L., Baker R.P. and Martin J. (1997): The Effect of New Data Collection Technologies on Survey Data Quality. In "Survey Measurement and Process Quality", Lyberg L. et al (eds.). Wiley, New York.

Statistics Finland (1999): Labour Force Survey publications and working papers. (In Finnish)

Statistics Finland (2000): Consumer Survey publications and working papers. (In Finnish)

Statistics Sweden (2000): Hushållens inköpsplaner, januari 2000. Pressmeddelande från SCB Nr 2000:021. (In Swedish)

Chapter V

Analysis of Survey Data
with Advanced Econometric Methods

Signalling Bottlenecks/Overcapacities and Low/High Dynamics

A New Approach to Indicate Changes in Business Cycles in the Manufacturing Industry Using Markov Switching Models on Business Survey Indicators

Marlene Amstad and Richard Etter

1 Introduction

There are three striking advantages to business tendency surveys (BTS): (a) BTS offer a range of business cycle information not supplied by official quantitative statistics. Company representatives give indications about changes in orders, production and inventories. In addition, they assess technical and personal capacities, inventories and order books. Moreover, their expectations and plans provide early information on changes to a whole set of important economic indicators. (b) BTS data have a publication lead on official data and, in addition, (c) are not revised. The latter two advantages constitute ideal prerequisites for the use of the data as leading indicators.

The three response categories are usually quantified by the balance technique. This simple method is widely used and has shown itself to be robust and at least equivalent to other techniques. But BTS information is generally limited if there is no close relation to official quantitative statistics (e.g. GDP). Moreover, the value of a balance indicator at one particular point in time can be compared only with its average - but the average is not a particularly interesting phase in a business cycle.

For this reason we present a method which enables the identification of extreme situations which cannot be sustained over a longer period, thus signalling an impending turning point. This method does not depend on a reference series. Thus, it is not only applicable to the whole economy but also provides signals analysing business cycles for particular regions, branches of business or other specific aspects. The used approach is developed and illustrated in Amstad (2000a). In this paper[1] we apply the methodology on a broader set of series of business tendency surveys and comment on their appropriateness in each case.

[1] This conference paper shows research until spring 2000 when both authors were at Swiss Institute for Business Cycle Reseach. For revised versions contact: marlene.amstad@snb.ch.

This paper is organised as follows. In the next section (Ch. 2) we describe the underlying concept of the used approach of early signals in business cycles. Chapters 3 to 5 deal with the prerequisites for the application of the approach. First of all, a description of the used data and the relation of BTS to different business cycle concepts is given. In Chapter 4 we discuss the statistical characteristics of the balance indicators; kernel density estimations and a normality test indicate that the traditional assumption of a unique normal distribution of the balance indicator is being undermined. In Chapter 5 a method for the identification of different averages of the indicators is presented - the Markov switching model.

In Chapter 6 the estimated parameter values are presented and their plausibility considered by examining whether they are in conformity with the stylised facts of the Swiss business cycle. Chapter 7 gives an interpretation along the lines of "bottlenecks/-overcapacities" and "high/low dynamics" using different indicators for business activities. In Chapter 8 signals are derived which serve as leading indicators of changes (in the growth cycle and acceleration/deceleration) in the business cycle of industrial production and of GDP. A summary and conclusions are given in Chapter 9.

2 The Concept

Forecasting turning points in the business cycle is often the focal point of empirical work on economic policies. The objective is to find indicators with a lead as long and stable as possible to a given reference series, in the course of which false signals should not be given. As the lead provided by the indicators does not usually exceed three quarters, economic policy measures are often instigated at too late a stage and may even generate procyclical effects.

Just as interesting in terms of economic policy is information which shows that the business cycle is in danger of leaving its sustained path or has even left it already. This means that it is not purely an issue of forecasting a specific point in time, but much rather of identifying the beginning of an excessive upward or downward movement in a cyclical phase. To us it appears to be of greater advantage from a macro-economic point of view to recognise when an economy starts to become overheated than to forecast the end of the boom phase. A method applied to BTS data is presented which is designed to identify extreme cyclical situations at an early stage. This method enables a much better use of the existing information in particular in the current situation.

The proposed method may be applied by using a reference series as an orientation (see Ch. 8) but this is not essential. It is also possible to make statements on the chosen indicator directly (see Ch. 7). This clearly expands the field of application. Information can be given on sensitive data such as on the profit situation and not only on industrial production or on GDP. Moreover, a major advantage is that it is possible to differentiate between different lines of business, regions, company sizes or export orientations.

The concept presented in this paper profits from the following circumstance: the business cycle is considered a two-part phenomenon according to most definitions.[2] A good indicator of the business cycle therefore has a distribution with two peaks. For this reason, in BTS questions are asked particularly on this dichotomous structure (+, -).[3] And indeed, preliminary work indicates that in most cases the balance indicators follow not only one but two normal distributions (Amstad (2000a)). The mean of both of the normal distributions signals the respective normal behaviour within a regime (boom or recession). Should the balance indicator exceed the mean boom value (recession value) on the current margin, it is leaving the "sustained" path of the boom (recession) phase which points to an overheating (excessive downward movement) in the business cycle.

Two prerequisites must be met so that the concept proposed in this paper may be applied:[4] (a) the BTS indicators must have a twin-peaked distribution and (b) a method for determining the mean values of the regime must be found. To begin with, the data situation is described in Chapter 3. The two subsequent chapters deal with both of the above prerequisites. In Chapter 4 the first prerequisite is evaluated by means of kernel density functions and the Markov switching model is presented in Chapter 5 to estimate the mean values.

3 The Data Base

3.1 Monthly and Quarterly BTS Data

This study uses data gained from BTS of the Swiss manufacturing industry conducted by KOF/ETH Zurich. The institute has been carrying out surveys by mail in this sector since the 1950s.[5] Firms are asked questions on a monthly basis and additional questions on a quarterly basis.[6] The period under review spans from January (1st quarter) 1979 to December (4th quarter) 1999.

---

2 Most definitions of the business cycle assume that it consists of two parts according to Burns and Mitchell (1946): "recurrent sequences of altering phases of expansion and contraction in the levels of a large number of economic and financial time series".

However, attention must be drawn to the fact that, for example, an M-shaped business cycle ("saw-tooth pattern of economic activity"), which would imply a third regime, is not taken into account.

3 The "same" reports may be considered a transition regime.

4 Apart from the prerequisite that the observed time series must cover several business cycles.

5 For international experiences with BTS see e.g. Oppenländer (1996). For Swiss experiences with BTS see e.g. Etter (1985).

6 For questionnaires see http://www.kof.ethz.ch/.

Table 1. BTS indicators of the monthly survey in the manufacturing industry

Abbr.	Variable	Type of question
NO_Y	New orders	compared to previous year
NO_E		expectations
OB_M	Order books	compared to previous month
OB_J		appraisal
PR_Y	Production	compared to previous year
PR_E		expectations
FP_M	Inventories of finished products	compared to previous month
FP_J		appraisal
IP_E	Purchase of intermediate products	expectations
PF	Plans of firms	synthetic indicator
BD	Business development	synthetic indicator [1)]
BD_E		medium term expectations

1) Constructed from appraisals of order books and inventories of finished products and of year-to-year changes of new orders and production.

Table 2. BTS indicators of the quarterly complementary survey in the manufacturing industry

Abbr.	Variable	Type of question
E_Q	Employment	compared to previous quarter
E_J		appraisal
TC_Q	Technical capacity	compared to previous quarter
TC_J		appraisal
TC_PC		percentage changes
PS_Q	Profit situation	compared to previous quarter
PIP_E	Prices of intermediate products	expectations

There are three possible answers to each qualitative question. With respect to the appraisals, the answers can be "too high", "normal" and "too low"; regarding the changes, "up", "unchanged" and "down" are the possible answers. The weighted[7] answers are aggregated to form percentages of each category of the total. The three response categories are quantified using the balance technique, i.e. the difference between the

[7] Weights on the micro-level: employment of the firm; weights on aggregated levels: weights calculated out of the population.

positive part and the negative part. This simple method is widely used and has proved itself to be robust and at least equivalent to other techniques (Dasgupta and Lahiri (1992)).

The two questionnaires used in the manufacturing industry include more questions than presented here. However, most of these were introduced only recently and are therefore unsuitable for this project. Questions on monthly, quarterly or future changes are deseasonalised.[8]

3.2 BTS and Different Business Cycle Concepts

When working with BTS data one has to bear in mind that BTS questions are based on different business cycles concepts. Here only a brief overview is given. The business cycle concepts discussed in this paper are illustrated in Amstad (2000a,b) in greater detail.

Growth cycle concept

Categorical judgements (too high, normal, too low) provide information on normal, over- or undercapacities. Firms base their judgements on the long-term average of the degree of utilisation. The judgements are based on a "normal" situation, but the "normal" situation can change over time. Thus, appraisals include a detrended view of an indicator. This corresponds with the growth cycle (GC) concept which goes back to the work of Mintz (1969). Growth cycles are defined as different states between points of extreme deviation from an average trend.

Acceleration/Deceleration concept

The balance which aggregates and quantifies the micro data provides information on the different intensities of changes. So the balance can be interpreted on the basis of the business cycle concept of acceleration/deceleration (A/D). This concept looks at points of extreme growth rates. The A/D concept can indicate more turning points than the growth cycle concept, but every turning point of the growth cycle has to show up in the acceleration/deceleration concept.

The relation between the BTS and business cycle concepts discussed above are illustrated in Table 3.

---

[8] CENSUS-X-11 procedure

Table 3. BTS and Business Cycle Concepts

	Growth Cycle	Acceleration/ Deceleration
Changes (past and future)		X
Judgements	X	

3.3 Reference Series: GDP and IP

In Chapter 8 we will compare the results with the reference indicators GDP and IP. Both are published by the Swiss Federal Statistical Office on a quarterly basis. We used them as a detrended series[9] or a year-to-year change series.

4 Statistical Properties of BTS Indicators

As a first step, we present the characteristics of various business survey indicators and examine whether the hypothesis that such indicators have a two-peak normal distribution does not need to be rejected. In the case of a simple normal distribution, the mean value provides little information as it does not represent an interesting phase of the business cycle. Estimating the kernel density of the different balance indicators from the Swiss business survey conducted in the manufacturing industry is one way of checking whether they follow only one fixed data generating process. This provides no formal evidence of the presence of two or more data generating processes. However, because at least the time series with two clearly separate mean values should show twin-peaked density functions, they are estimated for the balance indicators by means of kernel functions.[10]

The typical twin peaks of those indicators which have clearly separate mean values can be seen when viewing the chart displaying the kernel densities (cf. Figure 1 for the indicators FP_J, PR_J, NO_E). The different heights of the peaks show that each indicator remains in the two states for a different period of time (cf. also Ch. 6.2).

---

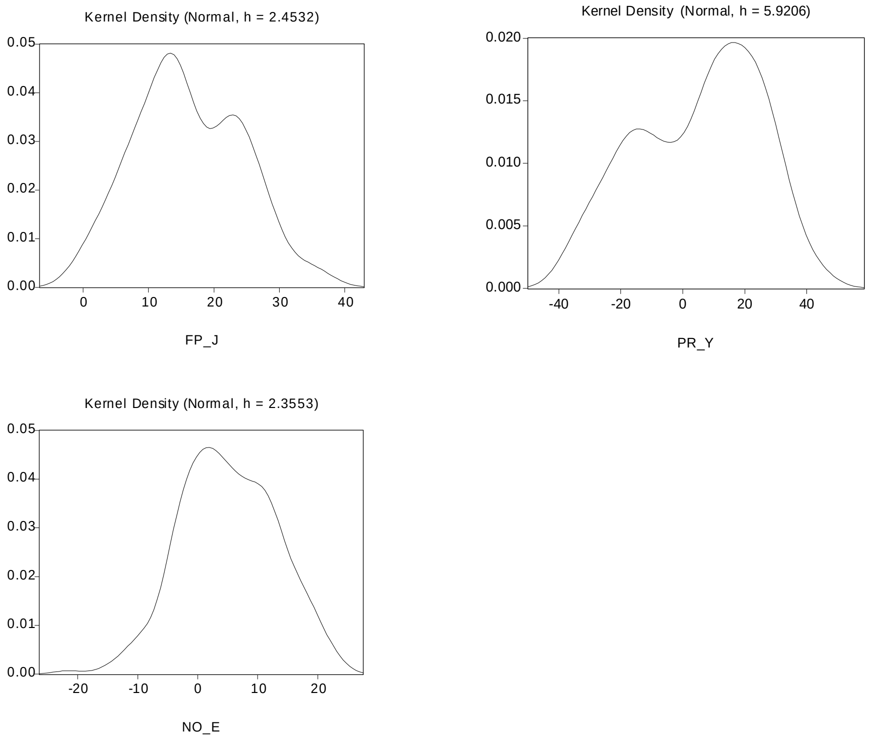
9 Phase Average Trend Method (PAT) used by OECD.

10 A kernel appraisal is steered by two parameters:

(1) Bandwidth: this determines the size of the window which is shifted over the data. The two-stage, adaptive procedure according to Silverman (1986) is applied. The value used is shown as "h" (the larger the bandwidth, the smoother the estimate).

(2) Core function: it weights each of the observations made in the window on the basis of the relative distance to the middle of the window. As the MSM models a mixture of normal distributions (cf. chapter 5), the normal distribution is used as the core function.

Figure 1. Estimation of Kernel Density of the Series FP_J, PR_J, NO_E



The Jarque Bera Test serves to check whether the hypothesis that an indicator has a simple normal distribution is to be rejected.

According to the Jarque Bera Test (Table 4) a single peak normal distribution can be dismissed for most monthly indicators. However, the case is somewhat different with the quarterly indicators: for most indicators, the single peak normal distribution cannot be rejected. In summary it is possible to state that the analysis of the charts and the Jarque Bera Test provide indications of a twin-peaked distribution for the majority of the balance indicators.

Table 4.

Indicator	Jarque Bera Test	Probability
monthly:		
NO_Y	11.71	0.003
NO_E	0.70	0.703
OB_M	7.55	0.023
OB_J	7.48	0.024
PR_Y	12.34	0.002
PR_E	15.27	0.000
FP_M	12.97	0.001
FP_J	4.33	0.115
IP_E	11.09	0.004
PF	15.91	0.000
BD	11.69	0.003
BD_E	0.07	0.962
quarterly:		
E_Q	4.60	0.100
E_J	0.70	0.705
TC_Q	3.08	0.214
TC_J	14.99	0.001
TC_PC	1.99	0.370
PS_Q	3.83	0.147
PIP_E	4.81	0.090

5 The Markov Switching Model (MSM)

To estimate both of the regime averages, a Markov switching model (Hamilton 1989)[11] was applied. Two averages were estimated for all the series shown in Table 5a and 5b (Appendix, column m1 and m2) with the variance modelled regime-independent. From the MSM one vector results, which provides information on the probability of economic activity persisting in a particular state (boom or recession).[12] However, this output is not taken into

11 In Hamilton and Susmel (1994) an intuitive way of understanding MSM is presented by introducing it in the context of state-space models. For an overview of attempts to model asymmetry in business cycles see e.g. Mills (1991).

12 A turning point signal based on this recession probability applied to the Swiss business tendency survey is estimated in Amstad (2000a S.125ff, 2000b).

consideration here. In this paper the focus is on estimating the density-parameters of the MSM. In the following chapters comments will be made on the latter and its use in signalling business cycle changes will be demonstrated in Chapters 7 and 8.

6 Interpretation of the MS Results

It becomes apparent that estimating parametric values is largely independent of whether the (much more volatile) original values or a seasonally adjusted input series are modelled. This can be interpreted as being the first indication of the MS modelling of survey data being suitable for forecasts, because forecasting procedures which are directly applied to the original series gain in lead in comparison to other methods which are based on seasonally adjusted and thus unstable marginal data.

It shall be shown below that the estimated parametric values provide plausible results – in the spirit of the stylised facts of the Swiss business cycle – for all modelled balance indicators.

6.1 Regime averages

The twin peaks which became apparent upon visual examination of the charts of the density functions of the balance indicators for various indicators are also confirmed by MS results for the other indicators.

For the **monthly series**, the MS results show, on the one hand, that the rejection of the simple normal distribution by the Jarque Bera Tests is supported. On the other hand, the remaining indicators (NO_E, FP_J, BD_E) show two means which significantly differ from one another (Appendix, Table 5a, column Diff). It can therefore be assumed that all business survey indicators recorded monthly are subject to a two-fold normal distribution.

For the **quarterly indicators**, with the exception of E_Q, TC_J and PIP_E, a simple normal distribution was unable to be rejected on the basis of the Jarque Bera Test at a 10% significance level. However, they all show two significantly different means m1 and m2 (Appendix, Table 5b, column Diff), which allows the assumption that there are two normal distributions even in these time series.

6.2 The Probable Duration in Each Regime

The probability of remaining in the existing regime (with the mean m1 or m2) varies depending on the indicator.

With **change questions,** the probability of remaining in regime 1 (acceleration phase) is greater than that of remaining in regime 2 (deceleration phase). This applies both to the monthly and to the quarterly series. For example, for NO_Y the probability of continuing in regime 1 stands at 0.97, which corresponds to 35.7 months. The probability of continuing in regime 2 is merely 0.94, which corresponds to 17 months. The fact that the mean m1 shows a smaller distance[13] to the overall mean m* than m2 reflects the asymmetrical course of the indicators. The same picture is obtained when examining the synthetic indicator "Plans of Firms" (PF).

Only three indicators are exceptions to this rule.[14] (1) Considering changes in inventories of finished products (FP_M), rationalisation measures (just-in-time) obviously led to a steady decrease in the stock/turnover relation which has caused m to turn slightly negative. However, in times of cyclical slumps a strong short-term unintended inventory increase occurs, which causes m1 to show high positive values for a short period. This interpretation is also supported by the fact that the length of duration in regime 1 is clearly shorter than in regime 2. (2) The larger distance of m1 to m than of m2 to m concerning the question of the subsequent business development (BD_E) appears less plausible. One reason for this may be the short period of observation. (3) The only monetary indicator PIP_E (expected price changes of primary products) signals that the probability of remaining in a regime with high price increases is lower (average duration of 6 quarters) than in a low price regime (11 quarters). This concurs with the generally low price increases in Switzerland.

The **judgements** of inventories of primary products, finished products and order books (often called negative inventories) are subject to a shift at companies, as they are often very carefully estimated. This is why both the mean values m1 and m2 are positive for stocks and negative for the order books. With regard to the length of duration, the inventories of primary products and finished products follow the pattern of the questions on changes – the indicators persist longer in regime 1 than in regime 2 (inverse view). The same picture is obtained for the synthetic indicator business activity (BD), which includes two judgement questions apart from two questions on changes. However, for the order books (OB_J) the phases estimated as "too small" dominate.

In summary it can be said that the estimated parametric values were able to be shown to be plausible for all the modelled balance indicators on the basis of stylised facts of the cyclical development. In the two following chapters the use of the parametric values to signal changes in the business cycle is shown.

7 Use of Mean Values as Signals

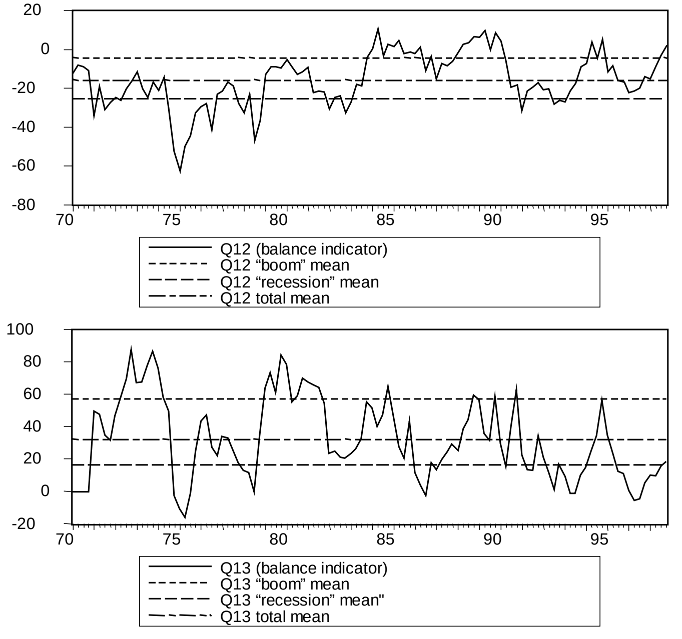
The two mean values calculated for each indicator provide an interesting array of interpretations in form of a signal. If the upper average value m1 (demarcation line 1) for balance indicators which are subject to the growth cycle concept is exceeded, this is interpreted as a signal of overheating (balance indicators according to the acceleration/-deceleration concept: strong growth phase); values falling below the lower mean value m2 (demarcation line 2) are interpreted as a signal of excessive downward movement in the

[13] E.g. balance indicator on the question of changes in production: m*= 3.7, m1= 15.0, m2= -18.3; distance of 11 resp. 22 points

[14] In addition to the following arguments, it should also be taken into consideration that the MS estimations presented in this paper are based on regime-independent standard deviations.

business climate (balance indicators according to acceleration/deceleration concept: marked contraction phase). First of all, two examples will show the use of the mean values as signals (from Amstad (2000a), pp. 114 ff), then all of the modelled series will be commented upon.

Figure 2. Unadjusted balance indicator values and MS equilibrium values of the series "profit situation compared with previous quarter" (Q12) and "expectations concerning purchase prices for the next quarter" (Q13)



7.1 Illustration on the Basis of Two Examples

Up to now, the mean of the balance indicators has been used – either implicitly or explicitly – to interpret survey results. It becomes obvious from Figure 2 that the parametric values gained from the MS model can serve as a further aid for interpreting survey results. Both of the estimated regime averages of a balance indicator series can be interpreted as a kind of "equilibrium value" in boom or recession phases. For example, the balance indicator of the profit situation attained a figure of 1.95 at the end of 1997. This figure is above the long-term mean of –15.8, as has been the case for all figures since the beginning of 1997. With the aid of both of the regime averages, the value 97:4 may also be interpreted as the profit situation not only being better than the long-term average but also being above the mean calculated for a boom phase (-4.3). Analogously it is possible to argue that at the end of 1997 the expectations concerning purchase prices for the next quarter (balance indicator: 18.6) had managed to recover to tally with the mean or equilibrium value in recession phases (16.6) even though they had not attained the long-term mean (32.3). The parametric estimates gained from MS analyses can thus be seen as an aid for interpreting companies' assessments of the cyclical situation. This can contribute to increasing the advantages gained from survey data and promoting their use in scientific fields as well as in the media and in politics.

7.2 The Results in Detail

When evaluating the results, it is again necessary to clearly distinguish between questions requiring an appraisal and questions about changes.

Judgement questions

For the balance indicator of appraisals and its underlying growth cycle concept, the respective mean value of the density functions signals the beginning (not the turning point) of under- or overcapacities. Passing the demarcation line, the balance indicator announces an unsustainable development in the underlying economic category. This is, in particular for economic policy recommendations, a very important insight.

The balance indicators of the various appraisals questions almost simultaneously signal the beginning of a phase of an overheating or cooling down of the business climate (see Figures 3 - 6 and Appendix, Figures 7 and 8). The first boom phase (apart from TC_J) started in 1979, but by 1982 a clearly negative development became apparent – first of all for FP_J, then also for E_J and then for OB_J and BD. The 1980s were marked by a sustained expansion phase which included two strong boom phases in 1985/86 and in 1988. The situation in the 1990s changed radically: a strong recession phase started in 1991, once again first indicated for BD (in conjunction with FP_J). After a slight recovery, a cyclical weak phase was again exhibited by all indicators. It was not until 1999/2000 that a clear phase of recovery was indicated. Only E_J did not yet exceed m1. It is well known that the job market always lags slightly behind the cyclical development.

Change questions

Questions on changes contain more white noise than questions requesting appraisals, which makes them more difficult to interpret. This applies most to monthly changes (no further comments will be made on these), slightly less to questions concerning expectations and plans and to a much lesser extent to comparisons with the previous year.

Change questions show more turning points than appraisal questions. This result reflects the statement theoretically arrived at in Chapter 3.2 that the A/C concept is able to show more turning points than the growth cycle concept.

The first distinct deceleration phase is indicated for the year 1979 in our set of data. Only NO_E and PS_Q did not exceed the m1 line. In 1982 a dramatic deterioration set in for all non-monetary indicators. Expected demand (NO_E) signalled the beginning of the next above-average acceleration phases first (2nd quarter 1983). Subsequently, the other indicators also showed a strong acceleration in growth. A slowdown in the pace of economic activity took place without actual slumps occurring. In 1988 a second surge in growth set in. The signs of above-average growth combined with reports of extreme situations in the questions requiring an appraisal indicated a clear contraction phase. An above-average contraction phase was first indicated by both of the expectation questions BD_E and NO_E. After a brief recovery phase, the industry reported another signal of contraction dynamics in 1992.

The following up and down phases of the change indicators were too short – apart from 1996 – to have an effect on the growth cycle. In the 1st quarter 1999 there were signs of a clear upturn in expected demand (NO_E). In the following quarter also various other balance indicators exceeded m1. As this is also confirmed by appraisal questions, it is possible that an economic problem is in the offing.

7.3 First Conclusions

The estimated regime averages can be used as demarcation lines. These enable a differentiated early assessment of a balance indicator value on the current margin: either (a) as an indication of an average boom (recession) phase or (b) as a signal for an exceptional phase which in the past did not persist to the same extent or did not persist for a longer period and preceded an upper (lower) turning point.

By comparing the signals developed here with the turning points in the IP and GDP, the use of this approach as a provider of important, previously unavailable information will be demonstrated.

8 Lead Characteristics to GDP and Industrial Production

If the upper average value m1 (demarcation line 1) is exceeded, this is interpreted as a signal of overheating, while a value falling below the lower average value m2 (demarcation line 2) is interpreted as a signal of an excessive slowdown in economic activity. A major advantage of this MS-based method is that no reference series or cycle concept are required for interpretations. Nevertheless, it is interesting to compare balance indicators with various reference series and cycle concepts.

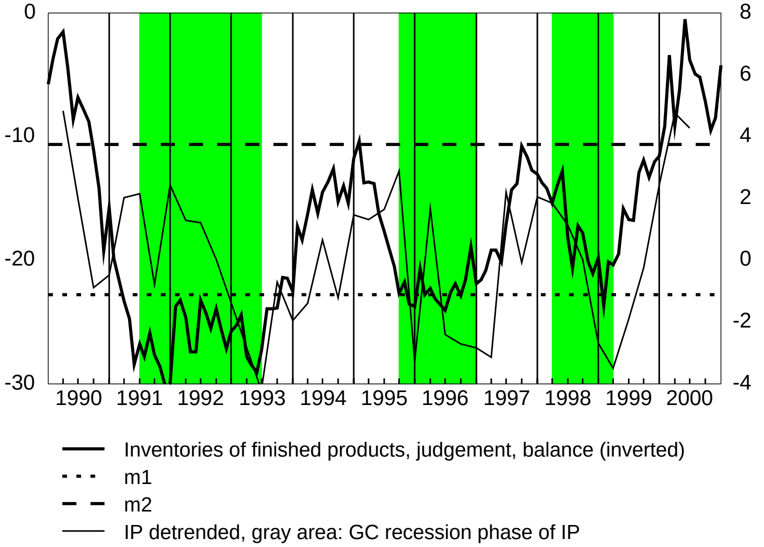
Two references series shall be considered below: (a) as the indicators are gained from the survey conducted in industry, the production index for industry is obviously suitable as a reference. (b) As the industrial sector has a strong effect on the volatility of the overall development and industrial statistics are only available from 1990 onwards, a comparison with the macroeconomic measure – GDP – is also advisable.

Two cycle concepts are considered for each of the reference series: (a) the growth cycle concept is the focal point for economic policy purposes (trend-adjusted series); (b) however, a comparison with the growth rates (acceleration/deceleration concept) is also of interest.

8.1 Early Signal for Growth Cycle in Industrial Production

According to the growth cycle concept, there were three contraction and four expansion phases in industry in the 1990s. The FP_J indicator provides an impressive reflection of the industrial cycle. Every time m1 is attained or figures fall below m2 - with a cyclical lead of variable length – a turning point in the growth cycle of industrial production is signalled, although there are only very slight distinctions in the cycles between 1994 and 1999. Warning signals only result after the demarcation lines have been noticeably exceeded. This occurred in mid- 1991 for the first time, with the indicator exceeding the value of m2 even before the beginning of the contraction phase. An early warning signal with a lead of over two years was thus created for the lower turning point. Although the following cycles were indicated, the ups and downs remained modest. It was not until the 1st quarter 2000 that a renewed distinct boom phase was heralded.

