OECD Journal: Journal of Business Cycle Measurement and Analysis
Volume 2010/1

Estimating and Forecasting the Euro Area Monthly National Accounts from a Dynamic Factor Model

by

Elena Angelini, Marta Bańbura and Gerhard Rünstler*

We estimate and forecast growth in euro area monthly GDP and its components from the dynamic factor model of Doz et al. *(2006), which handles unbalanced data sets in an efficient way. We extend the model to integrate interpolation and forecasting with cross-equation accounting identities. A pseudo real-time forecasting exercise indicates that the model outperforms various benchmarks, such as quarterly time-series models and bridge equations, in forecasting growth in quarterly GDP and its components.*

JEL Classification: C53, E37

Keywords: dynamic factor models, interpolation, nowcasting

* Angelini: Directorate General Research, European Central Bank, Kaiserstrasse 29, D-60316 Frankfurt am Main, Germany (e-mail: elena.angelini@ecb.europa.eu); Bańbura: Directorate General Economics, European Central Bank (e-mail: marta.banbura@ecb.europa.eu); Rünstler: Austrian Institute for Economic Research, Arsenal Objekt 20, A-1030 Vienna, Austria (e-mail: gerhard.ruenstler@wifo.ac.at).
The authors would like to thank G. Camba-Mendez, D. Giannone, L. Reichlin, C. Schumacher and two anonymous referees for useful discussions and suggestions. The opinions in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank.

Given the delays in the publication of national accounts, economic policy making in real time faces the difficulty of uncovering the actual state of the economy. For the euro area, a flash estimate of GDP is not available until six weeks after the end of the respective quarter. The first national accounts estimate is released about four weeks later. Meanwhile, observers must rely on high-frequency indicators that arrive within the quarter, such as industrial production, surveys and financial market data. However, the large number of available indicators, the noise present in many of the series and the different delays in their publication make the efficient exploitation of this information a difficult task. Under these circumstances, various approaches have been proposed to obtain measures of the economic status from the monthly data. Those include projections of quarterly GDP for the current and, possibly, next quarter, estimates of monthly GDP and monthly coincident indicators of economic activity.

In this paper, we investigate a unified approach to the interpolation and forecasting of GDP and of its demand and value-added components. The approach is based on a dynamic factor model (DFM) for a large monthly data set, and is suitable for dealing with asynchronous data releases in real-time application. Our objective is to obtain estimates and forecasts that satisfy temporal aggregation constraints with respect to quarterly data, as well as appropriate accounting identities.

We build on the DFM of Doz *et al.* (2006a), which differs from other approaches (*e.g.*, those of Stock and Watson, 2002; Forni *et al.*, 2000, 2005) in modelling factor dynamics in an explicit manner. From a state-space representation of the model, forecasts are obtained through application of the Kalman smoother. As a consequence, the model can deal with those irregular patterns of data availability at the end of the sample, which arise in real-time data sets because of differences in publication delays (Giannone *et al.*, 2008). In addition, the state-space framework allows for a comprehensive analysis of the contributions of individual data to the forecasts (Bańbura and Rünstler, 2010).

We combine the model with forecasting equations for monthly GDP and its demand components with appropriate temporal aggregation rules and the relevant accounting identities. Hence, our approach provides monthly estimates and predictions of quarterly GDP growth and its components, which are mutually consistent. This greatly facilitates communicating the results to policy makers, as it clarifies the implications of quarterly predictions for intra-quarter dynamics, and vice versa.

In the empirical part of the paper we evaluate forecasts for GDP and its demand and value components in terms of out-of-sample forecast performance against various alternative models. Banerjee *et al.* (2005), Bańbura and Rünstler (2010) and Angelini *et al.* (2008) report good forecasting performance of factor models for euro area GDP growth. Alternatively, GDP growth has been forecast from bridge equations using a small set of selected monthly indicators, notably measures of production and sales (*e.g.*, Rünstler and Sédillot, 2003; Baffigi *et al.*, 2004; Diron, 2008).

Thus far, estimates of monthly GDP have been derived primarily from bottom-up approaches based on estimates of monthly components (*e.g.*, Mitchell *et al.*, 2005a, 2005b; Proietti and Frale, 2007), which again are based on selected indicators. The bottom-up approach suffers, however, from the potential weakness that poor interpolates for certain components may hamper the aggregate GDP interpolate. Breitung and Schumacher (2008) have employed a diffusion index model for estimating monthly GDP directly.

We find that the factor model forecasts for euro area GDP beat the forecasts from alternative models, such as quarterly time series models and bridge equations. The same applies to forecasts for the demand and value-added components, with the exception of private and public consumption, for which none of the models does well. (For the latter series, further research is required in order to detect informative monthly indicators.) We also compare the monthly interpolates of GDP delivered by our model to those obtained from standard interpolation methods based on a small number of indicators. The resulting in-sample monthly estimates are similar. We argue, however, that this might not be the case for real-time GDP interpolates of the most recent periods.