Figure 3. Industrial production, detrended and Judgement of inventories of finished products, inverse (FP_J, balance indicator)

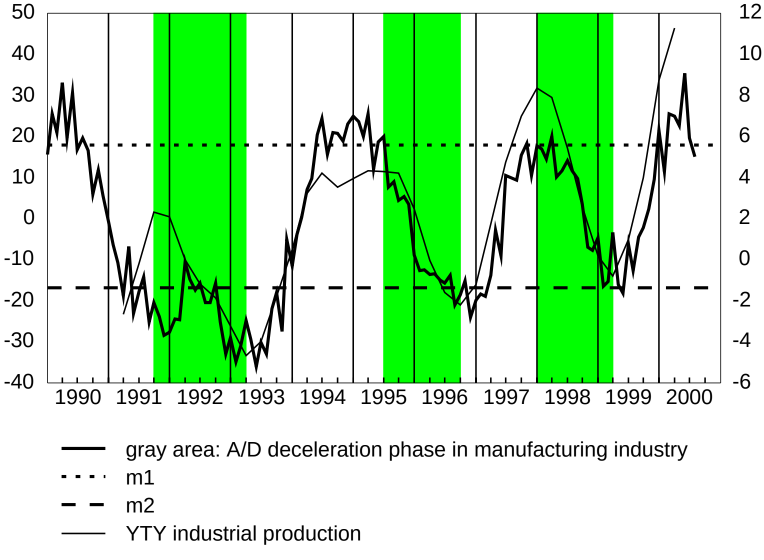


8.2 Early Signal of Acceleration/Deceleration in Industrial Production

In the 1990s the number of acceleration/deceleration cycles in industry was not higher than that of growth cycles. The **changes in production** lead one to fear a clearly negative development by mid-1991, although the contraction phase did not actually start until the 4th quarter 1991. A lively growth phase was indicated in 1994, which had to lead to a clear slowdown in economic activity at some stage. This finally occurred in the 3rd quarter 1995. The business cycle situation became dramatic in the 2nd half of 1996. The decline reported by the production index was confirmed by the BTS indicator exceeding m2 in the summer of 1996.

In the 2nd half of 1997 there was a brief danger of an economic overheating but the appraisal questions showed that the growth dynamics had not continued for long enough. The situation has become more tricky since the 2nd half of 1999 because the appraisal indicator – in addition to the change indicator – crossed its respective demarcation line. There are now clear signs of a possible economic overheating.

Figure 4. Industrial production, year-to-year changes and Production, year-to-year changes (PR_Y, balance indicator)



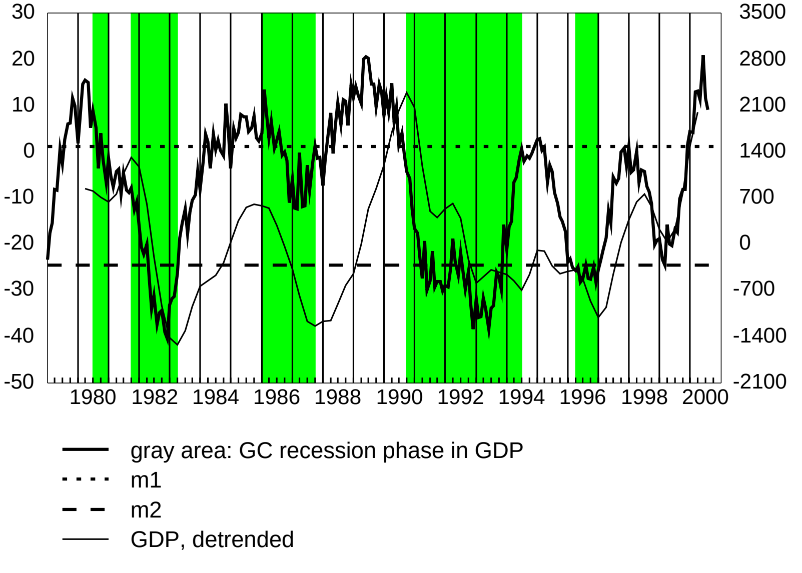
8.3 Early Signal for Growth Cycle in GDP

Displaying BTS data with GDP enables a comparison to be made for the 1980s too. GDP development continues to be strongly affected by industry, which justifies such a comparison. However, there are individual phases in the business cycle which only industry experiences and vice-versa. This must be taken into consideration when interpreting results.

In the 1980s the Swiss economy recorded three short growth cycle recessions, although expansion phases dominated. In the 1990s, by contrast, the economy was marked by a four-year recession phase. This was followed by a brief economic recovery and, subsequently, another short contraction.

The synthetic indicator of business activity (BD) clearly indicates three pronounced boom and two strong recession phases between 1979 and 1997. In the 3rd quarter 1979 an overheating phase began which developed into a growth cycle recession from the 4th quarter 1980 onwards. In the 2nd quarter 1982 the BD showed that the decline exceeded the usual extent of fluctuations. The bottom of the trough was consequently reached in the 1st quarter 1983. After this, signs of overheating were recognised in particular from 1984 and – after a short drawback - especially in 1988, with the latter phase leading to a marked recession from the end of 1990 onwards. The following light growth fluctuations were reflected only briefly by BD.

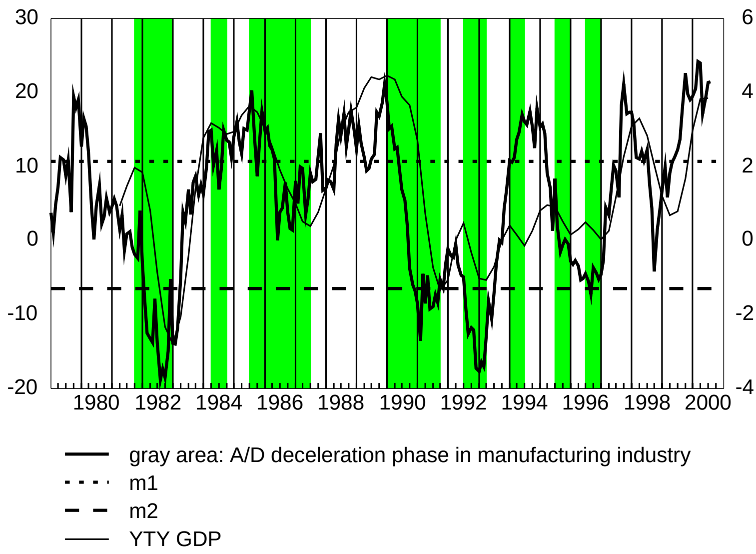
Figure 5. BIP, detrended and Business Development (BD, synthetic balance indicator)



8.4 Early Signal of Acceleration/Deceleration in GDP

According to the acceleration/deceleration concept, GDP shows eight acceleration and deceleration phases from 1979 to 1997. As Figure 5 shows, the productions plans (PR_E) indicate GDP development with some lead. Taking the demarcation lines into consideration, this indicator signals that only five accelerations and three decelerations exceeded the "normal" up and down of the economy. For example, PR_E fell below the value of m2 in the 1st quarter 1982, when slower economic growth set in properly and thus signalled a marked deceleration and even a decline. Lively dynamics in the business cycle set in 1984 and 1988. A phase of clearly slower GDP growth started in the 1st quarter 1990 and this negative development was intensified in the 3rd quarter of 1992. In 1994, a lively development was signalled, but this seems to have happened mostly in manufacturing industry.

Figure 6. BIP, year-to-year changes and Production, expectations (PR_E, balance indicator)



9 Summary and Conclusions

The presented approach generates two asymmetric means of the balance indicators of business surveys. If the upper mean value m1 (demarcation line 1) is exceeded, this is interpreted as an indication of business activity overheating, while a value falling below the lower mean value m2 (demarcation line 2) is interpreted as a signal of an excessive slowdown in economic activity. Using the derived demarcation lines considerably improves the content of the information of what is basically qualitative data, in particular for the most recent economic development (actual bound).

In business tendency surveys it is necessary to distinguish between appraisals with an underlying growth cycle concept and past or future changes, which follow the concept of acceleration/deceleration. By using the two demarcation lines m1 and m2, the growth cycle concept indicates strong boom or recession phases. In the acceleration/deceleration concept the two means signal strong dynamics within positive or negative changes. This gives an indication of a forthcoming bottleneck or overcapacity if not already indicated by the appraisal questions.

This approach enables signals to be derived which indicate changes (in the growth cycle and acceleration/deceleration concepts) in the business cycle of industrial production and of GDP at an early stage. With this additional information it is possible to expand the scope of monitoring of the business situation and it will facilitate the interpretation of business tendency survey results for the analysis of the business cycle and for conclusions on economic policy recommendations.

References

Amstad, Marlene (2000a): "Business Cycle Turning Points: Dating and Forecasting - Chronology of different types of turning points and development of an early warning system by means of Markov Switching Models applied on Swiss business tendency surveys", (in German), Dissertation University St. Gallen.

Amstad, Marlene (2000b): "On the use of Markov Switching Models applied to business survey data for the prediction of turning points". pp. 3-26 in: Use of Survey Data For Industry, Research And Economic Policy - Selected papers presented at the 24th CIRET Conference, Wellington, New Zealand, 1999, edited by Oppenländer, K.H. and G. Poser: Ashgate (2000).

Burns, A. / W.C. Mitchell (1946): "Measuring Business Cycles", National Bureau of Economic Research.

Dasgupta, S. / K. Lahiri (1992): "A Comparative Study of Alternative Methods of Quantifying Qualitative Survey Responses Using NAPM Data", Journal of Business and Economic Statistics, Vol. 4, pp. 391-400.

Etter, Richard (1985): "Die Unternehmensbefragungen als Analyse- und Prognoseinstrument der schweizerischen Konjunktur", Konjunktur, 30 Jahre Konjunkturtest, KOF/ETH Sonderbericht Nr. 165, Zurich.

Hamilton, J.D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle" Econometrica, Vol. 57, pp. 357-84.

Hamilton, James D. / Raul Susmel (1994): "Autoregressive Conditional Heteroskedasticity and Changes in Regime", Journal of Econometrics, Vol. 64, pp. 307-33.

Mills, T. (1991): "Nonlinear Time Series Models in Economics", Journal of Economic Surveys, Vol. 5, pp. 215-42.

Mintz, Ilse (1969): "Dating Postwar Business Cycle: Methods and their Application to Western Germany, 1950-67", Occasional Paper N°107, edited by NBER, .

Oppenländer, K. H. (1996): "Konjunkturindikatoren", Oldenbourg Verlag.

Appendix

Table 5a. Maximum Likelihood of Parameters for Markov-Switching Model of Monthly BTS [a), b)]

	m1	m2	Diff	s	P11	P22	lnLIK	D1	D2	D1/D2
NO_Y	14.993	-18.285	33.279	10.948	0.972	0.941	758.69	35.7	17.0	2.10
	(0.963)	(1.374)	(1.678)	(11.512)	(0.014)	(0.026)				
OB_M [c)]	1.592	-16.573	18.165	6.876	0.975	0.942	637.79	40.4	17.2	2.35
	(0.610)	(1.026)	(1.194)	(4.357)	(0.013)	(0.027)				
OB_J	-12.436	-37.827	25.391	8.862	0.978	0.984	692.12	44.4	61.6	0.72
	(0.945)	(0.826)	(1.255)	(7.150)	(0.014)	(0.011)				
PR_Y	18.046	-16.887	34.934	10.467	0.969	0.949	748.09	32.8	19.7	1.66
	(0.969)	(1.276)	(1.603)	(10.397)	(0.014)	(0.023)				
FP_M [c)]	4.862	-3.187	8.048	3.968	0.920	0.971	498.91	12.5	34.8	0.36
	(0.638)	(0.321)	(0.714)	(1.496)	(0.037)	(0.013)				
FP_J	22.787	10.581	12.206	3.748	0.956	0.977	369.79	22.8	42.7	0.53
	(0.476)	(0.377)	(0.607)	(1.466)	(0.023)	(0.014)				
NO_E [c)]	11.135	-1.158	12.293	4.979	0.947	0.944	563.67	18.8	18.0	1.04
	(0.616)	(0.574)	(0.842)	(2.428)	(0.022)	(0.022)				
PR_E [c)]	10.723	-6.514	17.237	5.430	0.983	0.958	574.10	59.1	23.8	2.49
	(0.435)	(0.666)	(0.796)	(2.688)	(0.010)	(0.022)				
IP_E [c)]	5.738	-12.292	18.030	5.960	0.977	0.956	599.32	43.8	23.0	1.91
	(0.543)	(0.823)	(0.986)	(3.307)	(0.012)	(0.023)				
BD_E [c)]	8.186	-3.404	11.590	4.913	0.950	0.966	433.00	20.2	29.0	0.70
	(0.752)	(0.488)	(0.896)	(2.608)	(0.030)	(0.017)				
BD	1.108	-24.453	25.561	7.603	0.972	0.954	(A/D)	6	21.5	1.66
	(0.668)	(0.886)	(1.110)	(5.342)	(0.013)	(0.021)				
PF	8.166	-9.601	17.767	5.291	0.982	0.959	567.44	54.7	24.4	2.24
	(0.443)	(0.680)	(0.812)	(2.554)	(0.010)	(0.022)				

a) standard errors in parentheses;

b) m1, m2 = regime-dependent means;
Diff = test statistics;
s = regime-independent standard errors;
P11,P22 = probability that regime 1 (2) will be followed by regime 1 (2);
lnLIK = likelihood value ;
D1, D2 = average regime persistence

c) Extreme values corrected

Table 5b. Maximum Likelihood of Parameters for Markov-Switching Model of Quarterly BTS [a), b)]

	m1	m2	Diff	s	P11	P22	lnLIK	D1	D2	D1/D2
E_Q [c)]	9.336	-19.841	29.177	9.579	0.935	0.875	248.72	15.3	8.0	1.9
	(1.977)	(2.614)	(3.277)	(15.950)	(0.040)	(0.059)				
E_J	14.383	-9.583	23.965	10.073	0.966	0.954	245.54	29.6	21.9	1.4
	(1.515)	(1.904)	(2.433)	(16.026)	(0.025)	(0.035)				
TC_Q [c)]	9.589	2.992	6.597	3.050	0.949	0.925	147.45	19.8	13.4	1.5
	(0.511)	(0.605)	(0.791)	(1.512)	(0.036)	(0.047)				
TC_J	9.263	-4.198	13.462	6.375	0.976	0.938	156.64	41.4	16.2	2.6
	(1.156)	(1.661)	(2.023)	(7.418)	(0.026)	(0.051)				
TC_PR [c)]	86.348	81.829	4.519	1.667	0.900	0.906	101.84	10.0	10.6	0.9
	(0.334)	(0.373)	(0.501)	(0.461)	(0.045)	(0.048)				
PS_Q [c)]	-3.326	-21.252	17.926	6.100	0.929	0.873	211.62	14.2	7.9	1.8
	(1.034)	(1.367)	(1.714)	(6.535)	(0.041)	(0.061)				
PIP_E [c)]	52.187	14.846	37.341	12.190	0.885	0.924	270.67	8.7	13.1	0.7
	(2.839)	(2.034)	(3.492)	(26.606)	(0.061)	(0.039)				
IIP_J	18.526	10.145	8.380	3.163	0.879	0.947	153.77	8.2	18.7	0.4
	(0.727)	(0.482)	(0.872)	(1.633)	(0.062)	(0.031)				

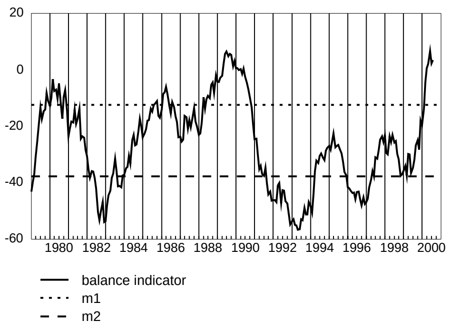
a) standard errors in parentheses;

b) m1, m2 = regime-dependent means;
Diff = test statistics;
s = regime-independent standard errors;
P11,P22 = probability that regime 1 (2) will be followed by regime 1 (2);
lnLIK = likelihood value ;
D1, D2 = average regime persistence

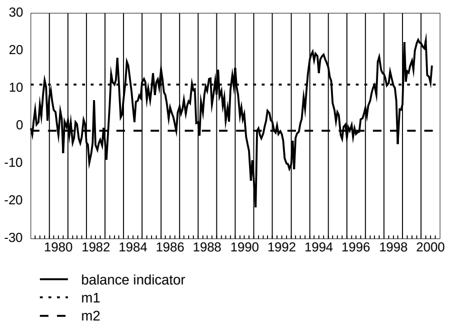
c) Extreme values corrected

Figure 8. Monthly Time Series (see also Figures 3-6)

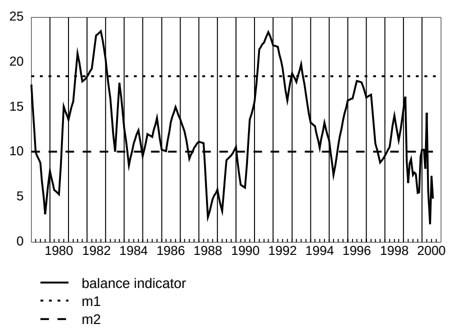
Judgement of Order Books



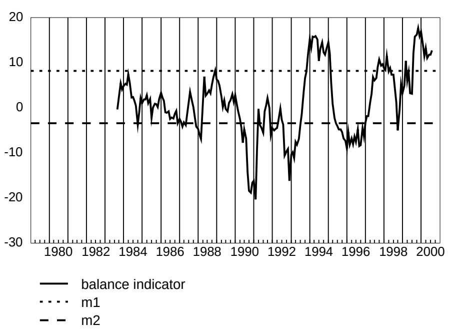
New Orders, Expectations



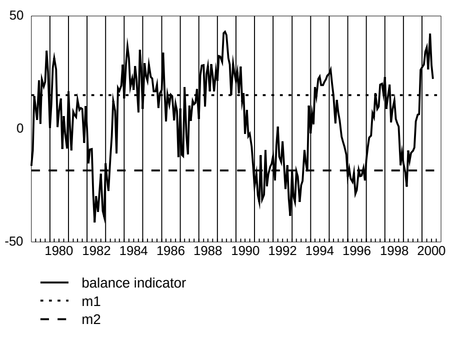
Judgement of Inventories of Intermediate Products



Business Development, Medium Term Expectations



New Orders, Year-to-Year Changes



Firms Plans, Synthetic Indicator

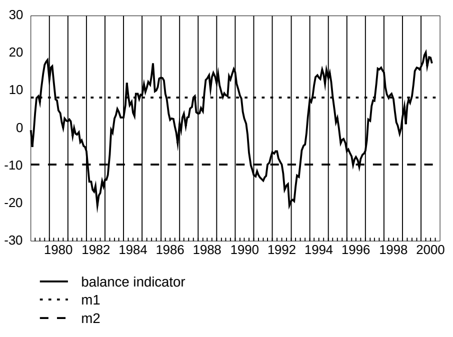
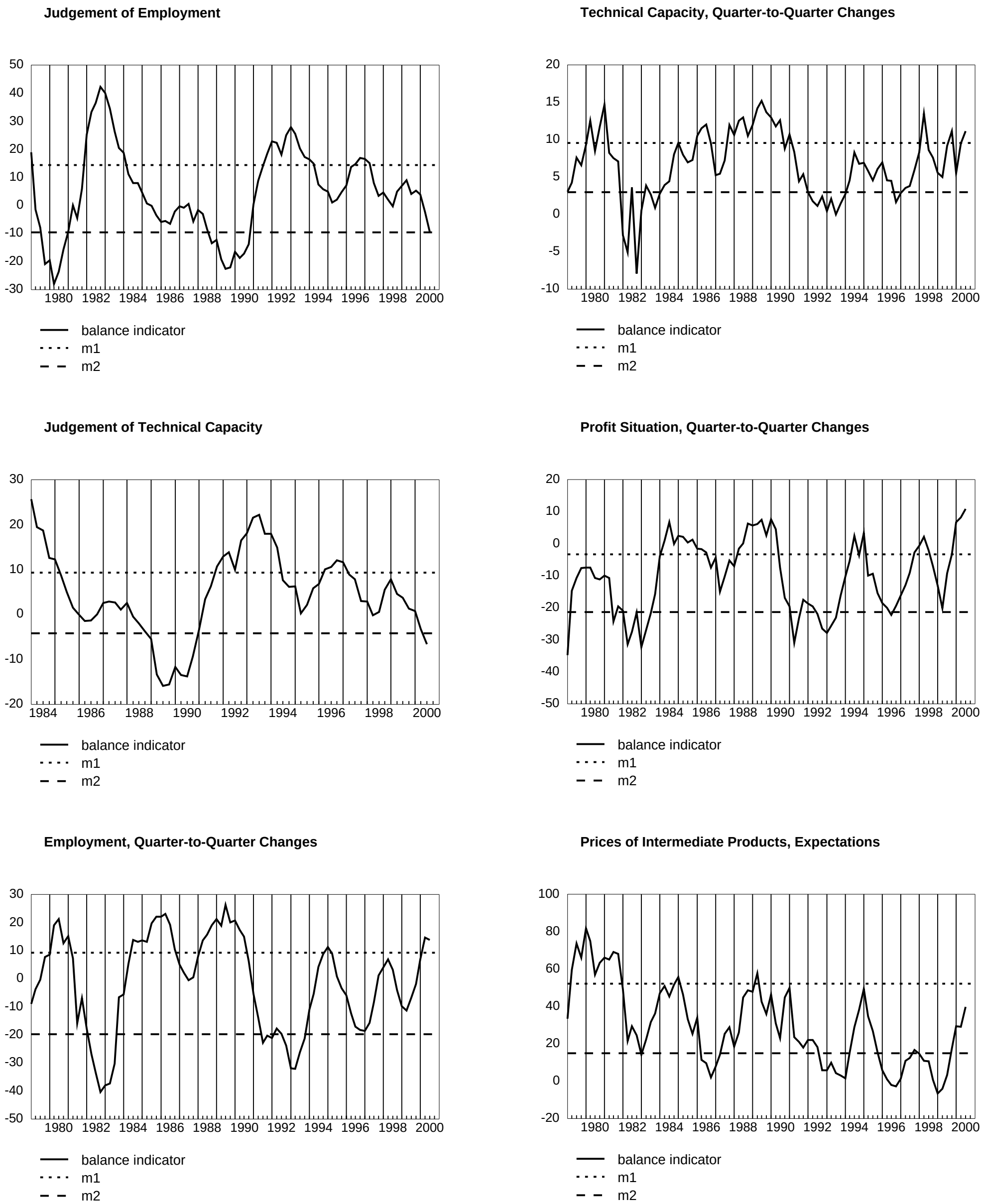


Figure 9. Quarterly Time Series



The Impact of Nonresponse Bias on the Index of Consumer Sentiment

Richard Curtin, Stanley Presser, and Eleanor Singer

Introduction

A basic tenet of survey research is the absolute preference for high response rates. A low response rate, more than any other single indicator, is considered to be a major threat to the usefulness of the collected data. The emphasis on high response rates stems from the belief that increases in non-response lead to greater bias in the resulting survey estimates. Survey organizations have devoted an increasing share of their budgets to reducing non-response rates by making multiple calls as well as attempting to convince respondents that had initially refused to agree to be interviewed. In the U.S. as well as many other countries, the trade-off between costs and the potential bias due to declines in response rates represents a critical issue for the measurement of consumer confidence.

This paper[1] explores the impact of survey non-response on estimates of the Index of Consumer Sentiment, based on the results from more than two hundred monthly surveys conducted by the Survey Research Center at the University of Michigan.[2] Two criteria were used to select the surveys included in this analysis, both involving sample design issues. The first involved the difference between face-to-face and telephone interviews, with the analysis restricted to the past few decades when the surveys were based on random digit dial telephone samples. Since the sample is designed as a rotating panel, the analysis of nonresponse was restricted to the initial interview. The full sample for each month consists of 60 percent new cases and 40 percent reinterviews. Each month's new sample is representative of all private households in the coterminous United States, with the respondent randomly selected from among all adults aged 18 or older living in the household. Each month about 300 initial interviews are conducted, although the number was somewhat larger in the earlier years. The 211 surveys included in the analysis were conducted between June 1979 and December 1996, with the number of interviews totaling more than 72,000.

---

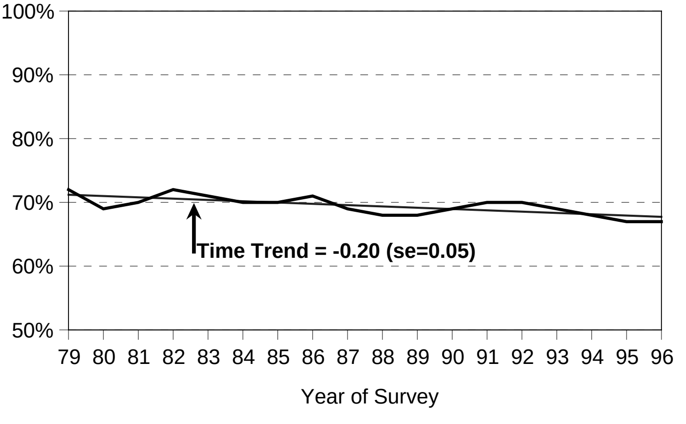
1 A different version of this research appeared in Public Opinion Quarterly, V. 64 (2000), entitled "Effects of Response Rates Changes on the Index of Consumer Sentiment," by Curtin, Presser, and Singer.

2 See the appendix for the question wording and the formula used to construct the Index of Consumer Sentiment.

Response Rate Trends for The Surveys of Consumers

The proportion of eligible respondents that refuse to participate in household surveys has risen in the United States, Europe, and Asia. While there are considerable differences across countries, by the type of survey, and by the sponsoring organization, the increases in refusal rates have been broadly based and persistent (Steeh, 1981; Groves and Couper, 1998; de Leeuw, 1999; de Herr, 1999; Synodinos and Yamada, 2000). The average yearly response rates for the Surveys of Consumers are shown in Chart 1 (the response rates take account of all sampled phone numbers with the sole exception of those known to be ineligible). The response rate over the period from 1979 to 1996 ranged from a high of 72% to a low of 67%, averaging about 70 percent. The data indicate a persistent slow decline over time, with the response rate falling one-fifth of a percentage point per year on average. The estimated rate of decline was highly significant (the time trend coefficient was four times its standard error), and exhibited only small year-to-year variations about the trend.

Chart 1. Response Rates for Surveys of Consumers



Low response rates are the combination of two factors: the failure to contact all eligible respondents in the sample and the failure to convince all contacted respondents to participate. Such failures are due to either the reluctance of respondents to participate in surveys or to the lack of effort devoted to contacting and convincing respondents to participate by survey organizations. As a result, the declines in response rates represent the interplay of changes in respondent reluctance and changes in survey efforts. Indeed, the Surveys of Consumers have increased efforts to counteract declines in response rates. Other than the constraint imposed by the month-long data collection period, the surveys have imposed no limit on the number of times sample telephone numbers were called, and attempts to convert all initial refusals were made by specially trained interviewers.

The proportion of all completed interviews that were initial refusals doubled from 1979 to 1996, rising form 7.4% to 14.6% (Chart 2), as did the mean number of calls to complete each interview, which rose from 3.9 to 7.9 (Chart 3). These are separate trends, as the mean number of calls to complete interviews that did not require refusal conversion also increased (from 3.7 to 7.4), with both increasing in parallel at 0.15 calls per year per completed interview. Taken together, these additional efforts translated into a substantial increase in the amount of interviewers' time per completed interview—2.1 hours in 1981 versus 2.7 in 1996, based on an average interview length of 33 minutes in both years. These figures suggest that the Surveys of Consumers were able to limit the annual response rate decline to 0.2% by increasing interviewer's hours by about 2% per year. Needless to say, it has been this yearly escalation of costs that has increasingly focused attention on the impact of nonresponse bias.