The paper is organised as follows. After a brief review of the DFM of Doz *et al.* (2006a), the integrated DFM version used to forecast and interpolate the national accounts is presented in Section 1. Section 2 reports the results of a pseudo real-time exercise to compare the performance of the model with various alternative models. Section 3 shows estimates of monthly GDP and compares them to results from standard interpolation methods. Section 4 concludes.

1. The model

1.1. A dynamic factor model

DFMs are designed to explain the dynamics in a panel of series by a few common sources of variation. Consider a vector of n stationary monthly series $x_t = (x_{1t}, \ldots, x_{nt})'$, $t = 1, \ldots, T$, which have been standardised to mean zero and variance one. The DFM by Doz *et al.* (2006a) is given by the equations

$$x_t = \Lambda f_t + \xi_t, \qquad \xi_t \sim \mathbb{N}(0, \Sigma_\xi), \tag{1}$$

$$f_{t+1} = \sum_{s=1}^{p} A_i f_{t-s+1} + B\eta_{t+1}, \quad \eta_{t+1} \sim \mathbb{N}(0, I_q). \tag{2}$$

Equation (1) relates the monthly series x_t to a $r \times 1$ vector of latent factors $f_t = (f_{1,t}, \ldots, f_{r,t})'$ from a matrix of factor loadings Λ, plus an idiosyncratic component $\xi_t = (\xi_{1,t}, \ldots, \xi_{n,t})'$. The latter is assumed to be multivariate white noise with covariance matrix Σ_ξ. Equation (2) describes the law of motion for the latent factors f_t, which are driven by q-dimensional white noise η_{t+1}, where B is a $r \times q$ matrix. The stochastic process for f_t is assumed to be stationary.

Stock and Watson (2005) refer to the specification given by (1) and (2) as the static representation of a DFM, since series x_t load only on current values of factors f_t. However, the specification can be derived from a restricted version of a more general DFM with q so-called dynamic factors. Static factors f_t then contain current and lagged values of dynamic factors, while series x_t may load on current and lagged values of the latter. In this case, it holds that $q < r$.[1]

The specification above differs from the representation of, *e.g.*, Stock and Watson (2002) in that the dynamics of the factors is explicitly modelled through equation (2). Exploiting the dynamics may lead to efficiency improvements in real-time forecasting (Doz *et al.*, 2006a; Rünstler *et al.*, 2009).

1.2. Interpolation

Following Bańbura and Rünstler (2010), we use a mixed-frequency approach to combine the monthly factor model with equations to model monthly GDP growth within a single state-space form. For this purpose, we introduce monthly growth in GDP and its components as latent variables.[2]

More precisely, denote with $Y_t = (Y_{1,t}, \ldots, Y_{m,t})'$ the $m \times 1$ vector of monthly national accounts, which satisfy the accounting identity

$$Y_{1,t} = \sum_{i=2}^{m} Y_{i,t} + K_t,$$

where K_t denotes a possible error term. At the end of each quarter, t = 3*k*, we find the quarterly national accounts as the sum of the respective monthly values,

$$Y_{i,3k}^{Q} = \sum_{s=0}^{2} Y_{i,3k-s},$$

where $k = 1, 2, \ldots, \lfloor T/3 \rfloor$.[3] We form monthly growth rates $y_{i,t} = \ln(Y_{i,t}) - \ln(Y_{i,t-1})$ and three-month growth rates $y_{i,t}^{(3)} = \ln(Y_{i,t}) - \ln(Y_{i,t-3})$. Let $y_t = (y_{1,t}, \ldots, y_{m,t})$. From logarithmic approximation, the following familiar aggregation rules apply to the growth rates of the monthly national accounts (see also, *e.g.*, Mariano and Murasawa, 2003; Breitung and Schumacher, 2008):

$$y_t^{(3)} = y_t + y_{t-1} + y_{t-2}, \tag{3}$$

$$y_{3k}^{Q} = \frac{1}{3}\left(y_{3k}^{(3)} + y_{3k-1}^{(3)} + y_{3k-2}^{(3)}\right). \tag{4}$$

The national accounts identity for monthly growth rates becomes

$$\chi_{t-1}' y_t = \kappa_t, \quad \kappa_t \sim \mathbb{N}(0, \sigma_\kappa^2), \tag{5}$$

where weights χ_{t-1} represent the shares of components in GDP, *i.e.*, $\chi_{1,t-1} = -1$, and $\chi_{i,t-1} = Y_{i,t-1} / Y_{1,t-1}$ otherwise.