Chart 2. Refusal Conversions

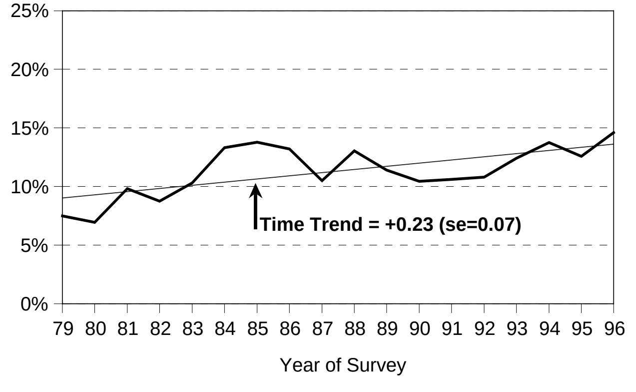
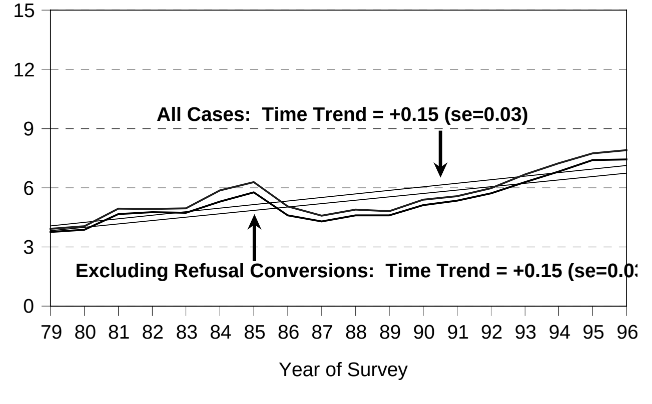


Chart 3. Number of Calls to Completion



Nonresponse Bias: Conceptual Framework

Nonresponse errors arise because not all sampled households are interviewed. Abstracting from other sources of error[3], the "true" mean level of the Index of Consumer Sentiment (ICS) can be defined as the weighted sum of the mean for households that were interviewed and the unobserved mean for the nonrespondents, with the weight (B) defined as the proportion of nonrespondents in the total sample (Cochran, 1977; Groves and Couper, 1998):

$$ICS_t = (1-\pi_t)ICS_t^o + \pi_t ICS_t^u$$

or, in terms of the observed value of the ICS:

$$ICS_t^o = ICS_t + \pi_t\left(ICS_t^o - ICS_t^u\right)$$

Defining the nonresponse bias as the expected value of the difference between observed mean and the "true" mean yields:

$$BiasICS_t^o = \pi_t\left(ICS_t^o - ICS_t^u\right)$$

The nonresponse bias is thus a multiplicative function of the size of the difference in the means and the proportion of nonrespondents. How these two factors are related is of some consequence.[4] If the two factors were completely independent, the factors underlying nonresponse would be unrelated to the factors that determine the response variable. In this fortunate circumstance, the expected value of the mean difference in the response variable would be zero, thus eliminating the nonresponse bias. More commonly the two factors are assumed to be related at least to some degree. The direction of the relationship could vary, however, so that as the proportion of nonresponse increases, the difference in means could become larger or smaller. Indeed, it is possible for the overall bias to be smaller at higher nonresponse rates, if nonrespondents become increasingly similar to respondents as the nonresponse rate rises. It may be that the difference in the means increase as the nonresponse rate diverges from 50% - in either direction.

Since the primary focus of the consumer surveys is the measurement of change in expectations, the nonresponse bias of estimates of change can be expressed as:

---

3 Other sources of potential bias include sampling, coverage, and measurement errors. The properties of sampling errors are well understood and can be controlled by probability sampling techniques. In addition, telephone surveys incur coverage errors, and like nonresponse errors, are usually addressed by the use of statistical adjustments via sample weights.

4 The ability of sample weights to counter nonresponse bias depends on this relationship. If within weighting classes both respondents and nonrespondents share similar likelihoods of participation and values on the survey variables, then the nonresponse error in the weighted estimates are lower than for the unweighted estimates.

$$Bias\left(\Delta ICS_t^o\right) = \pi_t\left(ICS_t^o - ICS_t^u\right) - \pi_{t-1}\left(ICS_{t-1}^o - ICS_{t-1}^u\right)$$

The methodological advantage of change measurements is readily apparent: if the nonresponse rate and mean difference are relatively constant over short periods of time, the two terms are equal and thus the nonresponse bias vanishes—even if the nonresponse bias is relatively large in each of the two time periods.

Analytic Strategy: Proxy Nonrespondents

The true extent of nonresponse bias is not known, since no data for the ICS questions are available for nonrespondents. While the absolute level of nonresponse bias cannot be determined, relative changes from the current level of nonresponse bias were simulated using the collected data. The strategy was to partition the completed interviews into two groups by selecting some cases to represent proxy nonrespondents. While there is no sure method to identify which respondents would have been nonrespondents under alternative designs and interviewing procedures, potential nonrespondents were identified by the amount and type of effort actually expended to contact or convince the respondent to participate.

Table 1. Proxy Nonresponse Groups

Proxy Nonresponse Groups	Percent of Interviews	Implied Response Rate
Initial Refusals	11.2%	61.8%
6 or more calls	33.4%	46.3%
2 or more calls	63.8%	25.2%

Three different comparison groups were used to simulate lower response rates: the exclusion of refusal conversion cases, which lowered response rates by about 10 percentage points, the exclusion of cases that required more than five calls to complete the interview, which lowered response rates by about 25 percentage points, and the exclusion of interviews that required more than two calls to complete, which lowered response rates by nearly 50 percentage points (see Table 1, the "implied response rate" represents the survey's average response rates if the cases in the nonresponse groups were treated the same as the survey's true nonrespondents). The comparison groups were purposely selected to cover a very broad range of response rate reductions, using simulated designs that are admittedly more characteristic of commercial surveys than those sponsored by governments or conducted by academic organizations. To be sure, the simulation of a 5-call design is not equivalent to what would have been done if the study was initially designed to limit the number of calls to 5; such designs would more carefully control the time and days when each of those 5 calls were made. It was this recognition of the difficulty of identifying potential nonrespondents that prompted the simulation of rather large reductions in response rates.

The expected bias induced by the exclusion of the proxy nonresponse cases essentially adds an additional term to the above equations. The observed ICS^{o} can be partitioned into ICS^{o^*} and ICS^{u^*}, where the latter term represents the cases designated as proxy nonrespondents. Similarly, the total nonresponse rate can be partition into B and $B*$, with the later representing the proportion of the sample designated as proxy nonrespondents. The expected value of the bias would then be expressed as:

$$Bias\left(ICS_t^{o*}\right) = \pi_t^*\left(ICS_t^{o*} - ICS_t^{u*}\right) + \pi_t\left(ICS_t^{o*} - ICS_t^{u}\right)$$

The total nonresponse bias is thus partitioned into two components: The portion that is due to proxy nonrespondents (the first term) and to true nonresponse (the second term). As the proportion of proxy nonrespondents approaches zero, the above equation reduces to the former since when $B* = 0$, $ICS^{o^*} = ICS^{o}$. This equation makes it clear that the analysis reported in this paper only focuses on the impact of the additional nonresponse not the total bias.

Impact of Nonresponse Bias: Total Sample Estimates

Index of Consumer Sentiment averaged 83.2 when calculated on the entire pooled 1979-1996 sample. The differences in the ICS attributable to the three comparison groups of proxy nonrespondents are shown in Table 2. The differences represent OLS regression estimates that also included controls for trends over time in the ICS.[5] Respondents that initially refused to be interviewed were not as optimistic, while respondents that required more calls to contact were more optimistic. Interestingly, the absolute values of the differences were similar, ranging from 2.5 to 3.0 Index points (see columns 1 to 3). Moreover, the data clearly indicate that nonrespondents are not all alike, and the impact from efforts to reduce nonresponse will critically depend on which types or sources of nonresponse are the focus of additional efforts.

Since the interviews with respondents that initially refused took more calls to complete (8.2 versus 5.2, on average), the regressions in columns 4-5 are joint estimates of the differences due to the combination of refusals and limited calls. When jointly estimated, the size of the estimated differences both increased in absolute value by about one-half to one Index point. Overall, the data clearly indicate the presence of a significant nonresponse bias if the sample was restricted to exclude any of the three proxy nonresponse groups.

[5] No statistically significant differences were found in the variances in the ICS between the cases designated as proxy nonrespondents and the remaining respondents for each of the three partitions.

Table 2. Differences in Index of Consumer

Sentiment Across Comparison Groups	OLS Regression Coefficients (Standard errors)				
	(1)	(2)	(3)	(4)	(5)
Initial Refusal=1	-2.9 (0.4)			-3.7 (0.4)	-3.9 (0.4)
6 or more calls=1		2.5 (0.3)		3.0 (0.3)	
3 or more calls=1			3.0 (0.3)		3.5 (0.3)

Note: Dependent variable was ICS; regressions also included time trend variables and a constant.

Table 3. Adjusted Differences in Index of Consumer Sentiment Across Comparison Groups

Groups	OLS Regression Coefficients (Standard errors)				
	(1)	(2)	(3)	(4)	(5)
Initial Refusal=1	-0.1 (0.4)			-0.5 (0.5)	-0.6 (0.5)
6 or more calls=1		1.2** (0.3)		1.2** (0.3)	
3 or more calls=1			1.5** (0.3)		1.6** (0.3)

Note: Dependent variable was ICS; regressions also included age, education, log income in constant dollars, gender, race, region, time trend variables and a constant. **=p<.001

It is important to note that the data were not weighted to account for differential nonresponse or other factors as is the usual practice.[6] The published figures for the Index of Consumer Sentiment are also based on the full sample, including the panel portion of the rotating sample design. The impact of the full sample weights on these differences was estimated by using the same variables that are used in the weighting procedure as control variables in the regressions. As shown in Table 3, when the respondents' economic and demographic characteristics are entered as control variables, the differences in the ICS that were associated with initial refusals were reduced to insignificance. Respondents that were harder to contact, however, still recorded

[6] All of the analysis reported in this paper is based on raw data. The published ICS is computed from data post-stratified to Census demographic totals as well as weighted to reflect differential selection probabilities (due to variation in household size and number of residential phone lines).

significantly higher Index values, although the difference was approximately cut in half when the demographic controls were included in the regressions. Also note that the bias was still slightly larger when more cases were excluded under the two-call design.

Economic and demographic controls reduced the nonresponse differences, because refusal conversion cases were more likely to be of lower socioeconomic status (and expressed less optimism about their economic situation), and the harder-to-contact cases were more likely to have higher socioeconomic status (and expressed more optimism about their economic prospects). Separate analyses were conducted to determine whether the probability of an initial refusal or the probability that the interview took more than 3 or 6 calls to complete was significantly associated with the economic and demographic characteristics of respondents. Logistic regressions indicated that younger respondents required more calls to complete the interview, while older respondents were more likely to initially refuse; higher income households were harder to contact, but lower income household were not more likely to initially refuse to participate; female respondents required fewer calls to complete the interview and were more likely to initially refuse; and more educated respondents required fewer calls to complete the interview and were less likely to initially refuse. The overall associations were quite small, however, with pseudo r-squares in the 5% to 8% range.[7]

Aside from the economic and demographic characteristics of the individual respondents, nonresponse may also be influenced by more general factors that reflect changes in the overall economic, social, or political environment. For example, (1999) found that changes in the aggregate environment also influenced people's willingness to participate in the Current Population Survey, conducted by the U.S. Census Bureau. Indeed, the Harris-Kojetin and Tucker found that changes in the Index of Consumer Sentiment were related to changes in refusal rates, in particular when the aggregate level of optimistic about economic prospects increased, the refusal rate was also likely to increase. When the likelihood of participation is itself a function of the survey variable of interest—bias can be relatively high even with low nonresponse rates since the difference in means between the observed and unobserved cases can be quite large.

Following the analysis by Harris-Kojetin and Tucker, *changes* in the initial refusal rate were related to *changes* in the ICS, and similarly to *changes* in the proportion that required more than five or two calls were related to *changes* in the ICS. Both month-to-month as well as quarter-to-quarter changes were examined over the 1979 to 1996 time period. As shown in Table 4, all of the correlations were quite small, none being significantly different than zero. To be sure, as with all of the analyses reported in this paper, these results only indicate that marginal increases in nonresponse above the prevailing levels showed no correspondence. Nonetheless, it could have been reasonably expected that marginal

[7] For the logistic regressions the dependent variables were defined as the probability of an initial refusal, the probability that the interview required six or more calls, or the probability that the interview required three or more calls. All regressions included age, log income in constant dollars, education, gender, region, race, time trend variables; the regressions on refusal conversions also included controls for the number of calls, and the regressions on the number of calls included whether the respondent had initially refused.

increases in nonresponse would have corresponded with increases in optimism under the Harris-Kojetin and Tucker hypothesis.

Table 4. Correlations of Change in Nonresponse Rates and Change in the Index of Consumer Sentiment

Correlations of changes in ICS with changes in . . .	Monthly Change (n=210)	Quarterly Change (n=70)
Initial Refusals	-0.012	0.069
6 or more calls	0.010	0.108
3 or more calls	-0.026	0.017

Note: All insignificant at p<.10

Table 5. Correlations of Size of Nonresponse Bias and Trends in the Index of Consumer Sentiment

Correlations of $\left(ICS_t^{o*} - ICS_t^{u*}\right)$ with rates of . . .	Monthly Change (n=210)	Quarterly Change (n=69)
Initial Refusals	0.081	0.163
6 or more calls	0.023	0.040
3 or more calls	0.086	0.120

Note: All insignificant at p<.10

Another possibility mentioned earlier was that higher nonresponse rates could be associated with larger or smaller differences in the ICS. If the size of the difference was negatively correlated with the nonresponse rate, then the bias would decline as the nonresponse rate rose. The data provide no evidence that this is the case. Nor do the data indicate a statistically significant positive relationship. Overall, the data suggest that the size of the difference is relatively independent of the level of nonresponse for the ranges observed. It must be again noted that this analysis only pertains to the additional increase in nonresponse bias, not the total bias. As a result the data only indicate that any changes in the characteristics of respondents as the proxy nonresponse rate increases over the actual prevailing levels are not reflected in responses to the questions that are included in the ICS.

Impact of Nonresponse with Variations in Sample Size

The analysis of the pooled 72,424 cases does not provide a realistic assessment of nonresponse problems for the usual context in which these surveys are used. While the very large sample sizes provide a robust test of the presence of nonresponse bias, the sample sizes that are actually used to calculate the ICS are considerably smaller. To gauge the impact of variations in sample size, tests were conducted based on monthly, quarterly, and yearly observations. The monthly samples averaged 333 interviews (which is smaller than the full monthly samples since only the initial interviews are utilized in this paper.) When the independent monthly samples were pooled to quarters, the average sample size was 1,000, and when pooled to years, the average sample size was 4,000.

As shown in Table 6, the average of the monthly, quarterly, and annual estimates of the ICS were nearly identical, as would be expected. The precision of the estimates, of course, increased along with the sample sizes. Overall, the standard errors of the ICS estimates based on quarters were about half the size of the monthly figures, and the annual estimates were about half the size of the quarterly estimates.

The mean differences in the ICS between respondents and the proxy nonrespondents were relatively constant whether calculated by months, quarters, or years. For example, respondents that initially refused were about 3.4 to 3.5 Index-points lower, while those that required three or more calls were about 2.8 to 2.9 Index-points higher. Whether these differences reached the level of significance, however, did depend on the sample size. For example, just 11% of the monthly differences between respondents requiring six or more calls and those requiring five or fewer calls were significant at the 5% level, which rose to 23% for quarterly estimates, and 35% for annual estimates.[8]

The primary analytic focus is usually on the change rather than the level of the ICS. Each of change scores reported in Table 6 was calculated as the mean difference between two changes—for example, the difference between the change in the ICS for those that required five or fewer calls and the change in the ICS for those that required six or more calls. The estimates of the period-to-period change in the ICS were not constant over the different survey frequencies, reflecting the simple fact that change accumulates over time. While the size of the period-to-period change increased as the periods increased from months to years, the standard errors decreased due to the larger sample sizes. In comparison to the mean level differences, the mean differences for the change estimates were quite small, mostly less than one-tenth of an Index-point. Notably, in nearly all cases the proportion of differences in change scores that were significant was very close to the 5% that would be expected by chance. The one exception was the difference between more or less than six calls based on the yearly samples. Overall, these results indicate that the level of nonresponse bias remains relatively constant across months and quarters, and as a result had little if any impact on estimates of change in the ICS.

---

[8] A more detailed discussion of these results in contained in Curtin, Presser, and Singer (forthcoming).

Table 6. Estimates of Impact of Nonresponse Based on Within Sample Comparisons

	Survey Frequency		
	Monthly (n=211)	Quarterly (n=70)	Yearly (n=17)
Overall Mean Level (Standard error)	84.2 (1.93)	84.2 (1.12)	84.8 (0.56)
Cooperators vs. Initial Refusals [1)]	3.45 (10%)	3.42 (13%)	3.35 (47%)
1-5 Calls vs. 6 or more calls [1)]	-2.33 (11%)	-2.32 (23%)	-2.86 (35%)
1-2 Calls vs. 3 or more calls [1)]	-2.88 (11%)	-2.87 (23%)	-2.86 (77%)
Overall Mean Change (Standard error)	0.14 (2.74)	0.47 (1.59)	1.79 (0.80)
Cooperators vs. Initial Refusals: [2)]	-0.40 (4%)	0.03 (0%)	-0.09 (0%)
1-5 Calls vs. 6 or more calls [2)]	-0.02 (6%)	-0.01 (7%)	-0.19 (31%)
1-2 Calls vs. 3 or more calls [2)]	-0.01 (4%)	-0.01 (6%)	-0.07 (6%)

1) Mean Differences in Levels (Proportion of surveys with significant mean differences at 5% level)

2) Mean Differences in Changes (Proportion of surveys with significant mean differences at 5% level)

Impact of Nonresponse Bias on Estimates from Independent Samples

The prior analysis was rooted in cross-section tests of nonresponse bias within one sample. Even the replication of the tests across time essentially relied on multiple within sample comparisons. It is of some interest to determine the impact of nonresponse bias between two independent samples which differ significantly in their nonresponse rates. The most convincing evidence of the impact of nonresponse would be derived from a true experimental design, for example, by conducting two surveys simultaneously over time that were methodologically identical in all respects except in the amount of effort expended to contact and convince respondents to participate.[9]

[9] An example of a cross-section experiment was conducted by Keeter et al. (1999). The study compared estimates from the same omnibus questionnaire administered using two different designs that differed in the length of time allotted to complete the study: one was conducted over five days and the other conducted over two months. The five-day survey had a response rate of 37%, well below the 61% response rate for the two-month survey. The study found very few statistically significant differences in the results across a large set of demographic, behavioral, attitudinal, and knowledge items.

In the absence of such a true experimental design, this approach was simulated by dividing each monthly sample into two random subsamples, using one as the base sample and the other to simulate the results that would be obtained from reduced effort. In order to achieve two equal sized subsamples, the random allocation was done separately for each type of simulated nonresponse, using the actual rate observed in each monthly survey. For example, suppose 12.5% of the cases in a monthly survey initially refused. The random allocation to the base sample (which included initial refusals as well as cooperators) would equal 46.7% of the total. The independent comparison sample would be randomly allocated 53.3% of monthly cases, so that when the 12.5% initial refusals were eliminated, the size of the comparison group would be reduced to 46.7% of the original monthly sample.[10] Since the use of subsamples cuts the monthly samples to less than half, the analysis was restricted to quarterly and half-year estimates.

Defining the first independent sample to include all cases, and the second to exclude the proxy nonresponse cases, estimates of the time-series relationship given by

$$ICS_{1t}^{o} = \alpha + \beta \ ICS_{2t}^{o*}$$

would indicate any systematic divergences due to the nonresponse bias in the means of the second sample. The appropriate test for the presence of bias would be to determine whether ∀=0 and ∃=1. Table 7 shows the results for the three proxy nonresponse groups, each calculated for quarterly, half year, and yearly frequencies. The results overwhelmingly rejected the hypothesis that the samples which excluded the proxy nonrespondents produced biased estimates in either the level or change in the ICS. In no regression was the constant term significantly different than zero, and in only two regressions was the beta coefficient significantly different than 1.0. The two exceptions were for the quarterly change regressions, when the restricted sample excluded the initial refusals or interviews that took more than two calls to complete. In both cases, the underlying cross-section samples were quite small - from 250 to 475 cases. Perhaps even more impressive was the proportion of the time-series variance that could be accounted for by the restricted samples: in nearly three-fourths of the regressions, the r-squared was above 0.90 and more than half were above 0.95. The clear exceptions were for the quarterly change regressions based on the smallest cross-section samples.

These results indicate that the nonresponse bias was trivial under nearly all circumstances. How can these results be reconciled with the cross section tests that indicated the presence of a significant bias? The cross-section analysis tested whether the difference in the ICS between the respondents that cooperated and those that initially refused, for example, was significant. In terms of the prior notation, this meant that ICS^{o*} significantly differed from ICS^{u*}. The expected value of the bias, however, depends not only on this difference but also on the proportion of nonresponse in the total sample.

---

10 Define r_t as the rate of initial refusals (or interviews that required more than two or five calls). The proportion allocated to the base sample was (1 - r_t) / (2 - r_t), and the proportion included in the comparison sample was equal to (1 - (1 - r_t) / (2 - r_t)), which was then reduced by (1 - r_t) when the proxy nonresponse cases were eliminated from the subsample.

Focusing only on the additional bias introduced by restricting the sample to exclude the proxy nonrespondents, the expected value of the bias is:

$$Bias\left(ICS_t^{o*}\right) = \pi_t^*\left(ICS_t^{o*} - ICS_t^{u*}\right)$$

Table 7. Impact of Nonresponse Bias on Comparisons Between Two Independent Samples

Survey Frequency	# Time Periods	# Cases Per Period	Level Regressions			Change Regressions		
			∀	∃	Rsqd-adj	∀	∃	Rsqd-adj
A) All call sample predicted by sample with no initial refusals								
Quarterly	70	475	1.624	0.975	.957	0.026	0.836[b]	.590
			(2.119)	(0.025)		(0.434)	(0.084)	
Half-Year	35	950	0.767	0.985	.985	0.074	0.940	.873
			(1.778)	(0.021)		(0.358)	(0.062)	
Yearly	18	1900	0.993	0.983	.993	0.0001	0.991	.958
			(1.747)	(0.021)		(0.346)	(0.052)	
B) All call sample predicted by 1-5 call sample								
Quarterly	70	400	0.020	1.011	.960	0.031	0.860	.604
			(2.101)	(0.025)		(0.429)	(0.084)	
Half-Year	35	800	-1.418	1.028	.990	0.013	1.018	.922
			(1.512)	(0.018)		(0.288)	(0.052)	
Yearly	18	1600	-1.323	1.027	.993	0.005	1.027	.957
			(1.737)	(0.021)		(0.353)	(0.055)	
C) All call sample predicted by 1-2 call sample								
Quarterly	70	250	3.442	0.982	.928	0.058	0.690[b]	.460
			(2.743)	(0.033)		(0.515)	(0.090)	
Half-Year	35	500	0.865	1.012	.970	0.061	0.934	.783
			(2.547)	(0.031)		(0.490)	(0.085)	
Yearly	18	1000	0.733	1.014	.985	0.094	0.975	.914
			(2.506)	(0.030)		(0.517)	(0.075)	

a=significantly different than 0.0 at $p < .05$; b=significantly different than 1.0 at $p < .05$.

Note: Case counts varied depending on the actual proportions of proxy nonresponses observed. Case counts give average numbers in each of the independent subsamples for the indicated periods. Each of the paired independent samples were randomly selected to be equal in size.

The analysis conducted on the independent random samples tested whether the product of the proxy nonresponse rate and the difference was significant, while the cross-section tests focused only on the significance of the difference. Since the proxy nonresponse rate ranged from about 10% to 50%, the effective size of the bias when comparing two independent samples was reduced by those same proportions. Given that the difference was relatively small - too small to be significant in a majority of the cross-section tests reported in Table 6 - it is not surprising that the difference weighted by the proxy nonresponse rate was nearly always reduced to insignificance.

It may be useful to describe how these results may explain a rather common observation when comparing the results of two different surveys. One survey achieves high response rates, and can demonstrate that more restrictive procedures, such as not attempting to convert initial refusals or by limiting the number of attempts to contact eligible respondents, would result in a significant bias. Another survey targets a much lower response rate and claims that the results from their surveys rarely differ from the other more rigorous and more expensive procedures. The simple truth, at least for measures like those that are included in this analysis, is that both claims may be correct.

Summary and Implications

The strong preference for high response rates is based on sampling theory. Probability samples, which assign a nonzero chance of selection to every member of the population, provide the means to draw inferences about the entire population from the small subset selected for interviews. This inferential capability depends on achieving response rates that are high enough to insure that the realized sample accurately reflects the selected sample. Since sampling theory provides no mechanism to judge what would be the lowest acceptable response rate, any decrease in response rates must be regarded with suspicion.

The unfortunate fact of survey research, however, is that refusal rates are rising despite strenuous and expensive efforts to reverse the trend. To be sure, devising more effective methods to reduce nonresponse represents the optimum strategy. This absolute preference for high response rates reflects as much the strength of sampling theory as the absence of any comprehensive theory of nonresponse. It is this lack of a theoretical model of nonresponse that limits our ability to generalize the findings in this paper about the potential bias resulting from higher nonresponse rates in other surveys or about other topics. Nonetheless, consumer confidence is a widely measured economic indicator worldwide, and the potential nonresponse bias is of considerable interest in its own right.[11] Moreover, while there is nothing more useful than good theory, there is perhaps nothing more productive of theoretical developments than data that helps clarify and quantify the underlying issues.

A summary of this paper's analysis must begin with the clear recognition of the limited range of nonresponse that was studied. No information was available on the "true" nonrespondents to the Surveys of Consumers. Rather, the study focused on the bias induced from marginal additions to the prevailing levels of nonresponse. The three groups

[11] Other countries that conduct consumer surveys include Austria, Australia, Belgium, Canada, China, Czech Republic, Denmark, Estonia, Finland, France, Germany, Great Britain, Greece, Hungary, Ireland, Italy, Japan, Netherlands, Norway, Poland, Russia, Spain, South Africa, Sweden, Switzerland, and Taiwan.

of proxy nonrespondents were identified based on whether they initially refused or were difficult to contact. The presumption was not that these cases were similar to the "true" nonrespondents, but that they would be similar to the nonresponse cases in the range of simulated response rates. The proxies were simply assumed to be similar to the type of nonresponse cases that would be observed at the margin, all other things being equal. All other things are unlikely to be completely equal, however. Surveys designed with five-call limits, for example, are likely to strictly control the day and time of each of the five calls. Nonetheless, in the absence of a true experimental design, the use of the proxy nonrespondents does provide a reasonable approximation of potential nonresponse bias.

The analysis presented in this paper gave clear evidence of nonresponse bias. Consumer confidence was found to be significantly lower among initial refusals, while confidence was significantly higher among respondents that were harder to contact. The bias was found to be related to the economic and demographic characteristics of the respondents, with initial refusals more likely to have lower socioeconomic status, and the harder-to-contact cases more likely to have higher socioeconomic status. While the difference in socioeconomic status was plausibly related to the observed differences in optimism about economic prospects, economic and demographic characteristics alone could only account for a small share of the variance in the probability of being a nonrespondent. Nonetheless, the bias due to initial refusals was reduced to insignificance by controlling for differences in the type of economic and demographic characteristics usually incorporated into sample weights. The bias from the exclusion of respondents that were difficult to contact was greatly reduced but not eliminated by those same control variables. As a result, it is likely that sample weights have an asymmetrical impact, eliminating the refusal bias but not the bias from failing to reach respondents that are harder to contact.

Three general characteristics of the nonresponse bias served to limit its impact on the Index of Consumer Sentiment. First, the overall bias was relatively small. Second, the data provided no evidence that the likelihood of participation in the survey was itself a function of the prevailing level of consumer confidence, what is sometimes termed non-ignorable bias. Third, the size of the nonresponse bias did not systematically vary with the proportion of nonrespondents within each of the proxy nonresponse groups.

The small size of the bias meant that variations in cross-section sample sizes were critical to whether the bias in the estimated *level* of the ICS proved significant. Only when samples approached 4,000 was the bias likely to be found significant in half the samples tested. Of greater analytic importance for measures of consumer confidence, the relatively constant bias meant that the estimated period-to-period *change* in ICS was found to be unaffected regardless of the sample size. Given that the major use of measures of consumer confidence is for the analysis of change, the data confirm the well-known methodological advantages of trend surveys.

Finally, it was shown that independent samples that differed to a considerable degree in nonresponse gave essentially equivalent estimates of trends in both the level and change in the ICS over time. The equivalence was due to the combination of a small nonresponse difference weighted by the fractional increase in nonresponse rates. Perhaps more than any of the other results, this analysis appears to indicate that nonresponse bias has little if any practical impact on estimates of the ICS.

Does nonresponse really not matter for measures of consumer confidence? The answer is more complex than the results suggest. To be sure, the analysis suggests that at least for some types of surveys, the tradeoff between survey costs and response rates might be reconsidered. The analysis, however, provides little guidance on how to best implement such a deliberate strategy of accepting higher nonresponse. Nonresponse is the result of many different behaviors on the part of respondents as well as survey organizations. Exactly which behaviors should be modified, and to what extent, is not entirely clear. Not all sources of nonresponse may be as benign as those investigated in this study. More research is needed on the factors that determine each source of nonresponse, and how each source differs in terms of its potential bias. Perhaps the most important caveat stems from what was excluded from this analysis, namely, the unobserved bias already incurred at the prevailing 70% response rate. Rather than tempting us to lower costs and lower response rates, that missing analysis might just as strongly confirm the wisdom that high response rates should be our highest priority.

References

Cochran, W. Sampling Techniques. New York: Wiley; 1977.

Curtin, Richard, Presser, Stanley, and Singer, Eleanor. The effects of changes in response rate on the Index of Consumer Sentiment, Public Opinion Quarterly, forthcoming.

de Herr, Wim. International response rate trends: Results of an International Survey. Journal of Official Statistics. 1999 Jun; 15(2):129-142.

de Leeuw, Edith, editor. Journal of Official Statistics. Special issue on Survey Nonresponse. Vol. 15, 1999.

Groves, Robert M. and Couper, Mick P. Nonresponse in Household Interview Surveys. New York: Wiley; 1998.

Harris-Kojetin, Brian and Tucker, Clyde. Exploring the relation of economic and political conditions with refusal rates to a government survey. Journal of Official Statistics. 1999 Jun; 15(2):167-184.

Steeh, C. G. Trends in Nonresponse, 1952-1979. Public Opinion Quarterly. 1981; 45:40-57.

Synodinos, Nocolaos E. and Yamada, Shigeru. Response rate trends in Japanese surveys. International Journal of Public Opinion Research. 2000; 12(1):48-72.

Appendix

Index of Consumer Sentiment

The Index of Consumer Sentiment is based on the answers to five questions:

- We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?
- Now looking ahead – do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?
- Now turning to business conditions in the country as a whole – do you think that during the next 12 months we'll have good times financially, or bad times, or what?
- Looking ahead, which would you say is more likely – that in the country as a whole we'll have continuous good times during the next 5 years or so, or that we will have periods of widespread unemployment or depression, or what?
- About the big things people buy for their homes – such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or a bad time for people to buy major household items?

The basic formula for the Index of Consumer Sentiment is:

$$ICS_t = \sum_{j=1}^{5}\left(P_{jt}^{f} - P_{jt}^{u}\right) \quad 100 + 100$$

where

P_{jt}^{f} = the sample proportion giving favorable replies to the j^th question at time t

P_{jt}^{u} = the sample proportion giving unfavorable replies to the j^th question at time t.

Equivalently, the formula can be expressed in terms of the individual responses:

$$ICS_t = \sum_{J=1}^{5} \quad \sum_{i=1}^{n}\frac{x_{ijt}}{n}(100) + 100$$

where

X_{ijt} = 1 if favorable response to j^th question by i^th respondent at time t,

X_{ijt} = -1 if unfavorable response to j^th question by i^th respondent at time t,

X_{ijt} = 0 for all other responses to j^th question by i^th respondent at time t.

The final figures are published as a proportion of the base year value (1966).

Quantification of Qualitative Data Using Ordered Probit Models

With an Application to a Business Survey in the German Service Sector

Ulrich Kaiser and Alexandra Spitz

Abstract

This paper[1] aims at providing business survey analysts with simple econometric tools to quantify qualitative survey data. We extend the traditional and commonly applied method proposed by Carlson and Parkin (1975) to capture observable survey respondent heterogeneity. We also discuss specification tests.

The empirical analysis is based on business survey data taken from the ZEW‘s "Service Sector Business Survey", a quarterly business survey in the German business-related service sector carried out since 1994.

1 Introduction

Whenever present day information on the development of an economy or parts of the economy is missing, the public interest in information gathered from business surveys receives heightened attention. A major advantage of business surveys is that first results can usually be published within three months after the data collection period has ended. Many economists, such as Oppenländer (1997), claim that this up–to–dateness makes business surveys at least as important as official statistics. A synoptic table provided by the Centre for International Research on Economic Tendency Surveys office (CIRET, 1998) highlights the influence of business surveys: while there were 34 surveys in 15 countries collected in 1960, the number increased to 318 surveys in 57 countries by the end of 1997.

The informational content of business surveys is, however, often limited. Most surveys simply ask questions on the state of the respondents' business on an ordinal scale. A frequent question is, for example, "Did your total sales increase, decrease or remain the same in the current quarter with respect to the preceding quarter". In order to aggregate the information contained in the individual responses, balances – the share of

[1] This research has been inspired by discussions with Robert Dornau and Winfried Pohlmeier. Ulrich Kaiser gratefully acknowledges financial support by the German Science Foundation (Deutsche Forschungsgemeinschaft, DFG) under grant PF331/3-3.

firms reporting increased sales minus the share of firms reporting decreased sales – are calculated. In addition to the more formal aspect that the information contained in the "no change" category is neglected,[2] people may find it difficult to assess the implication of a balance of 20 percent, for example. In particular, if a time dimension is lacking, it is difficult to assess whether this value signals confidence or stagnation.

A quarter of a century ago Carlson and Parkin (1975) developed a fairly simple technique to quantify the qualitative information collected in business tendency surveys. Their method has been extended in many different ways; comprehensive surveys are presented by Geil and Zimmermann (1996), Seitz (1988) and Zimmermann (1985 and 1997).

In this paper, we suggest a simple alternative to the basic Carlson and Parkin (1975) procedure, which has several advantages with respect to "direct" tests for the crucial assumption of normality and with respect to the incorporation of individual–specific variables that allow control for observed survey respondent heterogeneity.

This paper also introduces a comparatively new dataset, the "Service Sector Business Survey" (SSBS) to the literature. The SSBS is a quarterly business survey that is collected by the Centre for European Economic Research (Zentrum für Europäische Wirtschaftsforschung, ZEW) in cooperation with Germany's largest credit rating agency Creditreform since June 1994.[3] Roughly 1,100 firms of the fast growing German business–related services sector regularly take part in the SSBS. The SSBS is unique in the sense that it provides information on an increasingly important part of the German economy that is substantially underrepresented in official statistics. Herbert Hax (1998), the chairman of the German Council of Economic Advisors, recently criticised the lack of appropriate data on the service sector that severely hampers business cycle forecasts and economic policy advice. The lack of data for the observation of business cycles in the German business–related service sector appears to even more severe since Kaiser and Voss (2000) have shown, using Granger causality analysis, that manufacturing generally does not lead business–related services and manufacturing in the business cycle. That inadequate data availability on services is not only a particular German problem but also a worldwide problem, as has been stressed by Waller (1997).

We aim at closely linking quantification methodology with practical implementation and, hence, start by describing a somewhat "typical" business survey, the SSBS, and proceed with a discussion of quantification methods. Finally, we present quantification results and perform specification tests.

Our discussion focuses on the standard ordered Probit model. Although it is well established that quantification in an ordered Probit context is feasible and simple, it is scarcely applied in practice. In this paper we demonstrate that it is worthwhile to use the ordered Probit model for quantification since the inclusion of respondent–specific variables – which is infeasible in the Carlson and Parkin (1975) method – helps to increase the precision of the estimates and substantially reduces the width of the confidence bounds that correspond to the quantified survey results.

---

[2] See Ronning (1984, 1990) for details on this issue.

[3] Details on the sample design and the survey design are given in Kaiser et al. (2000).

2 Data

The SSBS has steadily gained in terms of media attention since its implementation in the second quarter of 1994. It focuses on ten branches of the service sector, which are often referred to as "business–related services". Although no clear–cut and broadly accepted definition of business–related services exists, researchers have agreed upon definitions based on the enumeration of certain sectors. Our definition of business–related services closely follows Hass (1995), Klodt et. al. (1997), Miles (1993) and Strambach (1995). It is displayed in the table below with the corresponding German industrial classification WZ93.[4]

Sector	WZ 93
Computer Services	72100, 72201–02, 72301–04, 72601–02, 72400
Tax consultancy & Accounting	74123, 74127, 74121–22
Management Consultancy	74131–32, 74141–42
Architecture	74201–04
Technical Advice & Planning	74205–09, 74301–04
Advertising	74844, 74401–02
Vehicle Rental	71100, 71210
Machine Rental	45500, 71320, 71330
Cargo handling & Storage	63121, 63403, 63401
Waste and Sewage Disposal	90001–07

Every three months, ZEW and Creditreform send out a single page questionnaire to about 3,500 firms that belong to the ten sectors listed above. The survey is constructed as a panel data set and currently covers 25 waves. It is a stratified random sample, stratified with respect to the ten sectors, five size classes (two for Eastern Germany and three for Western Germany), and regional affiliation (Eastern/Western Germany). The stratified target population thus consists of 50 cells. A sample questionnaire, the questionnaire of the first quarter of 2000, is printed in on the last page of this paper.

A sample refreshment takes place on an annual basis. Firms that have not taken part in the survey for more than six times in a row are removed from the panel. First survey results of the study and a general description of the survey can be found in Saebetzki (1994). Current survey results are released in the media and in ZEW publications.[5]

The SSBS starts three weeks prior to the end of a quarter. Questionnaires and a personal letter to the prospective survey respondent are sent out by mail. The questionnaires are mostly returned to the ZEW by fax. After two weeks, those firms that

[4] The WZ93 industrial classification code is a classification system developed by the German Federal Statistical Office in accordance with the European NACE Rev. 1 standard that classifies economic units according to their sector of concentration.

[5] The ZEW sends current survey results to an interested public. Send an email to konjunkturumfrage@zew.de to receive copies.

have not replied are sent a reminder. Altogether, the response rate amounts to about 30 percent. As a thank you for filling out the questionnaire, the participating firms receive an analysis in the form of a four page report that contains the main findings of the survey. In addition, they can obtain further information over the Internet.[6]

The questionnaire is divided into two parts. In the first part, the firms are requested to indicate on a three point Likert scale whether their sales, prices, demand, returns and number of employees have decreased, stayed the same, or increased in the current quarter in comparison to the previous quarter. Moreover, they are supposed to give an assessment for the forthcoming quarter. The second part of the survey is dedicated to current economic and political issues. Topics cover on–the–job–training, wage negotiation and dispersion of general wage agreements (Kaiser and Pfeiffer 2000; Kaiser and Pohlmeier 2000), innovation and the demand for heterogeneous labour (Kaiser, 1998a), the adjustment to demand fluctuations (Kaiser and Pfeiffer, 2000) and the implications of the introduction of the Euro on firms' export propensity (Kaiser and Stirböck, 1999).[7]

A detailed description of the data set is presented by Kaiser et al. (2000). An overview and selected survey results are reviewed in Kaiser (1999).

3 Quantification Methods Reconsidered

People in charge of collecting business survey data are often hesitant to ask directly for sales, prices, profits, demand or employment. In practice, survey respondents are asked to give a qualitative assessment on their business development on a three or five point Likert scale. There are three main reasons for proceeding this way instead of asking for quantitative assessments. First, firms may be reluctant to report actual figures due to privacy reasons. Second, an inherent risk of asking quantitative questions is that there is a high potential of ending up with information with "spurious precision", for example respondents may be either unable to report precise figures or they may purposely misreport the actual figures. The third reason may be the most compelling one in terms of practical relevance: it is simply easier and faster to give qualitative instead of quantitative assessments. Asking ordinal questions helps to save the respondents' time and hence helps to improve the total response rate.

When survey respondent i answers questions on an ordinal scale, she implicitly has a threshold model in mind. She indicates increased ("+") sales if the actual change in sales, hereafter abbreviated by Y^*, is above a certain threshold μ_2. Likewise, if the actual change in sales is below a lower threshold μ_1, she indicates decreased sales ("–"). If the actual change is between the two thresholds, she reports unchanged ("=") sales figures. Clearly, these thresholds may vary across different survey respondents or groups of survey

6 The Internet address is: http://www.zew.de/aktuell/branchenreport/wb-BreportStart.html

7 In a related study, Kaiser (1998b) analyzes the impact of political events on answering patterns in business surveys.

respondents and also across time. In order to clarify things, it is useful to write the threshold model formally as:

$$Y_i = \begin{cases} + & if \quad Y_i^* > \mu_2 \\ = & if \quad \mu_1 < Y_i^* \leq \mu_2 \\ - & if \quad Y_i^* \leq \mu_1 \end{cases}, \tag{1}$$

where Y_i denotes the qualitative sales assessment of respondent i. Let N^+, $N^=$ and N^- denote the number of individuals who report increased, unchanged and decreased sales figures, respectively, and let N denote the total number of survey respondents. Then the relationship between the choice probabilities and the answering shares can be summarised by the following system of equations:

$$\begin{aligned} P[\,Y_i = \text{'+'}\,] &= P[\,Y_i^* > \mu_2\,] = N^+/N \\ P[\,Y_i = \text{'='}\,] &= P[\,\mu_1 < Y_i^* \leq \mu_2\,] = N^=/N \\ P[\,Y_i = \text{'–'}\,] &= P[\,Y_i^* \leq \mu_1\,] = N^-/N \end{aligned} \tag{2}$$

That is, the empirical probabilities to indicate increased, unchanged or decreased sales are simply equal to the shares of the respective answers.

The system of equations (2) nicely illustrates that a straightforward and simple nonparametric, e.g. distribution and parameter–free, estimator for the probability to report increased, unchanged or decreased sales simply is the share of answers for these categories.

In order to quantify qualitative information, a distributional restriction concerning the choice probabilities $P[\cdot]$ has to be imposed. Let the actual sales changes Y_i^* be dependent on a constant term, β_0 and an identically and independently distributed error term ε_i which follows a distribution function $F(\cdot)$ with mean zero and variance σ^2: $Y_i^* = \beta_0 + \varepsilon_i$. The choice probabilities $P[\cdot]$ are hence given by:[8]

$$\begin{aligned} P[\,Y_i = \text{'+'}\,] &= 1 - F(\tfrac{\mu_2-\beta_0}{\sigma}) \\ P[\,Y_i = \text{'='}\,] &= F(\tfrac{\mu_2-\beta_0}{\sigma}) - F(\tfrac{\mu_1-\beta_0}{\sigma}) \\ P[\,Y_i = \text{'–'}\,] &= F(\tfrac{\mu_1-\beta_0}{\sigma}) \end{aligned} \tag{3}$$

The choice of the distribution function, often also referred to as the "link" function is arbitrary provided that it is symmetric. However, one must test if the distributional assumption is correct. Common choices are the normal and the logistic distribution. The

[8] Since, e.g., $P[Y_i = \text{'+'}] = P[\beta_0 + \varepsilon_i > \mu_2] = P[\varepsilon_i > \mu_2 - \beta_0] = 1 - P\left[\frac{\varepsilon_i}{\sigma} < \frac{\mu_2-\beta_0}{\sigma}\right] = 1 - F\left[\frac{\mu_2-\beta_0}{\sigma}\right]$.

normal distribution leads to the ordered Probit model and the logistic distribution leads to the ordered Logit model.[9] In this paper we shall consider the normal distribution only since this is the distribution function considered by Carlson and Parkin (1975).[10] Since increases in β_0 and σ such that the ratio β_0/σ remains constant does not affect either probability and since changes in the parameter corresponding to the constant term in the mean function and in the thresholds such that their distance remains unchanged also do not affect the probabilities, identification restrictions have to be imposed. Standard software packages such as LIMDEP and STATA both set σ to one. LIMDEP furthermore restricts the first threshold parameter to zero and estimates a constant term in the mean function while STATA sets the coefficient of the constant term to zero and estimates all threshold parameters.

If both thresholds are known, the constant term in the mean function β_0 and the standard deviation of the error term σ can be estimated. In this case, quantification by an ordered Probit model with known thresholds and the Carlson and Parkin (1975) approach are exactly identical. In fact, such an ordered Probit model *is* the Carlson and Parkin method expressed in an alternative way. In the ordered Probit context, the estimated parameter $\hat{\beta}_0$ denotes the quantified sales growth rate and the estimated parameter $\hat{\sigma}$ denotes the standard error of the quantified sales growth rate.

An extension of this basic quantification method that uses ordered Probit models for one single survey to repeated surveys is straightforward. Let t denote the point in time in which individual i and its survey response is observed and let D_{it} denote a dummy variable that is coded "one" if individual i took part in the t^{th} survey. In order to find quantified sales changes for each of the $t = 1,...,T$ survey waves in an ordered Probit context, the latent variable is specified as $Y_{it}^* = \sum_{t=1}^{T} \alpha_t D_{it} + \varepsilon_{it}$. The threshold model is then given by:

$$Y_{it} = \begin{cases} + & if & & Y_{it}^* & = & \sum_{t=1}^{T} \alpha_t D_{it} + \varepsilon_{it} > \mu_2 \\ = & if & \mu_1 < & Y_{it}^* & = & \sum_{t=1}^{T} \alpha_t D_{it} + \varepsilon_{it} \leq \mu_2 \\ - & if & & Y_{it}^* & = & \sum_{t=1}^{T} \alpha_t D_{it} + \varepsilon_{it} \leq \mu_1 \end{cases} \quad (4)$$

The constant term β_0 is now made wave–specific by the inclusion of the dummy variables D. Estimates of the α_t*'s* represent the quantified sales changes at time t.

9 A discussion of whether ordered or unordered models are appropriate in this context is provided by Ronning (1990).

10 Choosing either the logistic or the normal distribution merely is a matter of convenience since the distributions are very similar to one another with the logistic distribution having more mass at the tails. It is therefore advisable to consider the logistic distribution instead of the normal distribution if the extreme choice categories, in this case "+" and "-" are heavily populated. The choice of the normal distribution by Carlson and Parkin (1975) was the source of wide criticism, e.g. see Maddala (1990).

Estimates for the standard error of the quantification can be obtained by specifying the standard error of the disturbance term as $\sigma_{it} = \exp(\sum_{t=1}^{T} \gamma_t D_{it})$, where γ_t are the estimated parameters.[11]

As opposed to the linear regression model in which the estimated parameters retain their consistency even when the error terms are non–normal, not identical and not independent, the parameters of the ordered Probit model become inconsistent in these cases. Specification tests are therefore advisable though rarely used in applied econometric work. We will return to this issue after having presented quantification results in section 4.

The standard error of quantified ordinal information usually is much lower when survey respondents give an assessment on *overall* economic issues compared to the situation when they judge their *own* business condition. In both cases, the variance in the answers is attributable to heterogeneity across the survey respondents. However, though opinions on the state of the overall economy may of course differ among survey respondents, the deviation of judgements on the state of their *own* businesses are likely to be much larger. In fact, variations of these opinions may be dependent upon firm size, regional affiliation (Eastern/Western Germany) or sector affiliation. It is thus straightforward to incorporate these differences within the specification of the standard deviation of the error term.

Let SC_{ik} denote the k^{th} firm size class of respondent i, let *East* denote a dummy variable for Eastern German firms and let $Sector_l$ denote the l^{th} sector. The standard error of the disturbance term is then given by $\sigma_i = \exp(\sum_{t=1}^{T} \gamma_t D_{it} + \sum_{k=1}^{K-1} \delta_k SC_{ik} + \kappa \ East_i + \sum_{l=1}^{L-1} \zeta_l Sector_l = \exp(\boldsymbol{\theta z_i})$ for $i = 1,...,N$, where the k^{th} size class and the l^{th} sector are the reference groups. Likewise, it seems reasonable that the same set of variables affects not only the variation of individual responses but also the growth rate and thus the choice of the answering category so that $Y_{it}^* = \sum_{t=1}^{T} \alpha_t D_{it} + \sum_{k=1}^{K-1} \rho_k SC_{ik} + \tau \ East_i + \sum_{l=1}^{L-1} \varphi_l \ \ Sector_l + \varepsilon_{it} = \boldsymbol{x_i \beta}$.

The inclusion of the explanatory variables is equivalent to moving the threshold parameters μ around. This implies that if explanatory variables such as firm size and regional and sector affiliation are included in the specification, this is equivalent to specifying group–specific threshold parameters.

It is straightforward to obtain sector–specific sales growth rates for example by simply interacting the wave dummy–variables with the sector dummy–variables. The coefficients obtained from such an estimation reflect the wave–specific and sector–specific sales growth rates.

Another extension of the standard ordered Probit model as described in this section is the ordered panel Probit model. Many business surveys are constructed as panel data

---

[11] The exponential function is taken in order to avoid negative standard deviations.

sets and it seems advisable to explicitly use this additional information. The main advantage of panel data models is that unobserved heterogeneity of the individuals i can be taken into account. In this case, the error term ε_{it} is specified as the sum of two components: $\varepsilon_{it} = \xi_i + v_{it}$. The term ξ_i is assumed to be a time independent individual–specific random variable, reflecting unobserved firm heterogeneity while v_{it} is assumed to be an error term that is independent both among individuals and over time. Both error terms are assumed to be normally distributed with zero means.

The ordered Probit model, as discussed above, is a so–called "pooled" ordered Probit model. That is, we do not take into account the additional information contained in our panel data set by assuming the error term ε_{it} to be independent and identically distributed with a mean of zero and variance σ^2 for all individuals i and over time t.

Two principles for estimating panel data models exist: the "fixed effects" and the "random effects" approaches. Fixed effects estimation assumes the presence of an individual–specific effect ξ_i and independence of the error term component v_{it}. In this nonlinear specification, the fixed effects ξ_i and the coefficients α_t are unknown parameters and have to be estimated. In this case, the maximum likelihood estimator is only consistent when T tends to infinity. When T is finite, as is usually the case, the *incidental parameter problem* (Neyman and Scott, 1948) occurs: there is only a limited number of observations of Y_{it} for each individual i, $t = 1,...,T$, that contain information about ξ_i. Furthermore, an increase of the cross–sectional units, N, provides no information about ξ_i, but it increases the number of parameters ξ_i. The result is that any estimation of ξ_i is meaningless if T is finite, even if N is large. Unfortunately, the maximum likelihood estimators α_t and ξ_i cannot be separated in the nonlinear qualitative response models as is the case for linear models. When T is finite, the inconsistency of the estimated ξ_i is transmitted into the estimation of α_t. Chamberlain (1984) suggested an approach to remove the unobserved heterogeneity in multinomial Logit models.[12] Such an estimator does not exist, however, for ordered panel data models due to the existence of the threshold parameters.[13] Random effects estimation in the ordered Probit context is feasible, even in standard software packages such as LIMDEP. Instead of estimating N parameters ξ_i as in the fixed effect model, only the mean and variance are estimated. It only leads to efficiency gains if significant random effect are present, e.g., if the error components v_{it} are correlated over time. The pooled panel ordered Probit estimator,

---

12 This approach is based on a conditional likelihood approach proposed by Anderson (1970, 1973). The baseline idea is to remove the incidental parameters by writing the multinomial logit model in terms of a conditional maximum likelihood function. In probit models, the conditional maximum likelihood method does not remove the individual specific effects, however.

13 Also note that time–invariant variables such as sector or regional affiliation have to be removed from the specification since they are absorbed by the fixed effect.

however, retains its consistency.[14] For the sake of brevity, we will therefore not discuss the random effects ordered Probit model in further detail. Comprehensive discussions are presented by Hamerle and Ronning (1995) as well as Tutz and Hennevogl (1996). A recent application of the random effects ordered Probit model is presented in Kaiser and Pfeiffer (2000).

To summarise, quantification of qualitative survey data by ordered Probit models has two main advantages: (i) it allows for group–specific thresholds by the inclusion of explanatory variables and (ii) it allows one to explicitly take into account the variation of survey responses among the responding individuals. Further advantages are that tests for normality and heteroscedasticity can fairly easily be implemented and tests of identity of sales changes in individual quarters can be easily conducted by using a Wald test. The latter two topics will be discussed in further detail below.

4 Quantification Results

A key question in any quantification context is the derivation of the threshold values. Carlson and Parkin (1975) estimated thresholds by assuming long–term unbiasedness.[15] It is common practice to directly ask the survey respondents for the minimum value to which actual sales have to increase (decrease) before they report increased (decreased) sales figures once and then to assume that these values remain constant during the next couple of months or years.[16]

Proceeding this way is, however, not a sensible approach for the SSBS since this data set is not well balanced, for example the fluctuation of responding firms is quite large so that a considerable share of firms that has answered in survey wave t when it is asked for the individual thresholds is likely not to answer at $t+s$ and vice versa.[17]

Therefore, the threshold parameters were obtained from another data set which was also compiled by the ZEW, the Mannheim Innovation Panel in the Service Sector (MIP–S).[18] The MIP–S covers very similar sectors as the SSBS and has up to now been conducted four times, in 1995, 1997, 1998 and 1999. In 1997, the participating firms were asked to indicate on a five point scale whether their sales improved, remained unchanged or decreased within the last three years. Due to the panel structure of this data set, we were able to compare this qualitative assessment with the actual changes in total sales. We have calculated the median changes in sales – corrected to take into account that the SSBS asks for quarterly sales changes – for those firms that reported increased (decreased) sales figures as the upper (lower) threshold parameters. The respective value

[14] This is, in fact, a main reason that the application of the random effect model is scarce in the empirical literature.

[15] That is, they estimated the threshold on the basis of time-series of observed rates of buying-price inflation

[16] Threshold values for the well known ZEW Financial Market Test (for more information see http://www.zew.deprojekte.epl?action=detail&nr=6&lang=eng) are, e.g., obtained that way.

[17] See Kaiser et al. (2000) for more information on the stability of the panel data set.

[18] A thorough description of this data set is presented by Janz et al. (2000).

for μ_2 is 1.3 and the value for μ_1 is –0.5 percent. That is, we have found evidence for the presence of asymmetric thresholds: actual sales changes have to exceed a considerably higher threshold before firms report increased sales figures than the other way around. Besides the obvious psychological explanation that people tend to overstate bad economic or personal situations compared to good ones, Batchelor (1986) argues that individuals' answers may by subject to strategic behaviour, e.g. firms are more likely to report pessimistic results, in the hope of getting subsidies for their industry. Our finding of asymmetric thresholds supports the criticisms of the Carlson and Parkin (1975) approach, which assume symmetric thresholds.[19]

Table 1. Descriptive statistics of the actual sales changes reported in the SSBS

			Positive sales changes			
Wave	Minimum	10%	Median	90%	Mean	Std. dev.
20	1.1	4.0	10	20	11.4	7.5
21	0.9	3.0	10	20	11.2	7.3
22	1.0	4.0	10	20	11.0	6.7
23	1.1	3.5	10	20	11.2	7.0
24	1.3	3.0	10	21	11.3	7.3
25	0.9	3.0	10	20	10.9	7.5
mean	*1.1*					
			Negative sales changes			
Wave	Maximum	10%	Median	90%	Mean	Std. dev.
20	-0.5	-25.0	-10	-5.0	-13.7	8.0
21	-0.8	-24.0	-10	-4.2	-11.5	7.6
22	-0.9	-20.2	-10	-3.0	-11.6	7.5
23	-0.6	-20.0	-10	-3.9	-12.3	7.4
24	-0.6	-25.0	-10	-5.0	-13.9	8.0
25	-0.7	-25.0	-10	-5.0	-12.9	7.0
mean	*-0.7*					

In order to compare the thresholds derived from the MIP–S and the SSBS, Table 1 displays descriptive statistics of the actual sales changes reported by the firms interviewed in the SSBS since wave 20. The minimum value corresponding to the positive sales changes (upper panel) can be regarded as the bound *above* which firms indicate increased sales changes. The mean minimum (maximum) value of the actual sales changes

19 Other studies explain the existence of the "stay the same" category by considerations concerning the cost–intensive information acquisition process (Fishe and Idson, 1989).

reported by firms with increased (decreased) sales changes are 1.1 (-0.7) so that they compare well to our estimated thresholds of 1.3 and -0.5.

Table 2. Ordered Probit estimation results: baseline model

Conditional mean			Conditional variance		
	Coeff.	Std. err.		Coeff.	Std. err.
α_1	1.1784	0.0965	σ_1	2.0049	0.4749
α_2	1.0411	0.0687	σ_2	1.5533	0.3246
α_3	1.2631	0.0676	σ_3	1.5128	0.3157
α_4	0.6948	0.0763	σ_4	1.5878	0.3646
α_5	0.9332	0.0626	σ_5	1.5443	0.2961
α_6	0.8596	0.0628	σ_6	1.4731	0.2943
α_7	0.8967	0.0649	σ_7	1.4915	0.3046
α_8	0.2012	0.0722	σ_8	1.7390	0.3549
α_9	0.5316	0.0661	σ_9	1.6052	0.3172
α_{10}	0.4370	0.0646	σ_{10}	1.5049	0.3040
α_{11}	0.6707	0.0637	σ_{11}	1.4122	0.2944
α_{12}	0.1455	0.0823	σ_{12}	1.7710	0.4058
α_{13}	0.7503	0.0603	σ_{13}	1.7233	0.2943
α_{14}	0.7565	0.0630	σ_{14}	1.7636	0.3098
α_{15}	0.8022	0.0626	σ_{15}	1.7325	0.3057
α_{16}	0.5690	0.0682	σ_{16}	1.8531	0.3413
α_{17}	0.8997	0.0606	σ_{17}	1.6252	0.2902
α_{18}	0.7522	0.0571	σ_{18}	1.5008	0.2689
α_{19}	0.9674	0.0574	σ_{19}	1.5978	0.2735
α_{20}	0.3785	0.0629	σ_{20}	1.7069	0.3080
α_{21}	0.7589	0.0492	σ_{21}	1.4557	0.2294
α_{22}	0.7428	0.0530	σ_{22}	1.4677	0.2479
α_{23}	0.9973	0.0548	σ_{23}	1.5622	0.2593
α_{24}	0.5529	0.0642	σ_{24}	1.7346	0.3155
α_{25}	0.8590	0.0521	σ_{25}	1.4218	0.2420

A crucial assumption of the Carlson and Parkin (1975) approach is that the threshold parameters are time–invariant. This also is a source of wide criticism. Batchelor (1986) argues that the threshold parameters should be allowed to be a function of the size and variability of the stimulus. Some empirical papers investigate the appropriateness of varying threshold parameters in the field of inflation expectations. In this context, Seitz (1988) does not find that the threshold parameters are dependent on the level or variance of inflation. Dasgupta and Lahiri (1992) demonstrate that, in the case of inflation expectations, although varying thresholds help to capture extreme values better, they do not improve the resulting quantitative series.[20] Having a glance at Table 1 shows, that time–invariant thresholds might be a sensible choice here: the 10, 50 and 90 percent percentiles as well as means and standard errors of the actual sales changes do not differ much through time.

Our baseline estimation is the one in which wave dummy variables are included in the quantification only. Results are shown in Table 2. Table 2 displays the estimated sales growth rates α and the corresponding standard errors σ (instead of the vector of parameters γ).[21] Each of the coefficients in the mean function and the variance function are highly significantly different from zero except for α_{12}. The weak significance of this wave dummy variable related to the 12th wave, the first quarter of 1997, implies that this is the quarter where sales growth was lowest (0.1455 percent). Inversely, the highest sales growth is dated back to the fourth quarter of 1994 (third wave, α_3, 1.2631 percent).

The standard deviation of the error term σ reflects the heterogeneity of the firms participating in the SSBS so that it is rather surprising that the precision of the quantification is quite low. The heterogeneity of sales growth rate was largest in the second quarter of 1994, which might simply reflect that firms had to get used to the SSBS questionnaire. Interestingly, heterogeneity of growth rates was lowest in the fourth quarter of 1996 (11th wave, 1.4122 percent) and hence coincides with a remarkable increase in the sales growth rates.

Quantified sales growth rates vary considerably across the period of investigation. There are two reasons for this pattern: (i) expansion factors have not been attached to the individual respondents and (ii) the figures have not been seasonally adjusted. The first issue can easily be implemented in maximum likelihood procedures,[22] the second topic

---

[20] The authors used the Producer Price Index for intermediate materials and components for manufacturing as their benchmark for the quantification results of the National Association of Purchasing Managers survey.

[21] The corresponding asymptotic standard errors for σ were obtained using the "Delta"–method (Greene 1997, ch.6.7.5). All estimation results displayed in this paper are obtained using our own procedure programmed for the standard software package STATA6.0. The program code (implemented as an "ADO"–file) can be downloaded from the internet at ftp://ftp.zew.de/quant.ado. GAUSS files can be downloaded from ftp://ftp.zew.de/quant.prg. The standard software package LIMDEP allows for ordered probit estimation with known thresholds without requiring its own programming efforts.

[22] The STATA–ADO file, which can be downloaded from the internet allows the inclusion of such expansion factors.

can be tackled using familiar seasonal adjustment methods.[23] In order to keep things as simple as possible, both issues are not considered here.

Figure 1. Quantified Sales Growth Rates and Corresponding Standard Errors

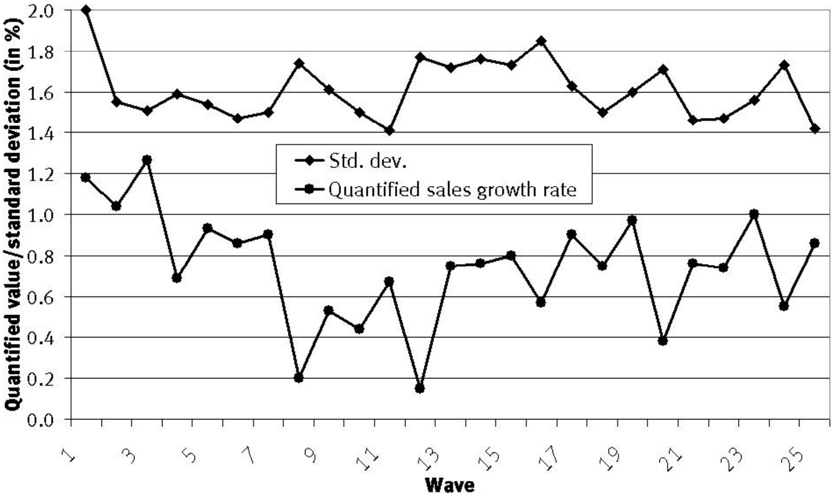


Table 3 displays estimation results of the extended model. In addition to the set of the wave dummy variables, we include control variables for observable firm heterogeneity. These variables include two firm size dummy variables (1–50 and over 100 employees with firms that have between 51 and 100 employees serving as the base category), a dummy variable for Eastern Germany and nine sector dummy variables (the sectors listed in section 2 have waste and sewage disposal as a base category). A comparison of both results shows only slight and unsystematic effects on the quantified sales growth rates. The standard errors $\hat{\sigma}_t$ of the quantified sales growth rates, however, are considerably reduced as displayed in Figure 2.[24] In order to retain the visibility of the firm size, the regional and the sector affiliation effect, Table 3 directly displays the coefficients of the wave dummy and the observable firm heterogeneity variables and not, as in Table 2, the values of σ.

[23] See Kaiser and Buscher (1999) for a suggestion to seasonally adjust short–time series.

[24] In order to maintain the comparability of results, the standard errors of the extended model displayed in Figure 2 refer to a model that included the additional explanatory variables in the variance function only.

Table 3. Ordered Probit estimation results: extended model

Conditional mean			Conditional variance		
	Coeff.	Std. err.		Coeff.	Std. err.
α_1	1.1908	0.1091	γ_1	-3.9195	0.4749
α_2	1.0514	0.0823	γ_2	-4.1742	0.3246
α_3	1.2316	0.0779	γ_3	-4.2467	0.3157
α_4	0.6651	0.0858	γ_4	-4.1990	0.3646
α_5	0.9444	0.0729	γ_5	-4.2522	0.2961
α_6	0.8623	0.0738	γ_6	-4.2812	0.2943
α_7	0.8928	0.0757	γ_7	-4.2719	0.3046
α_8	0.1986	0.0789	γ_8	-4.1720	0.3549
α_9	0.5213	0.0760	γ_9	-4.2144	0.3172
α_{10}	0.4308	0.0764	γ_{10}	-4.2487	0.3040
α_{11}	0.6338	0.0733	γ_{11}	-4.3643	0.2944
α_{12}	0.1386	0.0912	γ_{12}	-4.0956	0.4058
α_{13}	0.6804	0.0708	γ_{13}	-4.1760	0.2943
α_{14}	0.6899	0.0719	γ_{14}	-4.1571	0.3098
α_{15}	0.7328	0.0712	γ_{15}	-4.1911	0.3057
α_{16}	0.4996	0.0760	γ_{16}	-4.1188	0.3413
α_{17}	0.7974	0.0689	γ_{17}	-4.2776	0.2902
α_{18}	0.6820	0.0679	γ_{18}	-4.3242	0.2689
α_{19}	0.8658	0.0665	γ_{19}	-4.2869	0.2735
α_{20}	0.2961	0.0720	γ_{20}	-4.1966	0.3080
α_{21}	0.6567	0.0612	γ_{21}	-4.3067	0.2294
α_{22}	0.6767	0.0652	γ_{22}	-4.3146	0.2479
α_{23}	0.8908	0.0645	γ_{23}	-4.2660	0.2593
α_{24}	0.4377	0.0715	γ_{24}	-4.1738	0.3155
α_{25}	0.7576	0.0639	γ_{25}	-4.3276	0.2420

Table 3. continued

Conditional mean			Conditional variance		
	Coeff.	Std. err.		Coeff.	Std. err.
1–50 employees	-0.1139	0.0303	1–50 employees	0.0687	0.0227
> 100 employees	0.1568	0.0380	> 100 employees	0.0530	0.0284
Eastern Germany	-0.2943	0.0271	Eastern Germany	-0.0317	0.0200
Comp. services	0.6142	0.0526	Comp. services	0.1021	0.0375
Tax cons.	0.3280	0.0471	Tax cons.	-0.1528	0.0367
Management cons.	0.6515	0.0569	Management cons.	0.0950	0.0405
Architecture	-0.3726	0.0491	Architecture	-0.0501	0.0370
Technical advice	-0.0781	0.0432	Technical advice	-0.0366	0.0324
Advertising	0.2421	0.0580	Advertising	0.1745	0.0407
Vehicle rental	0.2529	0.0675	Vehicle rental	0.1657	0.0469
Machine rental	0.1742	0.0592	Machine rental	0.1561	0.0419
Cargo handling	0.2604	0.0493	Cargo handling	0.0045	0.0370

The coefficients related to the mean function are all significantly different from zero at the one percent significance level except for the wave dummy variable related to the 12th wave, which is insignificant, and for the dummy variable for technical planning, which is significant at the ten percent level only.

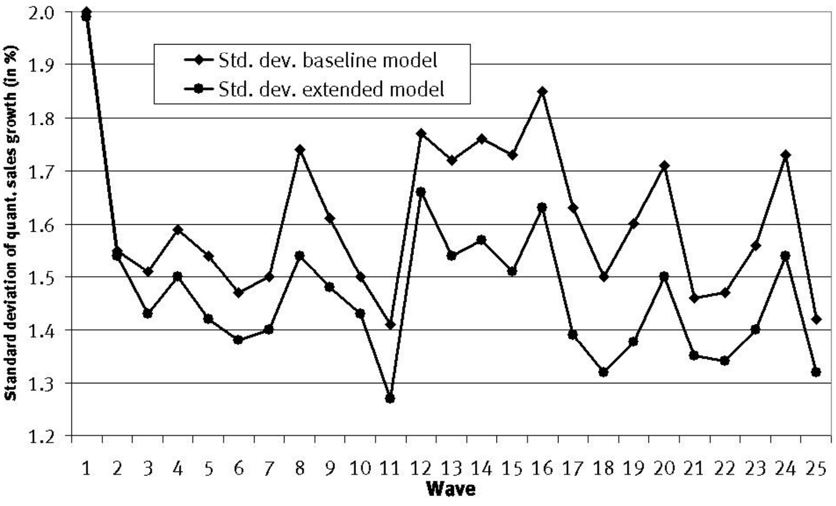
The estimation results for the mean function indicate that larger firm are more likely to grow than smaller firms. Eastern German firms usually have smaller sales growth rates than their Western German competitors. Growth rates are smallest for Management consultancy and Computer services, and are smallest for Architecture.

The estimation results for the conditional variance indicate that the heterogeneity of the business development is largest in a firm with 50–100 employees; a U–shaped effect of firm size on the variance is present. Eastern German firms do not significantly differ from their Western German competitors in the variation of survey answers. The variability of survey responses is smallest for tax consultants and largest for advertising firms.

The wave, size class and sector dummies are also jointly highly significant both in the conditional mean and the conditional variance.

The additional explanatory variables in the mean and in the variance are highly significant from zero as a Likelihood ratio test shows (χ^2_{12}=985.91 with critical values 18.55, 21.03 and 26.22 at the 10, 5 and 1 percent significance level, respectively).

Figure 2. Comparison of Standard Errors of Quantified Sales Growth Rates



5 Specification Tests

As noted above, heteroscedasticity and non–normality of the standard error of the disturbance term ε_i lead to inconsistent parameter estimates of the ordered Probit model.

Tests for heteroscedasticity and non–normality can easily be implemented in applied empirical work by initially calculated generalised residuals (Chesher and Irish, 1987) and by then calculating the appropriate test statistics. The generalised residuals of a q -categorical ordered Probit model are given by:

$$\hat{\varepsilon}_i^{G,q} = \sigma_i \frac{\phi(\frac{\mu_q - x_i'\beta}{\sigma_i}) - \phi(\frac{\mu_{q+1} - x_i'\beta}{\sigma_i})}{\Phi(\frac{\mu_{q+1} - x_i'\beta}{\sigma_i}) - \Phi(\frac{\mu_q - x_i'\beta}{\sigma_i})}. \tag{5}$$

Let z_i denote the vector of variables suspected of causing heteroscedasticity. The LM test statistic for heteroscedasticity can then be obtained by linearly regressing the interaction terms $\hat{\varepsilon}_i^G(\boldsymbol{x_i\beta})$ and $(\hat{\varepsilon}_i^G(\boldsymbol{x_i\beta}))z_i$ upon a vector of ones. The LM test statistic is N times the uncentered R^2 of this auxiliary regression and is χ^2 distributed with degrees of freedom equal to the number of variables potentially causing heteroscedasticity.

It is straightforward to apply this type of test to our baseline model from Table 2 assuming that firm size, sector and regional affiliation may cause heteroscedasticity. Since our control variables for unobserved firm heterogeneity include dummy variables only, we just obtain 37 different generalised residuals (25 wave dummies, 9 sector dummies, 2 size class dummies and 1 dummy for Eastern Germany) so that this type of test does not make much sense. If additional information such as the number of employees in absolute term is available, a test for heteroskedasticity as sketched above can simply be calculated.

An alternative test for heteroscedasticity is readily available by comparing the log–likelihood value of the baseline model with the model including the firm heterogeneity variables in the variance (but not in the mean) function. A simple Likelihood ratio test can then be performed. It turns out that the firm heterogeneity variables are jointly highly significantly different from zero in the variance function, which implies that these variables cause unobserved heteroscedasticity.

Another main source of criticism of the Carlson and Parkin (1975) method is their assumption of normally distributed price expectations — or, equivalently, non–normal error terms — which Carlson (1977) himself found to be non–normal. In this context, it seems advisable to test for the distribution of respondents' sales assessments. This test can be performed as well by using an auxiliary regression of the interaction terms $\hat{\varepsilon}_i^G \boldsymbol{x}_i$, $\hat{\varepsilon}_i^G (\boldsymbol{x}_i \boldsymbol{\beta})^2$ and $\hat{\varepsilon}_i^G (\boldsymbol{x}_i \boldsymbol{\beta})^3$ on a vector of ones. The corresponding LM test statistic is χ^2 distributed with two degrees of freedom. The coefficient related to the term $\hat{\varepsilon}_i^G (\boldsymbol{x}_i \boldsymbol{\beta})^2$ corresponds to skewness, the term $\hat{\varepsilon}_i^G (\boldsymbol{x}_i \boldsymbol{\beta})^3$ corresponds to kurtosis.

Unfortunately, such normality tests are infeasible if heteroscedasticity is present as indicated by simulation results by Davidson and MacKinnon (1992).

This reveals issues for future research, e.g. quantification in a non-parametric setting where the distribution is based on a kernel density estimation. Based on earlier findings of non–normal error terms, such as that by Carlson (1977), it seems likely that normality has to be rejected quite often in practice. It therefore seems advisable to non–parametrically estimate the link function F. This issue, however, has to be left to further research.

6 Conclusion

This paper reviews the probably most important technique to quantify qualitative survey data: the quantification method proposed by Carlson and Parkin (1975). We interpret their methodology in an ordered Probit context and show that respondent–specific variables can be easily implemented in this type of estimation approach. The ordered Probit model is particularly simple to apply since it is included in standard econometric software packages such as LIMDEP and STATA.

Using data taken from a quarterly business survey in the German business–related services sector, we demonstrate that the inclusion of such firm–specific variables such as regional and sectoral affiliation or firm size may substantially reduce the inaccuracy of the standard error of the quantified variables.

Quantification by means of an ordered Probit model also enables the analyst to test for significant effects of firm size for example, on survey responses and on their variability. Moreover, tests for mispecification such as non–normality of the error term or heteroscedasticity which lead to inconsistent parameter estimates can be implemented using standard econometric software packages.

Although the pace of the development of quantification techniques has slowed down remarkably within recent years, there still are avenues for further research. An important aspect in this context is to non–parametrically estimate the distribution function, linking individual survey responses to the quantified value. This issue will be discussed in our further research.

References

Batchelor, R.A. (1986), The Psychophysics of Inflation. Journal of Economic Psychologie 7, 269–290.

Carlson, J.A. (1977), Are Price Expectations Normally Distributed? Journal of American Statistical Association 70, 749–754.

Carlson, J.A., Parkin, M. (1975), Inflation Expectations. Economica 42, 123–138.

Chamberlain, G. (1984). Panel Data. In Handbook of Econometrics, Vol. II. (Grilliches, Z., Intriligator, M., Eds.) North-Holland, Amsterdam.

Chesher, A., Irish, M. (1987). Residual Analysis in the Grouped and Censored Normal Linear Model. Journal of Econometrics 34, 33–61.

CIRET (1998). International Business, Investment and Consumer Surveys: A Synoptic Table. CIRET office, Munich.

Dasgupta, S., Lahiri, K. (1992). A Comparative Study of Alternative Methods of Quantifying Qualitative Survey Responses Using NAPM Data. Journal of Business & Economic Statistics 10/4, 391–400.

Davidson, R., MacKinnon, J.G. (1992). A New Form of the Information Matrix Test. Econometrica 60, 145–157.

Fishe, R.P.H., Idson, T.L. (1989). Information–Induced Heteroscedasticity in Price Expectation Data. The Review of Economics and Statistics, 304–312.

Geil, P., Zimmermann, K.F. (1996). Quantifizierung qualitativer Daten. In Konjunkturindikatoren (Oppenländer, K.H., Ed.). Vahlen, München.

Greene, W.H. (1997), Econometric Analysis, Prentice–Hall, Upper Saddle River.

Hamerle, A., Ronning, G. (1995). Panel Analysis for Qualitative Variables. In Handbook for Statistical Modelling for the Social and Behavioural Sciences (Arminger, G., Clogg, C.C., Sobel, M.E., Eds), Plenum Press.

Hass, H.–J. (1995. Industrienahe Dienstleistungen: ökonomische Bedeutung und politische Herausforderung. Beiträge zur Wirtschafts- und Sozialpolitik 223, 3–39.

Hax, H. (1998). Anforderung an die Statistik zur Konjunkturbeobachtung durch den Sachverständigenrat. Allgemeines Statistisches Archiv 82, 15–24.

Janz, N., Ebling, G., Gottschalk, S., Niggemann, H., Peters, B. (2000). The Mannheim Innovation Panel (MIP): Survey Methodology and Empirical Research. ZEW mimeo, presented at the 9th International Conference on Panel Data at Geneva, June 22–23, 2000.

Kaiser, U. (1998). Die Wirkung wirtschaftspolitischer Ereignisse auf Erwartungen über den Starttermin der Europäischen Währungsunion, Ifo–Studien 2/1998, 103-120.

Kaiser, U., Buscher, H.S. (1999). The Service Sentiment Indicator–A business Climate Indicator for the German Business–related Services Sector. Center of Finance and Econometrics discussion paper 99–06.

Kaiser, U., Stirböck, C. (1999). The Impact of the Introduction of the Euro on Firms' Expectation Concerning Export Behaviour, Product Iinnovation and Foreign Competition: An Empirical Assessment of the German Business–Related Services Sector. intereconomics 34(3), 107–115.

Kaiser, U., Kreuter, M., Niggemann, H. (2000). The ZEW/Creditreform Business Survey in the Business–Related Services Sector: Sampling Frame, Stratification, Expansion and Results," ZEW Discussion Paper 00–22.

Kaiser, U., Pfeiffer, F. (2000). Collective Wage Agreement and the Adjustment of Workers and Hours in German Service Firms. ZEW Discussion Paper 00–33.

Kaiser, U., Voss, K. (2000). On the Lead/Lag Relationship between Manufacturing and Services in the Business Cycle. Ifo-Studien 2/2000, 161-176.

Klodt, H., Maurer, R., Schimmelpfennig, A. (1997). "Tertiarisierung in der deutschen Wirtschaft," Kieler Studien 283, Tübingen.

Maddala, G.S. (1990). Survey Data on Expectations: What Have We Learnt? In Issues in Centemporary Economics, Vol. 2: Macroeconomics and Econometrics (Nerlove, M., Ed.), Macmillan.

Miles, I. (1993). Services in the New Industrial Economy. Futures, 653–672.

Neyman, J., Scott, E.L. (1948). Consistent Estimates based on Partially Consistent Observations. Econometrica 16, 1–32.

Oppenländer, K. H. (1997). Characteristics and Classification of Business Cycle Indicators. In Business Cycle Indicators (Oppenländer, K.H., Hrsg.), Aldershot.

Ronning, G. (1984). Welche Informationen enthält die Antwortkategorie "unverändert" in Tendenzbefragungen? Ifo-Studien 18, 261–272.

Ronning, G. (1990). The Informational Contents of Responses from Business Surveys. In Microeconometrics: Surveys and Applications, (Florens, P.-J., Laisney, F., Eds.), Basil Blackwell.

Saebetzki, A. (1994). Die ZEW-Umfrage bei Dienstleistungsunternehmen: Panelaufbau und erste Ergebnisse. ZEW Dokumentationen 94–05.

Seitz, H. (1988). The estimation of inflation forecasts from business survey data. Applied Economics 20, 427–438.

Strambach, S. (1995). Wissensintensive unternehmensnahe Dienstleister: Netzwerke und Interaktion. Münster (Diss.).

Tutz, G. and W. Hennevogl (1996), Random Effects in Ordinal Regression Models, Computational Statistics & Data Analysis 22, 537–557.

Waller, S. (1997). Struktureller und sozialer Wandel — Konsequenzen für die amtliche Statistik: Ausgewählte Beiträge von der 23. CIRET–Konferenz in Helsinki. Ifo–Schnelldienst 29/1997.

Zimmermann, K.F. (1985). Innovationsaktivitäten, Preisinflexibilität, Nachfragedruck und Marktstruktur. In Industrieökonomik–Theorie und Empirie, (Brombach, G., Ed.), Tübingen.

Zimmermann, K.F. (1997). Analysis of Business Surveys. Handbook of Applied Econometrics, Vol. II: Microeconometrics. (Pesaran, M.H., Schmid, P., Eds.). Basil Blackwell.

Chapter VI

Use of Survey Data for Economic Policy Recommendations

Survey Data as a Source of Information for Policy Recommendations: A Logit Analysis

Elzbieta Adamowicz, Anna Gutkowska, Jacek Kotlowski, Konrad Walczyk

1 Introduction

More than ten years of Polish economic system transition lets us analyse the economic activity in terms of a business cycle. After the recession in the beginning of 90's the succeeding phases accelerating and slowing down the rate of economic growth occurred. There are some variables in the business survey that deliver the essential information about dynamics of economic activity. Most of the variables are procyclical: leading, coincident and lagging. The subject of this analysis is the examination of these variables' usefulness for the economic policy goals in the economy in transition.

It is interesting whether these variables of the qualitative nature can contribute to diminishing uncertainty and delay in economic policy decision taking. We are interested in two from four mentioned by M. Friedman sources of delay (Friedman M., 1953; Burda M., Wyplosz Ch., 1999):

1. reduction time necessary to recognise the phenomenon of the legitimate interference in an economy (diagnosis acceleration),
2. reduction time necessary to decision making (decision acceleration).

In the paper an analysis of the Research Institute of Economic Development of Warsaw School of Economics (RIED WSE) survey data usefulness for economic policy makers is carried out with binary Logit models application. As the reference point in the evaluation of qualitative business survey data usefulness the gross domestic product (GDP) was to be used. Unfortunately only quarterly data on GDP was available while the RIED data for the public and private sector is collected with monthly frequency. Instead we use the volume of sold manufacturing production published monthly by Central Statistical Office (CSO).

In the Polish economy the private sector grows rapidly, but the large public sector also exists. This is the reason why in this paper an analysis is provided for both sectors. We analyse the reaction of the economic subjects. Especially we observe the differences in reactions of public and private firms to the change in the economy.

The objects of our analysis are the following variables:

1. procyclical, changing with some leading in relation to the cyclical fluctuations: total orders (TOB_PRIV), export orders (EO_PRIV and EO_PUB), stocks (ST_PRIV and ST_PUB), general situation of Polish economy (PES_PRIV and PES_PUB),
2. procyclical, changing coincident in relation to the cyclical fluctuations: production (PO_PRIV and PO_PUB), employment (EMP_PRIV and EMP_PUB),
3. procyclical, changing with some lagging in relation to the cyclical fluctuations: prices (P_PRIV and P_PUB), financial situation (FS_PRIV and FS_PUB).

It is necessary to emphasise that an assumption taken by the authors was to analyse the raw business survey data.

2 Theoretical Assumptions

2.1 Yule's Coefficient

In order to measure a correlation between the binary dependent variable Y_t (see sections 2.2 and 3.2) and the RIED state balance statistics, Yule's coefficient of association φ was applied. However, it was necessary to transform survey data measurable on interval scale into dichotomous ones. We assumed that the RIED variable equalled 1 if the corresponding balance statistic was positive and took the value of 0 in the opposite case.

Yule's coefficient of association φ is defined as follows:

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}, \qquad [1]$$

where a, b, c and d are the total of observations for which the dependent variable and the RIED variable transformed into the binary one take values of (0, 0), (0, 1), (1, 0) and (1, 1) respectively. The coefficient falls into $\langle -1, 1 \rangle$ interval.

We considered the relationship between the variables examined as statistically significant if the value of the coefficient was less than -0.05 and more than 0.05. In the latter case the variables were meant to be positively correlated with each other, i.e. a move of the RIED variable from 0 to 1 was associated with the same move of the dependent variable.

2.2 Logit Approach

2.2.1 Logit Model

We examine the Logit model widely used as an analytical tool for description of the relationship between two or more variables, where dependent variable Y is dichotomous (e.g. the occurrence of an event, the choice between two alternatives, etc). We assume that Y_t equals 1 if the business activity in manufacturing sector increases in month t and Y_t equals 0 in the opposite case. The term business activity will be defined precisely in section 3.2.

A simple linear regression of Y on X is not appropriate, since among other things, the applied model implicates inappropriate restrictions on the residuals of the model. Furthermore, the fitted value of Y from a simple linear regression is not restricted to lie between zero and one.

Instead, we adopt a specification that is designed to handle the specific requirements of binary dependent variables. We define a theoretical continuous index I_t as a linear function of explanatory variables $X_1, X_2, ..., X_K$:

$$I_t = \beta_0 + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + ... + \beta_K \cdot X_{Kt} \text{ for } t = 1,...T \quad [2]$$

Observations on I_t are not available. Instead, we have data that distinguish only whether individual observations on Y are in one category ($Y_t = 1$) or in the second one ($Y_t = 0$).

Then we assume that for every t a critical cutoff value I_t^* exists which translates the underlying index into dependent dichotomous variable Y_t. Then the observed dependent variable Y is determined by the fact whether I_t exceeds a threshold value I_t^*:

$$Y_t = \begin{cases} 1 & \text{if } I_t > I_t^* \\ 0 & \text{if } I_t \le I_t^* \end{cases}. \quad [3]$$

In the Logit model we assume that the threshold value I_t^* is logistically distributed. Furthermore we can define the probability that the dependent variable will equal 1 ($Y_t = 1$) as:

$$P_t = P(Y_t = 1) = P(I_t > I_t^*) = F(I_t) = \frac{1}{1 + \exp(-I_t)}. \quad [4]$$

By construction, the variable P_t will lie in the (0, 1) interval. P_t represents the probability that an event in month t occurs, the increase of business activity in the manufacturing sector in this case.

Given such a specification, we can estimate the parameters of this model using the method of maximum likelihood. The log-likelihood function is given by:

$$L(\boldsymbol{\beta}) = \sum_{t=0}^{T}[Y_t \log F(I_t) + (1-Y_t)\log F(-I_t)] \quad [5]$$

As the maximum likelihood function is non-linear we have to employ an iterative maximisation procedure. As a result we obtain the consistent and asymptotically efficient estimates $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0\, \hat{\beta}_1 ... \hat{\beta}_K]'$ (Pindyck, 1991 p. 312).

2.2.2 Forecasting

The model we discuss can be applied directly for forecasting the probability that a given event will occur. By substituting the estimates for the true parameters we can calculate the predicted probabilities $\hat{P}_t$ that an event occurs (in our case the increase of business activity in the manufacturing sector). Naturally, the forecasted $\hat{P}_t$ will usually not be equal to 0 or 1. Then we choose the cutoff value of predicted probability at the level of 0.5 (Pindyck 1991 p. 317). If the predicted probability $\hat{P}_t$ exceeds the cutoff value 0.5, we predict that the event occurs ($\hat{Y}_t = 1$) and if predicted probability $\hat{P}_t$ is less or equal to the cutoff value 0.5, we predict that the event does not occur ($\hat{Y}_t = 0$).

"Correct" classifications are obtained when the predicted probability is lower than or equal to the cutoff value and the observed $Y_t = 0$, or when the predicted probability is higher than the cutoff value and the observed $Y_t = 1$.

There are a number of approaches to measure the forecast accuracy of the Logit model. One of them is based on a classification, computed from estimating a model that includes only the intercept term. If we subtract the percentage of correct classifications for model including only the intercept term from the percentage of correct classifications for the model with explanatory variables, we obtain the gain in the number of correct predictions. This gain is a measure of the predictive ability of the model.

2.2.3 Accuracy of Fit

The preferred measure of accuracy of fit in Logit model is the likelihood ratio index called McFadden R-squared. As the name suggests, this is the analogy to the R-squared reported in linear regression models. This index is constructed as follows:

McFadden R-squared = $1-\frac{L(\boldsymbol{\beta}^*)}{L(\boldsymbol{0})}$, [6]

where $L(\boldsymbol{0})$ represents the value of the log-likelihood function when all parameters are equal to 0 and $L(\boldsymbol{\beta}^*)$ represents the value when the log-likelihood function has been maximised.

By construction, if the maximisation procedure suggests that there is no gain from changing any of the estimated parameters from 0, then McFadden R-squared will equal 0. However if we estimate a likelihood function that forecasts every choice in the given sample correctly, the estimated likelihood function will be 1 and the log-likelihood index $L(\boldsymbol{\beta}^*)$ will be 0. Then McFadden R-squared will equal to 1.

Thus, McFadden R-squared ranges from 0 to 1 and the larger the value of this index, the higher the accuracy of fit.

2.2.4 Interpretation of Coefficients

Interpretation of the coefficient values in the Logit model is complicated by the fact that estimated coefficients from a binary model cannot be interpreted as the marginal effect on the dependent variable. Thus we can interpret the parameters in two ways:

1. as a marginal effect of X_{kt} on the probability of manufacturing sector expansion in month t (called P_t) measured as a partial derivative:

$$\frac{\partial P_t}{\partial X_{kt}} = \beta_k \frac{\exp(\beta_0 + \beta_1 \cdot X_{1t} + ... + \beta_K \cdot X_{Kt})}{(1+\exp(\beta_0 + \beta_1 \cdot X_{1t} + ... + \beta_K \cdot X_{Kt}))^2} = \beta_k \cdot P_t \cdot (1-P_t) \qquad [7]$$

2. as an elasticity of probability P_t regarding to X_{kt}:

$$\frac{\partial P_t \cdot X_{kt}}{\partial X_{kt} \cdot P_t} = \beta_k \cdot X_{kt} \frac{1}{1+\exp(\beta_0 + \beta_1 \cdot X_{1t} + ... + \beta_K \cdot X_{Kt})} = \beta_k \cdot X_{kt} \cdot (1-P_t) \qquad [8]$$

Because of the non-linearity of Logit model, both the marginal effect and the elasticity depend on the values of all of the regressors in model. We need to calculate them at different levels of the explanatory variables to observe the range of variation of the resulting changes in the probabilities. Formula [7] clearly shows that the direction of the effect of a change in X_{kt} depends only on the sign of β_k. Positive values of β_k imply that increasing X_{kt} will increase the probability of the response; negative values imply the opposite.

3 Preliminary Analysis of the RIED Survey Data

3.1 Yule's Coefficient

The highest values of Yule's coefficient have been noticed for EO_PRIV (0.274), PO_PRIV (0.259), TOB_PRIV (0.228) and EMP_PRIV (0.225). Weak correlation exists between the dependent variable and EO_PUB (0.172) and between the dependent variable and PO_PUB (0.131). Other RIED variables are either statistically insignificant or their relationship with Y_t can be neglected. Detailed results of Yule's coefficient calculations are presented in Table 1.

The results suggest that private sector data are the most useful in the prediction of Poland's economic activity changes. It is worth noticing that public enterprises' opinions about overall situation of the Polish economy are negatively correlated with Y_t. This may implicate excessive pessimism of the sector.

On the basis of the correlation results one may also conclude that stocks, prices and financial situation cannot be considered as good indicators of business performance changes. These implications are going to be taken into account when examining the proposed Logit model.

Table 1. Yule's coefficient of association

RIED Variable	a	b	c	d	φ
PO_PUB	28	13	25	20	0.131
PO_PRIV	27	14	18	27	0.259
TOB_PRIV	36	5	31	14	0.228
EO_PUB	32	9	28	17	0.172
EO_PRIV	34	7	26	19	0.274
ST_PUB	12	29	13	32	0.004
ST_PRIV	14	27	17	28	-0.038
P_PUB	3	38	2	43	0.061
P_PRIV	5	36	5	40	0.017
EMP_PUB	41	0	45	0	0.000
EMP_PRIV	34	5	32	15	0.225
FS_PUB	39	2	43	2	-0.010
FS_PRIV	37	4	42	3	-0.056
PES_PUB	30	11	37	8	-0.109
PES_PRIV	36	5	37	8	0.078

3.2 Logit Approach

As it has been already signalled in section 2.2 the advantage of Logit approach is the possibility to calculate the probability that an event represented by the value of 1 of the binary dependent variable happens assuming given values of explanatory variables. In the proposed application Logit analysis allows to calculate the probability that in a given month the volume of NACE D sold industry production increases in relation to the previous month (see section 3.2.1) or to the mean level over the last few months (see section 3.2.2) assuming given values of the RIED business survey statistics.

Let's assume that the economic phenomenon assessed in the RIED business survey is positively correlated with manufacturing production volume. If the RIED statistic is a good indicator of country's business activity changes then its high positive balance should be accompanied with the high probability of NACE D production positive change when in fact this change is recorded and the balance below zero with the low probability. As a result the binary Logit model with a good indicator of Poland's economic performance changes as an explanatory variable should be characterised by McFadden R-squared value close to 1 and the percentage of correct classifications close to 100 percent. In the subsequent sections the RIED variables are analysed with respect to these criteria.

3.2.1 Definition of Binary Dependent Variable

To define the dependent variable the index IN_t of NACE D industry real production change in month t was constructed

$$IN_t = \frac{PO_t}{PO_{t-1}}, \qquad [9]$$

where PO_t is the volume of NACE D industry production in month t in real terms. The index value above 1 means that the volume of industry production in month t increased, equal to 1 that it remained unchanged and below 1 that it decreased in relation to the volume in month $t-1$.

In the sample under examination there were only a few months in which $IN_t \approx 1$. The more it is sufficient to define the binary dependent variable in the following way

$$Y_t = \begin{cases} 1, & \text{if} \quad IN_t > 1 \\ 0, & \text{if} \quad IN_t \leq 1 \end{cases}. \qquad [10]$$

Such a definition implies that the direction of production volume changes, not their magnitude, is the subject of research interest. However it is expected that the strong increase should be accompanied with higher probability of $Y_t = 1$ calculated on the basis of binary Logit models than the weak one.

The estimation outputs of these among 135 binary Logit models in which the dependent variable Y_t is expressed by [10] and in which the RIED statistics enumerated in the introduction lagged by $k =$ -4, -3, ..., 0, ..., 3, 4 are significant as the explanatory variables at no more 5 percent significance level are presented in Table 2.

To summarise the results obtained it is enough to say that these models are characterised by poor goodness of fit as McFadden R-squared does not exceed 0.058. The percentage of months correctly classified on the basis of these models to two groups: periods in which $Y_t = 1$ and $Y_t = 0$, is never higher than 61.63.

Table 2. Estimation outputs of binary Logit models with the RIED balance statistics in the character of explanatory variables significant at no more than 5 percent significance level.

Explanatory Variable	Coefficient	z-Statistic	McFadden R-squared	% Correct	Total Gain
PO_PUB	0.048	2.487	0.058	55.81	3.49
P_PUB(3)	0.035	2.019	0.038	56.63	3.61
ST_PUB(1)	-0.063	-2.052	0.038	61.18	9.41
ST_PUB(4)	-0.073	-2.260	0.049	60.98	8.54
PO_PRIV	0.042	2.164	0.042	61.63	9.30
EO_PRIV(-1)	0.042	2.375	0.055	61.18	9.41
EO_PRIV(2)	0.039	2.211	0.048	54.76	2.38
TOB_PRIV(-4)	-0.038	-2.062	0.041	60.98	7.32
TOB_PRIV(1)	0.040	2.108	0.042	58.82	7.06

Due to unsatisfactory quality of results drawing any conclusions about the character of RIED variables and their usefulness for policy recommendations was resigned. Instead reasonable explanation why the raw business survey data appeared so poor as the indicators of industry production changes started to be searched.

3.2.2 Redefinition of Binary Dependent Variable: Respondents' Memory

Two hypotheses that could possibly explain the quality of results obtained in section 3.2.1 were formulated. The first one says that the respondents are too optimistic/too pessimistic in their assessments of country's economic performance. The second hypothesis can be summarised in the following way: the respondents in their assessments of country's economic performance monthly changes do not compare the economic situation in month *t* with the situation in the previous month as they are asked. Instead the comparison is made with a mean level over the last few months. In short one may say that respondents' memory longer than one month exists while answering the RIED business survey questions. As the latter appears to be more correct it is presented beneath in more detail.

Testing the hypothesis of respondents' memory is conducted with the use of binary Logit models. It is assumed that there is no basis to reject the hypothesis about respondents' memory if the redefinition of binary dependent variable proposed below does improve the estimation results both in terms of McFadden R-squared values and the percentage of correct classifications in relation to the results obtained in section 3.2.1. In the Logit models the redefined binary dependent variable Y_{mt} in the scenario of *m*-month memory is related to the index IN_{mt}, $m =$ 2, 3, 4, 12 expressed by

$$IN_{mt} = \frac{PO_t}{\frac{1}{m}\sum_{l=1}^{m} PO_{t-l}} \qquad [11]$$

in the following way

$$Y_{mt} = \begin{cases} 1, & \text{if} \quad IN_{mt} > 1 \\ 0, & \text{if} \quad IN_{mt} \le 1 \end{cases}. \qquad [12]$$

The value of 1 is assigned to the binary dependent variable Y_{mt} when there is a positive change in the volume of NACE D industry production in month *t* comparing with the mean level over the m previous months. The value of 0 is assigned when the change is non-positive. The explanatory variables are again individual RIED variables lagged by $k =$ -4, -3, ..., 0, ..., 3, 4 presented in the introduction.

The estimation outputs of the models with binary dependent variable expressed by [12] and explanatory variables significant at no more than 5 percent significance level are presented in Table 3 and 4. The results obtained are discussed in detail in the next section.

4 Results for the Individual Time Series

As the definition of binary depend variable expressed by [12] suggests there were four scenarios of respondents' memory considered: 2-, 3-, 4- and 12-month. Almost for all models estimated under the assumption of 3- and 4-month memory the results are better in terms of McFadden R-squared and the percentage of correct classifications than the results obtained under the assumption of 2-month memory. It seems that while answering the business survey questions respondents compare country's economic performance in month *t* with a mean level over more than last 2 months. That is why these results are not discussed in the paper. They are available on the request.

On the other hand almost for all the binary Logit models estimated under the assumption of 12-month memory the results are better in terms of McFadden R-squared and the percentage of correct classifications than the results obtained under the assumption of 3- and 4-month memory. However they are worse in terms of total gain. These results are not discussed in the paper either. The adequate models may be useful for the economic policy recommendations in the longer period, but the research interest is focused rather on the short time horizon. That is why they are available on the request as well.

Table 3. Estimation outputs of binary Logit models with the RIED balance statistics in the character of explanatory variables significant at no more than 5 percent significance level. Dependent variable Y_3.

Explanatory Variable	Coefficient	z-Statistic	McFadden R-squared	% Correct	Total Gain
PO_PUB(-4)	-0.074	-3.288	0.121	71.95	13.41
PO_PUB(-3)	-0.111	-4.095	0.219	73.49	14.46
PO_PUB	0.119	4.224	0.240	72.09	13.95
PO_PUB(1)	0.077	3.394	0.128	68.24	9.41
EO_PUB	0.054	2.504	0.060	60.47	2.33
EO_PUB(1)	0.042	2.009	0.038	63.53	4.71
P_PUB(-2)	0.040	2.350	0.053	63.10	3.57
P_PUB(4)	0.047	2.568	0.067	58.54	-1.22
FS_PUB(-3)	-0.060	-2.706	0.076	71.08	12.05
FS_PUB(-2)	-0.047	-2.222	0.048	67.86	8.33
FS_PUB(1)	0.044	2.147	0.043	62.35	3.53
EMP_PUB(-1)	0.069	2.359	0.052	65.88	7.06
EMP_PUB	0.074	2.547	0.060	63.95	5.81
ST_PUB(4)	-0.068	-2.068	0.042	65.85	6.10
PO_PRIV(-4)	-0.063	-2.834	0.085	65.85	7.32
PO_PRIV(-3)	-0.065	-2.909	0.090	68.67	9.64
PO_PRIV(-2)	-0.041	-2.049	0.040	58.33	-1.19
PO_PRIV	0.111	4.184	0.213	68.60	10.47
PO_PRIV(1)	0.064	2.957	0.088	65.88	7.06
EO_PRIV	0.043	2.380	0.056	60.47	2.33
P_PRIV(4)	0.057	2.716	0.075	62.20	2.44
FS_PRIV(-4)	-0.039	-2.074	0.042	67.07	8.54
FS_PRIV(-3)	-0.050	-2.553	0.066	65.06	6.02
FS_PRIV(-2)	-0.058	-2.866	0.086	63.10	3.57
EMP_PRIV(-1)	0.045	2.210	0.045	61.18	2.35
EMP_PRIV	0.059	2.837	0.077	61.63	3.49
ST_PRIV(1)	-0.048	-2.003	0.037	61.18	2.35
ST_PRIV(2)	-0.063	-2.484	0.060	63.10	3.57
ST_PRIV(3)	-0.049	-2.038	0.040	63.86	4.82
ST_PRIV(4)	-0.049	-2.025	0.040	62.20	2.44
TOB_PRIV(-4)	-0.047	-2.395	0.058	67.07	8.54
TOB_PRIV(-3)	-0.038	-2.033	0.040	62.65	3.61
TOB_PRIV	0.077	3.349	0.128	66.28	8.14
TOB_PRIV(1)	0.054	2.621	0.070	68.24	9.41

Table 4. Estimation outputs of binary Logit models with the RIED balance statistics in the character of explanatory variables significant at no more than 5 percent significance level. Dependent variable Y_4.

Explanatory Variable	Coefficient	z-Statistic	McFadden R-squared	% Correct	Total Gain
PO_PUB(-4)	-0.085	-3.501	0.149	70.73	6.10
PO_PUB(-3)	-0.099	-3.795	0.184	78.31	13.25
PO_PUB(-1)	0.043	2.179	0.047	67.06	2.35
PO_PUB	0.118	4.093	0.235	76.74	12.79
PO_PUB(1)	0.070	3.138	0.110	69.41	4.71
EO_PUB(-1)	0.042	1.992	0.038	65.88	1.18
EO_PUB	0.045	2.127	0.044	68.60	4.65
P_PUB(4)	0.043	2.312	0.055	64.63	0.00
FS_PUB(-4)	-0.042	-1.961	0.039	67.07	2.44
FS_PUB(-3)	-0.066	-2.782	0.086	67.47	2.41
FS_PUB	0.041	2.019	0.038	66.28	2.33
FS_PUB(1)	0.056	2.620	0.069	69.41	4.71
EMP_PUB(-1)	0.060	2.083	0.041	60.00	-4.71
EMP_PUB	0.063	2.206	0.045	59.30	-4.65
ST_PUB(1)	-0.085	-2.501	0.064	64.71	0.00
ST_PUB(2)	-0.077	-2.298	0.054	71.43	5.95
ST_PUB(3)	-0.081	-2.390	0.060	67.47	2.41
ST_PUB(4)	-0.090	-2.531	0.068	65.85	1.22
PO_PRIV(-4)	-0.069	-2.910	0.097	69.51	4.88
PO_PRIV(-3)	-0.060	-2.656	0.077	69.88	4.82
PO_PRIV	0.118	4.239	0.233	72.09	8.14
PO_PRIV(1)	0.068	3.011	0.096	70.59	5.88
P_PRIV(4)	0.052	2.491	0.064	64.63	0.00
FS_PRIV(-4)	-0.042	-2.153	0.048	65.85	1.22
FS_PRIV(-3)	-0.049	-2.431	0.062	69.88	4.82
FS_PRIV(-2)	-0.038	-1.983	0.039	64.29	-1.19
FS_PRIV(1)	0.052	2.607	0.069	68.24	3.53
EMP_PRIV	0.040	1.995	0.037	62.79	-1.16
ST_PRIV(2)	-0.075	-2.775	0.082	69.05	3.57
ST_PRIV(3)	-0.053	-2.105	0.045	69.88	4.82
TOB_PRIV(-4)	-0.059	-2.801	0.088	74.39	9.76
TOB_PRIV(-3)	-0.042	-2.133	0.046	65.06	0.00
TOB_PRIV	0.081	3.381	0.138	70.93	6.98
TOB_PRIV(1)	0.059	2.725	0.081	70.59	5.88

4.1 General Findings

In the scenarios of 3-month and 4-month memory the results obtained are similar and the conclusions drawn on the basis of both scenarios are nearing. That is why they will be discussed simultaneously. In the scenario of 3-month (4-month) memory the explanatory variables in 34 (34) of all the estimated binary Logit models appeared to be statistically significant at no more than 5 percent significance level. Among these variables 9 (9) represented the volume of industry production, 6 (8) financial situation of the enterprises, 5 (8) stocks and 4 (4) total order books of the private sector.

Under the assumption of 3-month and 4-month memory the set of best indicators of Poland's economic performance changes in terms of McFadden R-squared is the same and contains the following variables: PO_PUB (0.240, 0.235), PO_PRIV (0.213, 0.233), PO_PUB(-3) (0.219, 0.184), TOB_PRIV (0.128, 0.138), PO_PUB(1) (0.128, 0.119) and PO_PUB(-4) (0.121, 0.149). These variables are also among the best indicators in terms of the percentage of correct classifications and total gain.

On the basis of estimation outputs presented in Table 3 and 4 one may also conclude that: (1) stocks as the lagging variable are useless for the needs of economic policy makers, (2) under the assumption of 4-month memory prices and employment are not good indicators of Poland's economic performance changes due to the non-positive total gains of the models in which they are employed in the character of explanatory variables, (3) considering all the three criteria of models' quality financial standing of private enterprises is the best indicator of country's business activity changes when it is lagged by $k =$ -3, (4) export orders do not indicate the positive change in the industry production volume earlier than one month before and later than one month after it takes place.

4.2 Probabilities of NACE D Production Positive Change in 1993:01-2000:02

In the previous sections it was shown that PO_PUB and PO_PRIV are among the best indicators of Poland's economic performance changes. Figures 1-2 depict the values of binary dependent variable in the period 1993:01–2000:02 under the assumption of 3-month respondents' memory and the probabilities of NACE D production growth in each month of this period assuming the balance statistics of industry production volume recorded by RIED. The figures are analysed to identify the months in which the direction of country's business activity change is correctly and wrongly recognised.

Under the scenario of 3-month memory the year of 1993 is the worst in terms of PO_PRIV predictive power and equally bad as 1999 when PO_PUB is concerned. Although both indicators correctly predicted non-positive changes in January-February and July-August they completely failed to predict increases that started in March and ended in June. Moreover PO_PRIV did not indicate the growth that began in September and continued till December. In turn in 1994 these indicators gave only one false signal - non-positive change in July proceeded and followed by growths identified as an increase.

Figure 1. Probability of business activity increase in manufacturing sector (for Y_3 and PO_PRIV).

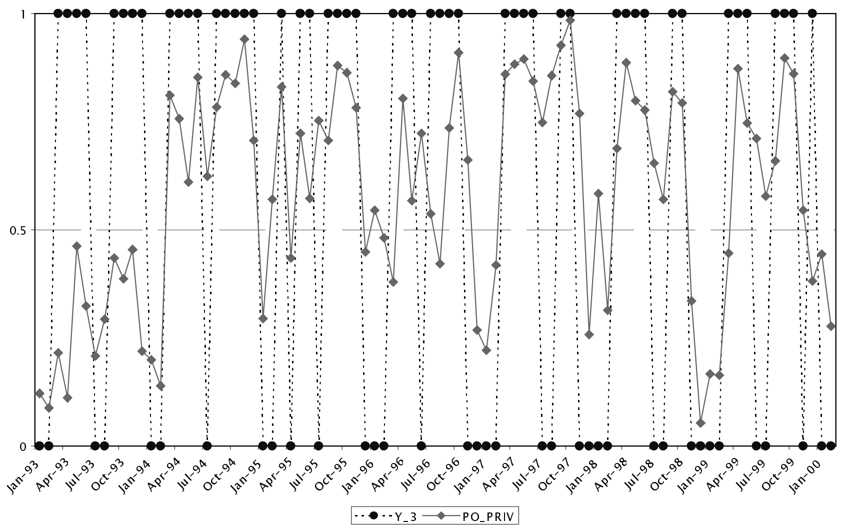
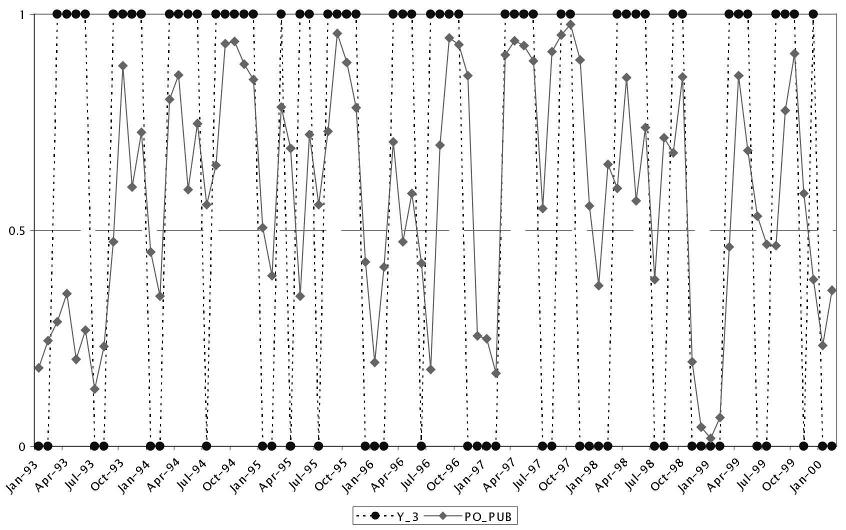


Figure 2. Probability of business activity increase in manufacturing sector (for Y_3 and PO_PUB).



Since August till December 1995 PO_PUB and PO_PRIV were perfectly correct in predicting the type of Poland's economic performance change. In the previous months PO_PRIV appeared better in terms of its predictive power indicating April non-positive and May positive changes what PO_PUB failed to do. In 1996 PO_PUB was late to signal the first positive change in the series from July to October while PO_PRIV did it well. Both variables did not correctly indicate the first non-positive change in the series since November 1996 till February 1997. The non-positive change that took place in June could be expected on the basis of PO_PRIV indications in August while PO_PUB predicted it correctly. Moreover PO_PRIV was late to signal the first positive change in the series from March to May and the first non-positive change in the period January-February.

In 1997 both PO_PUB and PO_PRIV failed to predict the months of non-positive changes in July-August that were observed after 4 months of correctly signalled growths and in November-December recorded after 2 months of correctly classified increases. Similar situation took place in 1998 when PO_PUB and PO_PRIV without mistake predicted growths in the same periods but failed to correctly assess the direction of July-August changes. On the basis of PO_PRIV and PO_PUB it was possible to classify rightly decreases that started in November 1998 and continued till February 1999. Besides in 1999 PO_PUB indicated correctly none of the first months of diversions while PO_PRIV did it well for August-October increase. PO_PUB and PO_PRIV predicted two months of increase in January and February 2000 after the growth in 1999.

4.3 Sensitivity Analysis

For all the RIED variables that turned out to be significant at no more than 5 percent significance level in the binary Logit models estimated under the assumption of 3-month memory the sensitivity analysis was carried out. The marginal probabilities and elasticities were calculated. They were to give an answer to the question about the effect of absolute and relative change in the values of RIED variables on the probability of NACE D industry production volume positive change. Both measures were calculated at the variables' mean and at the balances of -50, -25, -10, -5, 0 (marginal probability only), 5, 10, 25, 50. The means of marginal probability and elasticity absolute values at these balances were calculated to give a synthetic view on variables' usefulness as indicators of change. The results of sensitivity analysis are presented in Table 5 and 6.

4.3.1 Marginal Probabilities

In the scenario of 3-month memory the probability of positive change in the volume of industry production is the most sensitive to the increase of 1 from the mean value of PO_PUB and PO_PRIV (0.028 and 0.026 adequately). The other variables being among the phenomena whose unit growth from the mean level is associated with the highest change in probability are: PO_PUB(-3) (-0.026), TOB_PRIV (0.018), PO_PUB(1) (0.018), PO_PUB(-4) (-0.018), EMP_PUB (0.018). These results confirm conclusions from section 4.1 that PO_PUB, PO_PRIV, PO_PUB(-3) and TOB_PRIV are among the best indicators of Poland's economic performance change.

Table 5. Sensitivity analysis – marginal probabilities. Dependent variable Y_3

Explanatory Variable	Mean	Marginal probalities										Mean of (1)-(9) absolute values
		Mean	-50 (1)	-25 (2)	-10 (3)	-5 (4)	0 (5)	5 (6)	10 (7)	25 (8)	50 (9)	
PO_PUB(-4)	-2.855	-0.018	-0.001	-0.008	-0.015	-0.017	-0.018	-0.018	-0.017	-0.010	-0.002	0.012
PO_PUB(-3)	-2.970	-0.026	0.000	-0.005	-0.019	-0.024	-0.027	-0.027	-0.022	-0.007	-0.001	0.015
PO_PUB	-3.467	0.028	0.001	0.012	0.029	0.029	0.025	0.018	0.012	0.002	0.000	0.014
PO_PUB(1)	-3.366	0.018	0.003	0.013	0.019	0.019	0.017	0.015	0.012	0.005	0.001	0.012
EO_PUB	-6.148	0.013	0.006	0.012	0.013	0.013	0.012	0.011	0.009	0.006	0.002	0.009
EO_PUB(1)	-6.055	0.010	0.006	0.010	0.010	0.010	0.009	0.009	0.008	0.006	0.002	0.008
P_PUB(-2)	20.702	0.010	0.003	0.006	0.008	0.009	0.010	0.010	0.010	0.009	0.006	0.008
P_PUB(4)	21.740	0.011	0.002	0.006	0.009	0.010	0.011	0.012	0.012	0.011	0.006	0.009
FS_PUB(-3)	-17.796	-0.015	-0.005	-0.013	-0.015	-0.014	-0.013	-0.012	-0.010	-0.005	-0.001	0.010
FS_PUB(-2)	-18.079	-0.011	-0.005	-0.010	-0.012	-0.012	-0.011	-0.010	-0.010	-0.006	-0.003	0.009
FS_PUB(1)	-18.404	0.011	0.009	0.011	0.010	0.009	0.008	0.007	0.006	0.004	0.001	0.007
EMP_PUB(-1)	-28.102	0.017	0.013	0.016	0.010	0.007	0.006	0.004	0.003	0.001	0.000	0.007
EMP_PUB	-28.180	0.018	0.013	0.017	0.010	0.007	0.006	0.004	0.003	0.001	0.000	0.007
ST_PUB(4)	4.783	-0.016	-0.001	-0.005	-0.011	-0.013	-0.015	-0.016	-0.017	-0.013	-0.004	0.011
PO_PRIV(-4)	-2.170	-0.015	-0.002	-0.007	-0.013	-0.014	-0.015	-0.016	-0.015	-0.011	-0.003	0.011
PO_PRIV(-3)	-2.200	-0.016	-0.002	-0.007	-0.013	-0.015	-0.016	-0.016	-0.016	-0.011	-0.003	0.011
PO_PRIV(-2)	-2.310	-0.010	-0.003	-0.007	-0.009	-0.010	-0.010	-0.010	-0.010	-0.009	-0.005	0.008
PO_PRIV	-2.520	0.026	0.001	0.011	0.027	0.028	0.025	0.019	0.013	0.003	0.000	0.014
PO_PRIV(1)	-2.270	0.015	0.004	0.012	0.016	0.016	0.015	0.013	0.012	0.006	0.001	0.011
EO_PRIV	-4.100	0.010	0.006	0.010	0.011	0.011	0.010	0.009	0.009	0.006	0.003	0.008
P_PRIV(4)	17.900	0.013	0.002	0.006	0.010	0.012	0.013	0.014	0.014	0.012	0.005	0.010
FS_PRIV(-4)	-17.700	-0.009	-0.005	-0.009	-0.010	-0.010	-0.009	-0.009	-0.009	-0.007	-0.003	0.008
FS_PRIV(-3)	-17.500	-0.012	-0.005	-0.011	-0.012	-0.012	-0.012	-0.011	-0.010	-0.006	-0.002	0.009
FS_PRIV(-2)	-17.400	-0.014	-0.005	-0.012	-0.014	-0.014	-0.013	-0.012	-0.011	-0.006	-0.002	0.010
EMP_PRIV(-1)	-8.660	0.011	0.007	0.011	0.011	0.010	0.010	0.009	0.008	0.005	0.002	0.008
EMP_PRIV	-8.870	0.014	0.006	0.014	0.015	0.014	0.012	0.011	0.009	0.005	0.001	0.009
ST_PRIV(1)	3.490	-0.012	-0.002	-0.006	-0.009	-0.010	-0.011	-0.012	-0.012	-0.011	-0.006	0.009
ST_PRIV(2)	3.550	-0.015	-0.001	-0.006	-0.011	-0.013	-0.014	-0.015	-0.016	-0.013	-0.004	0.010
ST_PRIV(3)	3.530	-0.012	-0.002	-0.006	-0.009	-0.011	-0.011	-0.012	-0.012	-0.011	-0.006	0.009
ST_PRIV(4)	3.540	-0.012	-0.002	-0.006	-0.009	-0.010	-0.011	-0.012	-0.012	-0.011	-0.006	0.009
TOB_PRIV(-4)	-10.300	-0.011	-0.004	-0.009	-0.011	-0.012	-0.012	-0.011	-0.011	-0.008	-0.003	0.009
TOB_PRIV(-3)	-10.200	-0.009	-0.004	-0.008	-0.009	-0.009	-0.010	-0.009	-0.009	-0.008	-0.004	0.008
TOB_PRIV	-10.500	0.018	0.005	0.017	0.018	0.016	0.014	0.011	0.008	0.003	0.000	0.010
TOB_PRIV(1)	-10.200	0.013	0.007	0.013	0.013	0.012	0.011	0.010	0.008	0.004	0.001	0.009

When an average of marginal probability absolute values is compared, conclusions are similar in the sense that the probability reacts the most to the unit growth in the value of variables representing production volume: PO_PUB(-3) (0.015), PO_PUB (0.014), PO_PRIV (0.014), PO_PUB(1) (0.012) and PO_PUB(-4) (0.012).

Table 6. Sensitivity analysis – elasticities. Dependent variable Y_3

Explanatory Variable	Mean	Elasticities									Mean of (1)-(8) absolute values
		Mean	-50 (1)	-25 (2)	-10 (3)	-5 (4)	5 (5)	10 (6)	25 (7)	50 (8)	
PO_PUB(-4)	-2.855	0.084	0.075	0.213	0.209	0.134	-0.201	-0.467	-1.548	-3.582	0.714
PO_PUB(-3)	-2.970	0.124	0.018	0.138	0.240	0.180	-0.328	-0.795	-2.578	-5.516	1.088
PO_PUB	-3.467	-0.160	-5.908	-2.650	-0.689	-0.257	0.112	0.135	0.063	0.007	1.091
PO_PUB(1)	-3.366	-0.103	-3.692	-1.496	-0.404	-0.165	0.099	0.147	0.134	0.042	0.687
EO_PUB	-6.148	-0.136	-2.360	-0.883	-0.248	-0.107	0.075	0.122	0.157	0.090	0.449
EO_PUB(1)	-6.055	-0.103	-1.687	-0.627	-0.187	-0.083	0.063	0.109	0.166	0.131	0.339
P_PUB(-2)	20.702	0.336	-1.846	-0.812	-0.281	-0.132	0.112	0.205	0.365	0.348	0.456
P_PUB(4)	21.740	0.391	-2.250	-1.007	-0.348	-0.162	0.136	0.245	0.407	0.326	0.542
FS_PUB(-3)	-17.796	0.436	0.270	0.463	0.315	0.180	-0.220	-0.473	-1.355	-2.938	0.690
FS_PUB(-2)	-18.079	0.339	0.305	0.381	0.231	0.129	-0.155	-0.334	-0.977	-2.207	0.524
FS_PUB(1)	-18.404	-0.328	-1.601	-0.524	-0.141	-0.061	0.043	0.073	0.103	0.073	0.291
EMP_PUB(-1)	-28.102	-0.796	-2.626	-0.621	-0.114	-0.043	0.023	0.033	0.030	0.011	0.389
EMP_PUB	-28.180	-0.864	-2.883	-0.663	-0.115	-0.042	0.021	0.030	0.025	0.008	0.421
ST_PUB(4)	4.783	-0.131	0.055	0.139	0.134	0.087	-0.138	-0.332	-1.232	-3.174	0.588
PO_PRIV(-4)	-2.170	0.084	0.102	0.217	0.183	0.113	-0.161	-0.371	-1.236	-2.971	0.669
PO_PRIV(-3)	-2.200	0.124	0.092	0.210	0.184	0.115	-0.167	-0.387	-1.296	-3.106	0.695
PO_PRIV(-2)	-2.310	-0.160	0.175	0.213	0.135	0.077	-0.098	-0.217	-0.697	-1.767	0.422
PO_PRIV	-2.520	-0.103	-5.493	-2.457	-0.665	-0.256	0.123	0.156	0.084	0.011	1.156
PO_PRIV(1)	-2.270	-0.136	-3.014	-1.198	-0.338	-0.143	0.095	0.150	0.167	0.073	0.647
EO_PRIV	-4.100	-0.103	-1.804	-0.682	-0.204	-0.091	0.069	0.118	0.178	0.136	0.410
P_PRIV(4)	17.900	0.336	-2.737	-1.244	-0.429	-0.199	0.162	0.284	0.427	0.269	0.719
FS_PRIV(-4)	-17.700	0.391	0.321	0.332	0.187	0.103	-0.121	-0.259	-0.758	-1.752	0.479
FS_PRIV(-3)	-17.500	0.436	0.292	0.393	0.246	0.138	-0.168	-0.361	-1.056	-2.370	0.628
FS_PRIV(-2)	-17.400	0.339	0.257	0.424	0.288	0.165	-0.203	-0.440	-1.279	-2.810	0.733
EMP_PRIV(-1)	-8.660	-0.328	-1.826	-0.660	-0.189	-0.083	0.061	0.103	0.148	0.106	0.397
EMP_PRIV	-8.870	-0.796	-2.640	-0.958	-0.254	-0.106	0.070	0.110	0.127	0.062	0.541
ST_PRIV(1)	3.490	-0.864	0.119	0.178	0.127	0.075	-0.102	-0.233	-0.795	-2.087	0.465
ST_PRIV(2)	3.550	-0.131	0.070	0.155	0.137	0.087	-0.131	-0.310	-1.121	-2.893	0.613
ST_PRIV(3)	3.530	0.043	0.114	0.176	0.128	0.076	-0.104	-0.239	-0.818	-2.150	0.475
ST_PRIV(4)	3.540	0.029	0.111	0.171	0.124	0.074	-0.102	-0.234	-0.805	-2.132	0.469
TOB_PRIV(-4)	-10.300	-0.050	0.227	0.300	0.192	0.109	-0.136	-0.299	-0.912	-2.148	0.540
TOB_PRIV(-3)	-10.200	-0.065	0.247	0.266	0.155	0.087	-0.105	-0.227	-0.690	-1.663	0.430
TOB_PRIV	-10.500	-0.044	-3.590	-1.292	-0.302	-0.117	0.065	0.094	0.081	0.024	0.696
TOB_PRIV(1)	-10.200	-0.090	-2.300	-0.808	-0.216	-0.091	0.062	0.099	0.123	0.069	0.471

The marginal probabilities calculated under the assumption of 3-month memory show that: (1) a decrease in the probability of positive change is associated with an increase in the balance of leading production, leading financial situation, leading total order books and lagging stocks, (2) an increase in the probability of positive change is associated with an increase in the balance of coincident and lagging production, export orders, prices, lagging financial situation, employment, coincident and lagging total order books.

4.3.2 Elasticities

On average the probability of positive change calculated under the assumption of respondents' 3-month memory reacts more than proportionally to the increase of 1 percent in the balance of PO_PRIV (1.156), PO_PUB (1.091) and PO_PUB(-3) (1.088). When an increase of 1 percent from the mean value of variables is considered the probability appears to be the most sensitive to 1 percent change of EMP_PUB (-0.864), ST_PRIV(1) (-0.864), EMP_PUB(-1) (-0.796) and EMP_PRIV (-0.796).

Comparing the means of elasticity absolute values shows that a relative increase in the probability of NACE D industry production is associated with: (1) a relative increase from the negative balance and a decrease from the positive balance of leading production, leading financial situation, leading total order books and lagging stocks, (2) a relative decrease from the negative balance and an increase from the positive balance of coincident and lagging production, export orders, prices, lagging financial situation, employment, coincident and lagging total order books.

5 Results for the Combinations of Time Series

To increase the models' goodness of fit and raise the percentage of correct classifications the RIED variables were merged into the linear combinations. The RIED balance statistics were combined according to the following rules:

1. only coincident and leading variables that turned out to be significant in section 4 were considered in the character of explanatory variables,
2. there were 9 sets of the RIED balance statistics merged in the initial combinations: all the variables, all the variables except of these representing production volume and variables representing production volume exclusively assessed by the private sector, only by the public sector and reported by both sectors,
3. the number of variables in the initial combinations was reduced by subsequent removal of the variables characterised by the lowest absolute value of z-Statistics,
4. the final models were achieved when all the variables in the combinations became statistically significant at no more than 5 percent significance level,
5. in case of an overflow the set of initial variables in the combination was reduced by removal of these variables that performed the poorest while tested individually.

The estimation outputs of the binary Logit models that were constructed according to the presented rules under the assumption of 3-month memory are enclosed in the Appendix. The characteristics of models estimated under the assumption of 12-month memory are available on request.

5.1 General Findings

The results obtained under the assumption of 3-month memory confirm that merging variables into the linear combinations improves models' goodness of fit, increases the percentage of correct classifications and the value of total gain. Usually the best combinations in terms of McFadden R-squared and/or percentage of correct classifications are these in which the assessments of both sectors are employed.

The weighted sum of PO_PRIV, PO_PUB, PO_PUB(-3) is according to McFadden R-squared (0.522) the best linear combination of variables representing exclusively the volume of industry production. However the percentage of correct classifications (85.54) is the highest for the weighted sum of PO_PUB and PO_PUB(-3).

The model with the following set of variables FS_PRIV(-4), FS_PRIV(-2), TOB_PRIV, TOB_PRIV(-4) being in terms of goodness of fit the best combination of all the variables except of these representing production volume has the value of McFadden R-squared significantly lower (0.345). The combination of FS_PUB(-2), EMP_PUB(-1) and TOB_PRIV being the best in terms of correct classifications (80.95 percent) also performs worse than the weighted sum of PO_PUB and PO_PUB(-3).

When all variables were taken into consideration in the initial combination the final model being the best according to all the three criteria (0.643, 87.95, 28.92) contains PO_PRIV(-3), PO_PRIV, PO_PUB(-3), PO_PUB, EO_PUB and P_PUB(-2) in the character of explanatory variables. Combinations of all the variables assessed exclusively by the public sector and only by the private sector contain no other variables than these representing production volume.

5.2 Probabilities of NACE D Production Positive Change in 1993:01-2000:02

In the section 5.1 it was shown that the final models with a set of explanatory variables representing exclusively production and the models with explanatory variables chosen from the initial set of all the variables perform usually better than models in which variables apart from production volume are merged. Figures 3-4 are analysed to identify months of positive and non-positive changes in Poland's business activity in the period 1999:01-2000:02 that were correctly and wrongly recognised on the basis of binary Logit models that performed the best. They depict the values of binary dependent variable under the assumption of 3-month memory and probabilities of NACE D industry production growth in month *t* calculated on the basis of adequate models.

The model with the following set of explanatory variables PO_PRIV(-3), PO_PRIV, PO_PUB(-3), PO_PUB, EO_PUB and P_PUB(-2) estimated under the assumption of 3-month memory correctly associates the low probability of NACE D industry production positive change with actual decreases in the first two months of every year in the period 1993:01-2000:02. So does the model with PO_PRIV, PO_PUB, PO_PUB(-3) merged into the linear combination.

Figure 3. Probability of business activity increase in manufacturing sector (for Y_3 and significant lags of PO_PRIV and PO_PUB)

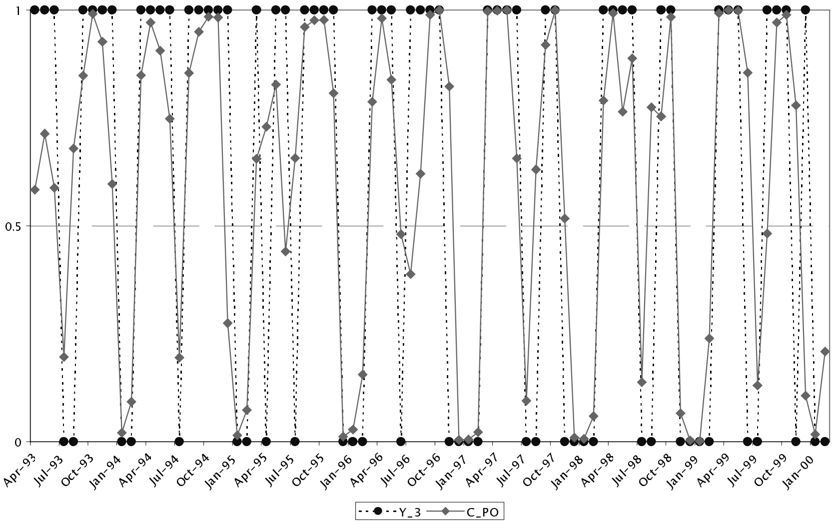
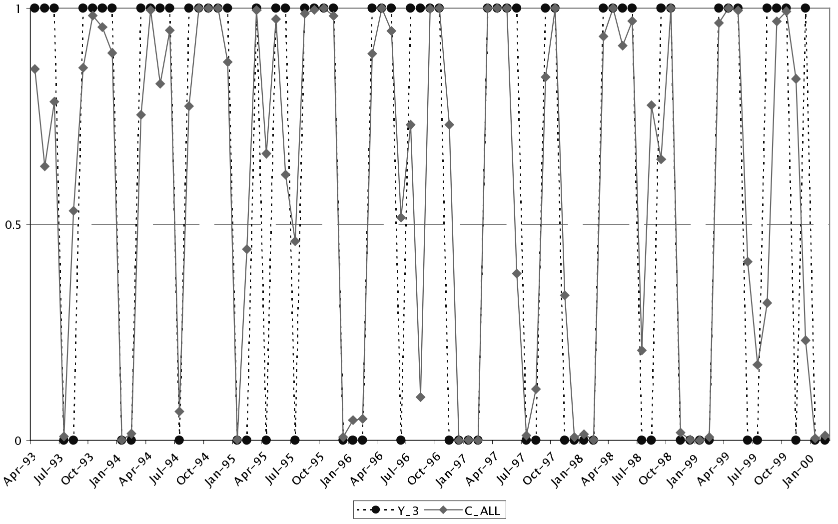


Figure 4. Probability of business activity increase in manufacturing sector (for Y_3 and all significant explanatory variables).



Both models indicated the beginning of series of positive changes in August 1993 while in fact it was reported one month later. In 1994 they correctly signalled the non-positive change that took place in July. The model containing exclusively variables representing production signalled that January-February decrease in 1995 would start one month earlier while in December 1994 the increase still continued. In 1995 both models wrongly classified April decrease. In addition the model with variables representing production volume exclusively indicated a decrease in June and an increase in July when in fact the changes in these months had the opposite directions.

Both models indicated that July-October increase in 1996 would continue still in November when actually it was the first month of a series of non-positive changes. The model with explanatory variables chosen among all the variables signalled that in June 1996 the growth would continue and August would be the month of decrease while in fact the character of changes in these months was the opposite. In turn the model based on variables representing production volume exclusively indicated that June decrease would continue still in July while actually it was the first month of a series of positive changes.

The model with explanatory variables chosen among all the variables signalled that July-August decrease in 1997 would start one month earlier while the model with variables representing production exclusively indicated that it would end one month earlier. It also signalled that September-October increase would continue still in November while in fact it was the first month of a series of decrease. Both models wrongly classified only one change in 1998 - the decrease in August.

The model with a set of explanatory variables chosen among all the variables indicated that June-July decrease in 1999 would last one month longer, August-October increase would continue still in November and December 1999 would be the first month of a series of decreases that actually started in January 2000. The model with a set of explanatory variables chosen among variables representing production volume exclusively signalled that March-May increase in 1999 would continue still in June, June-July decrease would take place still in August, August-October increase would continue still in November and December would be the first month of a series of decreases that actually started in January 2000.

6 Conclusions

In case of Poland, firms participating in business survey learn to act in changing market conditions, thus formulating of their opinions can not be the same as of companies which operate in the ripe market. The binary Logit analysis shows, that some of the business survey variables do not indicate with satisfactory accuracy the directions of economic performance changes. The major reason for weak performance of the RIED indicators is not, what was suspected, the respondents optimistic/pessimistic attitude while answering the business survey questions. Nevertheless their opinions better and better anticipate the directions of changes in the business activity. Thus we can observe the improvement of the survey answers accuracy in late 90's. So we could formulate the following hypotheses, relating the usefulness of business data survey for the economic policy:

1. The best results in the short term (3 months) are obtained for the production (PO_PUB, PO_PRIV), so we should assent these variables for the best coincident indicators, that show well the dynamics of the cyclical behaviour.

2. Other variables show the dynamics of the economic activity variously. Relatively good results are obtained for the following variables: total order books (TOB_PUB), export order books (EO_PUB, EO_PRIV) and employment (EMP_PUB, EMP_PRIV). The variables expressing the stock level (ST_PRIV, ST_PUB) react rather poor and with some lags. Considerably worse valuations concern the financial situation. It can be the result of the high inflation ratio, the problems with its suppression, frequent changes of the tax system and unwillingness of the respondents to reveal the true financial condition of the companies.

3. The worst result concerns the opinion about general Polish economy situation. In relation to the other variables for the whole period of transition the biggest pessimism among companies has been observed. As long as the pessimism exists in respondents opinion, this, so important for the business climate evaluation variable, can be relatively insufficiently helpful for the economic policy goals.

4. The business survey data are available with considerable outstrip in regard to the official statistic data and this is why they can be helpful to diminish the delay in diagnosing the economic situation and to accelerate the decisions. We hope that the importance of the survey results will be useful in a higher degree for Polish economics.

The analysis shows that the differences in behaviour between public and private sector are not so significant. Although it is clear that growing contribution of private sector in Polish economy stabilises the cyclical fluctuations, the experience from the Russian crisis can prove it (second half of 1998).

References

Burda M, Wyplosz Ch, Macroeconomics. A European Text. Oxford University Press, 1997, second edition 1999.

Di Mauro F., Mazzola F.: LDC's repayment problems: a probit analysis, Temi di discussione No. 116, Banca d'Italia, 1989.

Döpke J.: Predicting Germany's recessions with leading indicators – evidence from probit models, Kiel Working Paper No. 944, Kiel Institute of World Economics, 1999.

Estrella A., Mishkin F.: Predicting U.S. recessions: financial variables as leading indicators, Working Paper No. 5379, National Bureau of Economic Research, Inc., 1995.

Friedman M, The Effect of Full Employment Policy on Economic Stability, in : Essays in Positive Economics, The University of Chicago Press, 1953.

Gruszczy•ski M.: Scoring logitowy w praktyce bankowej a zagadnienie koincydencji, Bank i Kredyt Nr 5/1999 (in Polish).

LeClere M.: The Interpretation of Coefficients in Models with Qualitative Dependent Variables, Decision Sciences, Vol. 23, 1992.

Maddala G.: Limited-dependent and qualitative variables in econometrics, Cambridge University Press, 1998.

Pindyck R., Rubinfeld D.: Econometric models and economic forecasts, 3rd edition, McGraw-Hill, Inc., 1991.

Stock J., Watson M.: A procedure for predicting recessions with leading indicators: econometric issues and recent experience, Working Paper No. 4014, National Bureau of Economic Research, Inc., 1992.

Train K.: Qualitative Choice Analysis. Theory, Econometrics and an Application to Automobile Demand, The MIT Press, 1986.

Appendix

Table 7. Estimation outputs of binary Logit models with a set of explanatory variables significant at no more than 5 percent significance level chosen among variables representing production volume. Dependent variable Y_3

Explanatory Variable	Coefficient	z-Statistic	Prob.	McFadden R-squared	% Correct	Total Gain
			Private Sector			
C	0.656	1.870	0.062	0.466	84.15	25.61
PO_PRIV	0.227	4.218	0.000			
PO_PRIV(-2)	-0.121	-2.841	0.005			
PO_PRIV(-4)	-0.081	-2.643	0.008			
			Public Sector			
C	0.726	2.008	0.045	0.482	85.54	26.51
PO_PUB(3)	-0.158	-3.864	0.000			
PO_PUB	0.162	4.027	0.000			
			Both Sectors			
C	0.706	1.879	0.060	0.522	83.13	24.10
PO_PRIV	0.104	2.024	0.043			
PO_PUB	0.103	2.150	0.032			
PO_PUB(-3)	-0.176	-3.812	0.000			

Table 8. Estimation outputs of binary Logit models with a set of explanatory variables significant at no more than 5 percent significance level chosen among all the variables. Dependent variable Y_3

Explanatory Variable	Coefficient	z-Statistic	Prob.	McFadden R-squared	% Correct	Total Gain
			Both Sectors			
C	-4.250	-2.265	0.024	0.643	87.95	28.92
PO_PRIV	0.234	2.433	0.015			
PO_PRIV(-3)	0.168	2.193	0.028			
PO_PUB	0.156	2.162	0.031			
PO_PUB(-3)	-0.345	-3.197	0.001			
EO_PUB	-2.202	-2.150	0.032			
P_PUB(-2)	0.209	2.681	0.007			

Table 9. Estimation outputs of binary Logit models with a set of explanatory variables significant at no more than 5 percent significance level chosen among all the variables expect of these representing production volume. Dependent variable Y_3

Explanatory Variable	Coefficient	z-Statistic	Prob.	McFadden R-squared	% Correct	Total Gain
			Private Sector			
C	0.583	0.923	0.356	0.345	76.83	18.29
FS_PRIV(-2)	-0.113	-2.947	0.003			
FS_PRIV(-4)	0.075	2.014	0.044			
TOB_PRIV	0.151	3.920	0.000			
TOB_PRIV(-4)	-0.072	-2.318	0.021			
			Public Sector			
C	0.260	2.410	0.016	0.211	72.29	13.25
FS_PUB(-3)	-0.097	-3.628	0.000			
EMP_PUB	0.130	3.449	0.001			
			Both Sectors			
C	1.391	1.303	0.193	0.331	80.95	21.43
TOB_PRIV	0.123	3.490	0.001			
FS_PUB(-2)	-0.146	-3.990	0.000			
EMP_PUB(-1)	0.086	1.965	0.050			

The Evolution of Expectations of Disinflation: The Case of Bulgaria

John A. Carlson and Neven T. Valev

Abstract

Currency boards are monetary regimes designed to significantly restrict discretion over monetary policy. With rational expectations, the introduction of a currency board should reverse expectations of high inflation thus making rapid and costless disinflation possible.

In Bulgaria as in other countries before, inflation had declined rapidly after the introduction of a currency board. Was this initial success due to rational expectations or did it occur despite expectations tied to past experiences? We use data from two national surveys in Bulgaria to address that and related questions. We find that expected inflation is indeed lowered to a certain extent by the prospect of a currency board but to a different degree for different agents. Priors of the "type" of the policymaker and familiarity with the operation of the currency board contribute to explaining the differences. Despite the high expected inflation at the introduction of the currency board, actual inflation declined rapidly. Once actual inflation stabilized at lower levels, so did expectations.

1 Introduction

Under rational expectations, high inflation can be reduced rapidly and without output costs by simply introducing a credible policy that is consistent with low inflation. Such policies involve, for example, fixing the exchange rate and tightening the budget. Indeed the experiences of countries with hyperinflation in the 1920's, discussed by Sargent (1982), have been interpreted as strong evidence for rational expectations.

Does inflation come down quickly in such cases because of rational expectations or in spite of expectations tied to past experience? Considerable evidence suggests that expectations are "not-quite-rational" (Roberts, 1997) and that they contain a strong adaptive component (Jonung 1987).

In this paper we employ data drawn from two national household surveys in Bulgaria to analyze the evolution of forecasts of inflation. One of the surveys was taken prior to the introduction of a currency board on July 1st, 1997, but after all of its parameters were announced. The other was taken 10 months later when inflation had stabilized.

The evidence supports a number of claims. First, we find that expected inflation is influenced by both forward and backward looking components. Recent personal inflation experience has a strong influence on expectations but agents also take into account future developments to some extent.

Second, at the introduction of the currency board, the forward-looking component was not sufficiently strong for expected inflation to decline to low levels. In fact, median monthly expected inflation was 10 percent, lower than expected inflation without a currency board but clearly significantly higher than desired.

Third, expected inflation differs significantly across agents in predictable ways. The strongest factor is agents' perceptions of past inflation rates. Other factors are agents' level of education, as a proxy for understanding the operation of the currency board, and their political affiliation, as a proxy for their perceptions about the inflation preferences of the policymaker.

Fourth, perceptions of actual inflation are more strongly influenced by observed price changes of individual products than by reports of changes in an index of consumer-goods prices.

Fifth, despite the high expected inflation at the time the currency board was introduced, actual inflation declined rapidly in Bulgaria with no detectable output cost during the first year. This indicates that rational expectations about inflation as the result of a new monetary regime are not necessary for a successful disinflation policy.

2 The Introduction of the Currency Board in Bulgaria

Bulgaria is one of the transition countries in Southeast Europe. As in most countries in the former East bloc, the process has had ups and downs. The most dramatic crisis came in late 1996 and the first half of 1997 when the local currency depreciated by close to 3000 percent and price increases touched hyperinflation levels. Several major banks failed. In reaction to these developments, IMF suggestions for a currency board first started to appear in the media in late 1996. The political opposition to the then current government embraced the notion. With the acceleration of price increases and further devaluation, public unrest and mass acts of civil disobedience grew to a point where the government stepped down and new parliamentary elections were held. The opposition which formed the new government in late Spring introduced the currency board on July 1st, 1997. Financial stabilization, along with structural reform, was a major issue on the agenda.

Currency boards are institutions that replace central banks and are designed to ensure that domestic currency can be purchased on demand at a fixed exchange rate.[1] Some monetary aggregate, usually the monetary base, is backed by foreign exchange reserves. In its Orthodox form, a currency board has no responsibilities regarding the provision of liquidity to the banking system, financing the budget deficit, or reacting to rising unemployment. Inflation convergence with recent currency boards (Argentina, Estonia, and

[1] See Schwartz (1993) and Williamson (1995).

Lithuania) has been fairly rapid with inflation decreasing to low levels within one year of their introduction.

In the last few months before July 1st, a considerable debate about the nature of a currency board and whether it is appropriate in Bulgaria occupied the media. The opinions were very politicized and conflicting. Nevertheless, the amount of information appeared sufficient for a person with some knowledge of how the economy works to understand what a currency board is, in what ways it constrains the government, and what it implies about inflation and the exchange rate. Argentina and Estonia were also often discussed in the media.

There were some concerns regarding the currency board. First, the local currency was fixed to the German mark even though a major part of Bulgarian trade, such as oil and natural gas, is in US dollars. It was also the dollar to which savers ran during the financial crisis. The country however is a candidate to join the European Union and that aspiration seemed to dominate the choice. Second, the level of the peg, 1000 lev = 1 German mark, was considered inappropriate by some influential analysts including research economists at the central bank. They argued that the lev should be fixed at a more devalued level. Third, it was broadly discussed that the currency board might impose significant strain on the banking system, which had just undergone a major crisis and was considered vulnerable. Fourth, much uncertainty remained regarding privatization and structural reforms in general. Many expressed the view that the real reason for the financial crisis lay in this arena. Fifth, the Bulgarian currency board, like those in Argentina, Lithuania and Hong Kong, is not Orthodox by design. The liability side of its balance sheet includes items often held by central banks, such as government deposits and funds for emergency liquidity to the banking system. Therefore, monetary discretion is limited rather than eliminated.[2]

Many factors made a confident forecast of inflation difficult. Yet the currency board was supposed to deliver low expected inflation immediately. Was it successful? Did agents who understand better what a currency board is expect lower inflation? Did the currency board alleviate concerns about who controls monetary policy? Such questions prompted the surveys described in the following section.

3 Description of the Surveys

A national polling organization conducted a survey in Bulgaria during the last two weeks of June 1997 immediately before the introduction of a currency board on July 1st. By mid June, the fixed level of the exchange rate and the members of the currency board were announced. The survey (with 1022 respondents) is considered representative for the country. Two of the questions from the survey relate to expected inflation. In the first, each respondent was asked about her/his expectation of the average monthly inflation over the following year if a currency board is introduced and, in the second, about her/his expectation of the average monthly inflation over the following year if a currency board were not introduced. Monthly rather than annual rates of inflation were chosen because at

---

2 Bulgarian National Bank (1997).

that time, after a period of high and unstable inflation, price movements were generally discussed and quoted in the media in terms of monthly rather than yearly changes.

The timing of this survey is unique in the sense that it captures a very specific moment. The introduction of a currency board is an event with very low frequency, and the survey was carried out immediately before its introduction but after the legislation and all parameters of the board were publicly announced. There was no uncertainty about the level at which the nominal exchange rate was to be fixed or about the members of the board.

The second national survey analyzed in the paper was conducted in the last week of May 1998, several months after inflation had stabilized at lower levels. In addition to questions about expected inflation, questions were introduced about perceptions of recent inflation and about purchasing experience that might influence perceptions. The intent of the survey was to examine influences on expected inflation as a result of unfolding disinflation, perceptions of the process, and personal observations of relative price changes.

In both surveys, respondents also indicated their age, education level, gender, and political attitudes. Income data were provided by too few respondents to be usable in the estimations.

4 Expected Inflation at the Introduction of the Currency Board

CPI inflation in Bulgaria averaged close to 40 percent per month in the year leading to the introduction of a currency board on July 1, 1997. Starting twelve months before that date, it increased from around 2 percent per month to about 20 percent per month and continued to accelerate. The peak was in February when consumer prices increased by 242 percent before subsiding to a level of 3 percent monthly 2-3 months before the introduction of the currency board.

The survey at that time enables us to address the following questions. Was expected inflation reduced by the introduction of the currency board? Was the effect of introducing a currency board stronger on the expectations of some subgroups of agents?

A potential factor for differences among agents is the degree to which they are informed about what a currency board is and what it implies. Those with better information (those who form more forward-looking forecasts) should expect lower inflation than those who have less information (those who form more backward-looking forecasts). In the survey, the level of education is a variable that can convincingly proxy for such differences across agents. More educated agents have greater contact with the media where discussions of the currency board took place and occupy positions which offer them better resources to avail themselves of relevant information.

At the time the currency board was introduced, much uncertainty remained about economic policy in general, making forecasts difficult even for those with more information. Despite the restrictions imposed on the currency board, the intentions of the policymaker were still a valid concern. We therefore predict that political attitudes of respondents should explain some differences in expected inflation. Such attitudes proxy for agents' beliefs about the "type" of the policymaker. In particular, those who support the party in office, which introduced the currency board, are predicted to expect lower inflation.

To summarize our hypotheses, let π_{RS} denote the expectation of those who form more-rational expectations and politically support the government; π_{NS} the expectations of those who form more naïve expectations and support the government; π_{RO} those who form more rational expectations and oppose the government; and π_{NO} those who form more naïve expectations and oppose the government. Given that a currency board is introduced we predict:

$$\pi_{NO} > \pi_{RO} > \pi_{RS} \text{ and } \pi_{NO} > \pi_{NS} > \pi_{RS} \quad (1)$$

In other words, the lowest expected inflation will be exhibited in the group of respondents who have more education and politically support the government. Those who oppose the government and are also not well-versed in currency boards will expect the highest inflation compared to all other groups. The "rational opponents" and the "naïve supporters" will fall in-between in terms of expected inflation. How those two groups might rank relative to each other is uncertain a priori.

If a currency board is not introduced, the policies that lead to high inflation continue. Therefore, those who form more-informed expectations have no reason to expect lower inflation than those who form more adaptive expectations. An interesting question is whether political affiliation would play a role. To a large extent, the answer depends on the announced objectives of the government. If there is no reason to believe that the government will pursue a disinflation policy in the absence of a currency board, then political affiliation is unlikely to influence expected inflation.

4.1 Survey results for expected inflation at the introduction of the currency board

Table 1 shows mean and median expected inflation with and without a currency board for the overall sample and for subsamples defined by education, political affiliation, gender and age. Average expected monthly inflation with a currency board is 25 percent, and without a currency board, 50 percent. Note however that median expected inflation is 10 percent if a currency board is introduced and 25 percent if it is not. Because of the positive skewness of the answers, the median may be a better measure of central tendency. Clearly, the introduction of a currency board is consistent with lower expected inflation but does not by itself lower expectations to desired levels (the inflation of the German mark). Note also that the coefficient of variation of expected inflation is quite large, indicating a substantial disagreement among respondents. Such disagreement has often been used in the literature to measure subjective uncertainty.[3]

[3] Batchelor and Dua (1996) provide a list of papers that have used disagreement as a proxy for uncertainty in macroeconomics.

Table 1. Summary statistics of expected inflation Survey results, June 1997, Bulgaria

		Mean	Median	Coeff. of Variation (a)	Obser-vations
Whole Sample	Inflation with currency board	24.96	10	1.75	691
	Inflation without currency board	50.36	25	1.46	701
More education	Inflation with currency board	22.4	10	1.77	460
	Inflation without currency board	51.8	25	1.47	475
Less education	Inflation with currency board	29.9	15	1.57	231
	Inflation without currency board	47.2	25	1.38	226
Political supporters	Inflation with currency board	21.6	10	1.72	339
	Inflation without currency board	51.6	20	1.46	339
Political opponents	Inflation with currency board	28.1	15	1.66	352
	Inflation without currency board	49.1	30	1.42	362
Female	Inflation with currency board	21.07	10	1.09	341
	Inflation without currency board	47.64	25	1.42	342
Male	Inflation with currency board	28.74	10	1.96	347
	Inflation without currency board	53.25	20	1.45	356
Older	Inflation with currency board	23.2	10	1.16	256
	Inflation without currency board	44.7	20	1.30	260
Younger	Inflation with currency board	26.0	10	1.90	435
	Inflation without currency board	53.7	30	1.49	441

Note: Each respondent was asked to provide a forecast of the average monthly inflation rate over the following year conditional on introducing or not introducing a currency board. Education: more educated if respondent has high school or higher education. Political affiliation: supporter if today would vote for the party currently in office. Age: older if respondent is more than 50 years old.

Approximately 30 percent of the respondents did not provide a forecast. With few exceptions, they provided either both forecasts, with and without a currency board, or none at all. It is likely that the selection process is not random and that respondents self-selected on the basis of observable factors. Hence, estimation of the effect of political affiliation and education may produce biased results.

To measure the effects of education and political affiliation, we employ Heckman's (1979) procedure to correct for self-selection bias. It involves the maximum likelihood estimation of a participation equation which explains the decision to provide a forecast and a regression equation relating expected inflation to education and political affiliation. The procedure produces consistent and asymptotically efficient estimates by taking into account the correlation of the error terms in the two equations.[4] The estimations were performed using different definitions of the dependent variable: the actual point estimates; a dummy variable that equals 1 if expected inflation is above the median and 0 otherwise; and a constructed variable which equals 1 if the point estimate is less than 10, 2 if 10-19,.., and 12 if >109. The purpose is to provide robustness checks for the possible effect of

[4] For information on the procedure see Heckman (1979), Greene (1991), and Stata Corporation (1997).

skewness and outliers. The results using the three definitions of the dependent variable are qualitatively similar. In Table 2 we report the estimates using the third construct. The reported estimates of λ in Table 2 indicate also that there is significant selection bias. This appears to be because women and older respondents were less likely and those with more education more likely to report expected inflation.

Table 2. The effect of political affiliation and education on expected inflation Heckman's procedure, Survey data, June 1997, Bulgaria

	Dependent variable: Expected inflation	
	If a currency board is introduced	If a currency board is not introduced
Education	-0.75*** (0.20)	-0.41 (0.29)
Political affiliation	-0.52*** (0.19)	-0.09 (0.26)
Constant	4.71 (0.19)	6.41 (0.29)
λ	-1.71*** (0.14)	-2.51*** (0.25)
	Dependent variable:1 if respondent provided an answer, 0 otherwise	Dependent variable:1 if respondent provided an answer, 0 otherwise
Education	0.17* (0.09)	0.31** (0.09)
Political affiliation	0.13 (0.08)	0.06 (0.08)
Gender	-0.15* (0.08)	-0.20** (0.09)
Age	-0.29** (0.09)	-0.26** (0.09)
Constant	0.50 (0.10)	0.49 (0.10)
Model Chi2(7)	-68.4	-34.7
Number of observations	1022	1022

Note: MLE. Mill's ratio estimates used as starting values. Standard errors in parentheses. ***(**, *) indicates significance at the 1(5, 10) percent level. Expected inflation: = 1 if <10, 2 if 10-19,…12 if >109. Education: 1 if respondent has high school or higher education. Political affiliation: 1 if today would vote for the party currently in office. Gender: 1 if female. Age: 1 if respondent is more than 50 years old.

A significant value for $\lambda=\rho\sigma$, where σ is the estimated standard error of the residuals of the regression equation and ρ is the estimated correlation of the residuals from the participation and the regression equations, indicates evidence of self-selection bias.

The significant negative coefficients on both education and political affiliation given that a currency board is introduced indicate that they do have separate and distinct influences on expected inflation.[5] The magnitude of the effect of education is somewhat greater than that of political affiliation, indicating that understanding the operation of the currency board may have contributed more to lowering expected inflation than beliefs about the type of policymaker.

If a currency board is not introduced, neither education nor political affiliation explain differences in expected inflation among agents. That result is particularly interesting with regards to political affiliation because it suggests that perceptions of the type of policymaker were conditional on the introduction of the currency board.

The overall explanatory power of the regressions is weak. Such unexplained noise, also reflected in the high coefficient of variation in expected inflation, suggests a high degree of subjective uncertainty associated with the point estimates of respondents. As discussed below, this may have played a role in the short-run success of the currency board in Bulgaria.

5 Perceived and Expected Inflation Ten Months after the Introduction of the Currency Board

Despite the high expected inflation at the introduction of the currency board, actual inflation declined rapidly. In the first four months with a currency board, inflation declined from the level of about 3 percent per month to a level of around 1 percent. The cumulative increase in the CPI for the period December 1997 – May 1998, which is the period of the six months before the second survey, was 5.86 percent. Thus, the relatively high expected inflation was not an apparent deterrent to the initial disinflation process.

Did low actual inflation transform into low expected inflation? The survey conducted in May 1998 asks for both perceived and expected inflation and allows an examination of that question.

5.1 Perceived and expected inflation.

There were two questions in the May 1998 survey regarding perceived inflation. The first question asks whether prices in general have increased, decreased or remained the same in the last six months (December 1997 through May 1998), and the second question requests a percentage estimate if the respondent perceived a decrease or an increase in prices. Similar questions were asked about expected inflation.

Four percent of the respondents did not indicate whether they had perceived an increase, decrease or no change in prices in the last six months. Of those who provided an answer, 40 percent reported an increase, 11 percent a decrease, and 49 percent reported no change.

[5] We did not find evidence of significant relationships between age or gender and expected inflation. On average women reported lower expected inflation than men but this did not show up after correcting for selection bias in a multiple regression specification.

Eighty-five percent of those who perceived a decrease and 77 percent of those who perceived an increase in prices provided a numerical estimate. If we assume that "no change" means zero inflation, then the weighted average for the whole sample is 6.59 percent (using the medians we get 4.9 percent). If "no change" means positive but low inflation, average perceived inflation is somewhat higher but still the estimate is very close to the actual inflation, 5.86 percent.

Six percent of the respondents did not indicate if they expected a decrease, increase or no change in prices over the next six months. However, fewer agents reported a numerical estimate of expected inflation (302) compared to the agents who reported an estimate of perceived inflation (380). This, following Fische and Idson (1990), may be interpreted as greater uncertainty regarding the future than the past. Of those who answered, 37 percent indicated an increase, 7 percent a decrease and 56 percent no expected change in prices.

If we assume that "no change" means zero, average expected inflation is 5.56 percent. It appears that despite the high expected inflation at the time the currency board was introduced, by May 1998 expected inflation had declined in response to the decline in actual inflation.

The events that unfolded after the survey was taken proved the expectations largely wrong. Prices actually declined by 2.35 percent over the next six months.

Figures 1 and 2 plot the percentage of respondents who provided quantitative estimates of perceived and expected inflation. Also plotted is a normal distribution bell curve using sample mean and variance. Note that the distribution of expected inflation is tighter than the distribution of perceived inflation. In fact, the equality of the standard deviation of the two distributions is rejected at the 0.01 level. This evidence of mean reversion is contrary to what is usually expected. Other things equal, agents will be more uncertain about the future than the past and generally produce more spread-out estimates of future than of past inflation. This is what Vartia and Mankinen (1984) find using Finnish consumer survey data. Why is that different in the Bulgarian survey? Perhaps the reason was the ongoing convergence in prices following the introduction of the currency board. Agents had different real income experiences as the result of relative price shifts but, on average, expected large fluctuations in prices to subside gradually as the currency board system settled in. Actual price movements over the six months following the survey reveal that relative prices indeed did not fluctuate as widely as in the six months before the survey.

Figure 1. Perceived Inflation, each bar represents a 10-percentage point range

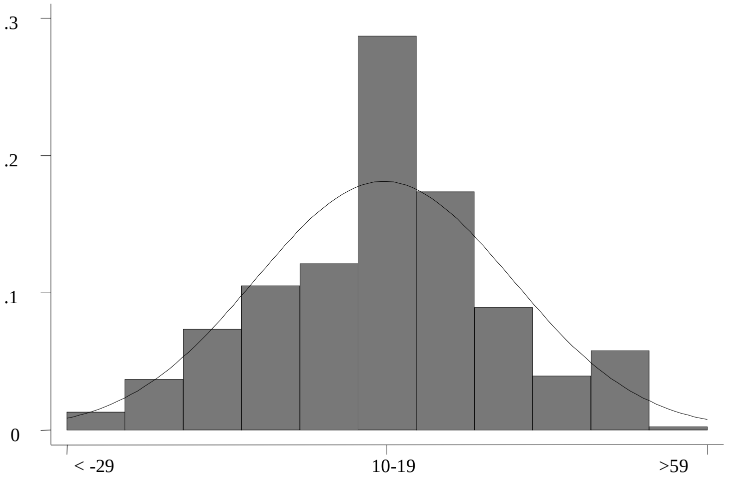
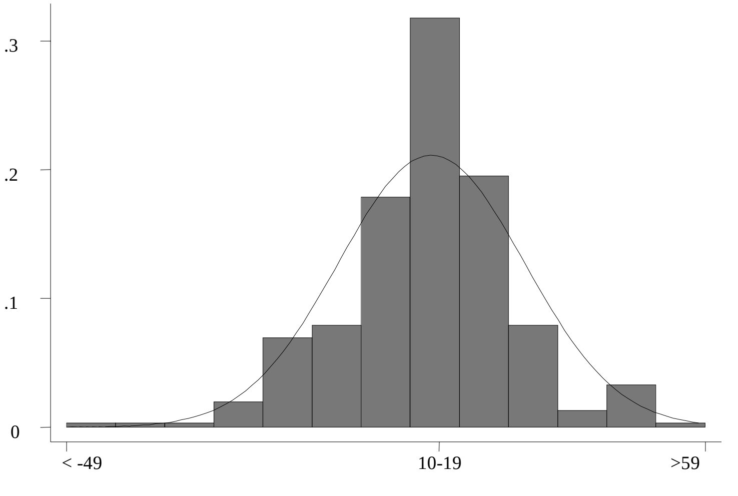


Figure 2. Expected Inflation, each bar represents a 10-percentage point range



5.2 Sources of differences in expected inflation across agents

Table 3 reports a cross-tabulation of respondents who reported both perceptions and expectations. The numbers in parentheses are the percentages of those in each group of perceived change who reported a particular direction for expected price changes.

Seventy one percent of those who perceived an "increase" also expect an "increase", 83 percent of those who perceived "no change" also expect "no change", and only 33 percent of those who perceived a "decrease" expect a "decrease". Thus, a higher percentage of those who perceived an increase are likely to expect a further increase in prices than those who perceived a decrease are likely to expect a further decrease. Presumably this asymmetry can be attributed to an anticipated upward trend in prices.[6]

Table 3. Perceived and expected inflation, Bulgaria, May 1998

	Respondents who expect a decrease in prices	Respondents who expect no change in prices	Respondents who expect an increase in prices	Totals
The group of those who perceived a decrease in prices	32 (33)	45 (46)	20 (21)	97 (100)
The group of those who perceived no change in prices	21 (5)	369 (83)	55 (12)	445 (100)
The group of those who perceived an increase in prices	12 (3)	88 (26)	244 (71)	344 (100)
Totals	65	502	319	886

Note: Numbers in parentheses are percentages.

To examine the influence of perceptions and other factors on expectations, we created two explanatory variables: "Perceived Decrease" equal to 1 if an agent perceived a decrease in prices and 0 otherwise, and "Perceived Increase" equal to 1 if an agent perceived an increase in prices and 0 otherwise. Those variables along with political affiliation, education, age, and gender were used in an order-probit regression to explain the qualitative responses (down, same, up) concerning expected price changes. The results are reported in Table 4.

Table 4 supports the hypothesis that expectations contain a strong adaptive component in that expectations are significantly influenced by perceptions of past price movements. This is particularly so among those who perceived a recent increase in prices. In addition, political affiliation strongly affects expected inflation. Those who support the government tend to expect lower inflation. This effect indicates that despite the

6 In the Appendix we show that differences in perceived inflation are explained to a large extent by the different individual price experiences of respondents.

successful initial convergence of inflation, agents' beliefs about the true intentions of the government were still polarized and still played a role in their forecasts.

Table 4. Expected inflation, Bulgaria, May 1998

	Dependent variable: Expected change in the general price level in the next six months, ordered down, same, up.
Perceived Decrease: equal to 1 if respondent perceived a decrease in prices in the last six months, 0 otherwise	-0.25* (0.13)
Perceived Increase: equal to 1 if respondent perceived an increase in prices in the last six months, 0 otherwise	1.35*** (0.10)
Variable equal to 1 if respondent would vote today for the party currently in office.	-0.51*** (0.09)
Gender: 1 if female, 0 if male	0.05 (0.08)
Age: 1 if over 45, 0 otherwise	0.001 (0.08)
Education: 1 if more than high school education, 0 otherwise	-0.17 (0.12)
Cut 1	-1.50
Cut 2	0.74
Pseudo R^2	0.21
N	857

Notes: Ordered Probit. Standard errors in parentheses. ***(*) significant at the 1(10) percent level.

Finally, note that unlike the survey taken at the introduction of the currency board, education does not explain differences in expected inflation. It appears that once prices have stabilized, personal price experiences and not the degree of knowledge about monetary policy play a more dominant role in forming forecasts.

6 Discussion

This paper uses data from two national consumer surveys in Bulgaria to study the evolution of expected inflation under the Bulgarian currency board. The first survey, which was taken immediately before the currency board came into operation on July 1st, 1997, reveals that expected inflation was lowered by the introduction of the currency board but was still much higher than desired. Respondents differed in their forecasts. Agents with greater education and those who politically support the government that introduced the currency board reported lower expected inflation.

Despite the high expected inflation, actual inflation declined very rapidly without a detectable cost in terms of slowdown in economic activity during the first year of the currency board. In fact, in 1998 output grew by close to 4 percent, the highest growth rate since transition. Clearly, agents did not take actions that would have made these expectations reality. A possible reason is that despite high mean expected inflation, most agents were aware that the monetary environment would change. To the extent that the outcome of that change was uncertain agents took a wait-and-see approach toward wage demands or in setting higher prices than they would if inflation expectations were held with greater certainty. We do not have a direct measure of agents' subjective uncertainty but there was a substantial dispersion in the forecasts at the time the currency board was introduced.

A follow-up survey 10 months later when inflation had stabilized at lower levels reveals that expectations were very much in line with the new environment. The survey also shows that the major determinant of expected inflation appears to be agents' perceptions of past inflation and that differences in perceived inflation among agents is driven by differences in price changes of individual items that households purchased, as shown in the appendix.

Our results suggest that recent inflation experience is an important influence on expected inflation when inflation is low and stable as well as when inflation is high and a new monetary regime is being introduced. Expected inflation in Bulgaria declined in response to the decline in actual inflation and not the other way around. Rational expectations cannot be the primary reason for the rapid and costless disinflation in the first year.

The stance of the central bank in terms of its dissemination of information also appears to have shifted through the process. At the introduction of the currency board, policymakers were engaged significantly in extensive explanations of how a currency board works and what it has done for other countries. Once inflation was lower, focus shifted to referring to the track record with low inflation rather than explaining how low inflation comes about or what policies the currency board has at its disposal. A possible reason is that the design of the currency board in Bulgaria, as that of most other currency boards, allows discretion over monetary policy. One example is the facilities for liquidity to the banking system. Understanding of the balance sheet of the central bank may raise concerns and, respectively, expected inflation on the part of rational agents.

Even though inflation has remained low throughout the period and real exchange rate appreciation does not appear to be a problem, in the years since the 1998 survey, unemployment in Bulgaria has risen substantially. There were several contributing factors. Wide scale privatization has accelerated the process of enterprise restructuring and liquidation. The war in Kosovo along with the financial crisis in Russia led to a sharp decrease in exports and slowdown of foreign investment. In addition, the complex and often changing regulatory environment contributes to the uncertainties faced by domestic and foreign firms.

The government has felt significant political pressure because of the rising unemployment with the popularity of key figures plummeting in recent months. One response to the pressure was the announced goal by the government of "10 percent annual *nominal* growth" in an environment of low and stable inflation. In principle, inflation targeting is not possible under a currency board but under the design of the Bulgarian currency board monetary policy is only limited and not eliminated. It is interesting that despite the high unemployment, the media was immediately saturated with sharp objections to such an approach to deal with the problem. The notion of a nominal income target was soon abandoned.

References

Batchelor, Roy and Pami Dua (1996). "Empirical Measures of Inflation Uncertainty: A Cautionary Note." Applied Economics, 28, 333.341.

Bulgarian National bank (1997) "Annual Report."

Fische, Raymond and Todd L. Idson (1990). "Information-Induced Heteroskedasticity in Price Expectations Data." Journal of Economics and Statistics, May, 304-312.

Greene, William H. (1993). Econometric Analysis, 2nd ed. New York: MacMillan

Heckman, James J. (1979). "Sample Selection Bias as a Specification Error." Econometrica 47(1). 153-161.

Jonung, Lars (1981). "Perceived and Expected Rates of Inflation in Sweden." American Economic Review, December, 961-968.

Roberts, John M. (1997). "Is Inflation Sticky?" Journal of Monetary Economics 39, 173-196.

Sargent, Thomas J. (1982). "The Ends of Four Big Inflations" in Inflation: Causes and Effects, R. E. Hall, ed., University of Chicago Press: Chicago.

Schwartz, Anna J. (1993). "Currency Boards; Their Past, Present and Possible Future Role," Carnegie-Rochester Conference Series on Public Policy 39, December, 147-187

Stata Corporation. Stata Statistical Software: Release 5.0. College Station, Texas, 1997.

Vartia, P. and R. Mankinen (1984). "Perceived and Expected Price Changes in Finland." Weltwirtschaftliches Archiv, 120, 121-32.

Williamson, John (1995). What Role for Currency Boards? Washington D.C.: Institute for International Economics.

Appendix

The Role of Changes in Individual Prices

Previous studies (Jonung, 1981, Vartia and Mankinen, 1984) have documented the close link between perceived past inflation and expected future inflation. Therefore, as here, dispersion of perceptions accounts for the dispersion of expectations. It has been hypothesized but not shown that dispersion in perceptions arises because agents make inferences about changes in the overall price level from their individual price experiences. In conclusion to this section, we briefly offer some evidence supporting that hypothesis.

In the May 1998 survey, respondents were also asked to name at most three products from the prices of which they draw their perceptions of inflation. They could name any goods or services or could say that "prices in general" have changed. The question was intended to test whether agents do think of particular products when they answer a question about the general price level. We test the hypothesis that agents who mention products that had large price increases (decreases) in the last six months will be more likely to report a perceived increase (decrease) in the general price level.

Of particular note, 67 percent named at least one product, but only 4 percent indicated an overall index of prices. We infer from this that perceptions of inflation generally depended more on price changes of individual items that households purchased than on reports about an index of consumer goods prices. The largest categories in the CPI are Foods (55.8 percent), Residential utilities (10.2 percent), Clothing and other small personal belongings (8.6 percent), Transportation and communications (7.9 percent). It is often argued that changes in consumption baskets, as real incomes fell post communism, were not adequately reflected in the CPI weights. Hence, official estimates of inflation may be regarded with less than full confidence and economic agents resort to personal experiences to estimate past inflation. Official data on inflation are published with an approximately one-month delay.

For a statistical analysis of influences on perceived inflation, we utilized an ordered-probit regression for perceived price changes in the last six months. Explanatory variables were "Increase" equal to one if the respondent named a product the relative price of which had increased and zero otherwise, "Decrease" if the respondent mentioned a product the relative price of which had actually decreased during the last six months, zero otherwise. Other explanatory variables were gender, education and age. The estimates from that equation are reported in Table 5. Respondents' assessment of the movements of prices in general were evidently strongly affected by the change in prices of individual products.

Finally, a question in the survey draws on the observation that food prices exhibited significant price volatility relative to other products. It asks what proportion of a respondent's household income is spent on food. We found that this proportion significantly affects whether a respondent reported a perceived positive or negative change in prices -- another indication of how changes in prices of specific goods affects perceptions of an overall change in prices.

Table 5. Perceived inflation, Bulgaria, May 1998

	Dependent variable: Perceived change in the general price level in the last six months, ordered down, same, up.
Decrease: equal to 1 if respondent mentioned a product that actually decreased in relative price in the last six months, 0 otherwise	-1.32*** (0.14)
Increase: equal to 1 if respondent mentioned a product that actually increased in relative price in the last six months, 0 otherwise	2.03*** (0.22)
Gender: 1 if female, 0 if male	-0.01 (0.08)
Age: 1 if over 45, 0 otherwise	0.09 (0.08)
Education: 1 if more than high school education, 0 otherwise	0.01 (0.11)
Cut 1	-1.32
Cut 2	0.40
Pseudo R^2	0.15
N	921

Notes: Ordered Probit. Standard errors in parentheses. *** significant at the 1 percent level.

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(30 2002 11 1 P) ISBN 92-64-09902-6 – No. 52763 2002