Finally, monthly growth rates y_t are related to the common factors by the static equation

$$y_{t+1} = \mu + \Lambda_y f_{t+1} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathbb{N}(0, \Sigma_\varepsilon), \tag{6}$$

where μ is a $m \times 1$ constant term and Λ_y is the $m \times r$ matrix of factor loadings for the national accounts variables. The idiosyncratic component $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{m,t})'$ is assumed to be multivariate white noise with covariance matrix Σ_ε. We further assume that the vectors of innovations $\xi_t, \eta_t, \varepsilon_t$, and κ_t are mutually uncorrelated.

Constraints (3) to (5) are log-linear approximations of the original identities and ignore issues related to chain-linking. In a recent paper, Proietti (2008) proposes an iterative non-linear estimator to interpolate the national accounts, which satisfies the national accounts constraints exactly and also allows for implementing exact (non-linear) temporal constraints for chain-linked data. The approach requires the data set to be balanced and, hence, is not applicable to forecasting in real time. Proietti and Frale (2007) have shown that ignoring chain-linking has only a very small impact on the interpolates. The log-linear approach, in turn, requires a small error term κ_t, but has the advantage that it allows for handling unbalanced real-time data sets.

1.3. State space form

Equations (1) to (6) can be cast in a single state-space form, which is illustrated below for the case of $p = 1$. The transition equation contains the dynamic law of motion for the state vector $\alpha_t' = (f_t', y_t', y_{t-1}', y_t^{(3)}, Q_t')$ comprising factor dynamics (2), temporal aggregation rules (3) and (4), and forecasting equations (6) for the monthly national accounts. In the state space form below, aggregation rule (4) is implemented in a recursive way from

$$Q_t = \Xi_{t-1} Q_{t-1} + \frac{1}{3} y_t^{(3)},$$

where $\Xi_{t-1} = 0_{m \times m}$ in the first month and $\Xi_{t-1} = I_m$ otherwise (see Harvey, 1989: 309ff). As a result, expression (4) holds in the third month of the quarter, $y_{3k}^Q = Q_{3k}$. The recursive implementation of the aggregation through Q_t reduces the size of the state vector, and thereby computation time, to a considerable extent.

The transition equation is given by

$$\begin{bmatrix} I_r & 0 & 0 & 0 & 0 \\ -\Lambda_y & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & -I & -I & I & 0 \\ 0 & 0 & 0 & -\frac{1}{3}I & I \end{bmatrix} \begin{bmatrix} f_{t+1} \\ y_{t+1} \\ y_t \\ y_{t+1}^{(3)} \\ Q_{t+1} \end{bmatrix} = \begin{bmatrix} 0 \\ \mu \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & \Xi_t \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ y_{t-1} \\ y_t^{(3)} \\ Q_t \end{bmatrix} + \begin{bmatrix} B\eta_{t+1} \\ \varepsilon_{t+1} \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

where I denotes the $m \times m$ identity matrix. The equation is to be pre-multiplied by the inverse of the left-hand matrix to achieve the standard state-space form.

The observation equation is

$$\begin{bmatrix} x_t \\ 0 \\ y_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 & 0 & 0 \\ 0 & \chi'_{t-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_m \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ y_{t-1} \\ y_t^{(3)} \\ Q_t \end{bmatrix} + \begin{bmatrix} \xi_t \\ \kappa_t \\ 0 \end{bmatrix}.$$

The final rows of the observation equation, related to y_t^Q, are defined only for the third month of the quarter and otherwise skipped in application.

1.4. Estimation, interpolation and forecasting

As shown by Doz *et al.* (2006a), under certain regularity conditions consistent estimates of the model parameters can be obtained as follows:

1. Apply principal components analysis to x_t to estimate the first *r* common factors $\hat{f}_t$, together with factor loadings $\hat{\Lambda}$ and variances of idiosyncratic components $\hat{\Sigma}_\xi$.
2. Estimate the VAR $\hat{f}_t = \sum_{s=1}^{p} A_s \hat{f}_{t-s} + \hat{\zeta}_t$ to obtain estimates $\hat{A}_s$ and $\hat{\Sigma}_\zeta$. Further, apply principal components to the estimated covariance matrix $\hat{\Sigma}_\zeta$ of residuals $\hat{\zeta}_t$ and extract the first *q* components to obtain $\hat{B}$.
3. Obtain quarterly aggregates $\hat{f}_t^Q$ of estimates $\hat{f}_t$ as from equations (3) and (4). Estimate a quarterly version of (6),

$$y_t^Q = \mu^Q + \Lambda_y \hat{f}_t^Q + \varepsilon_t^Q, \quad \varepsilon_t^Q \sim \mathbb{N}(0, \sigma_\varepsilon^Q),$$

by ordinary least squares (OLS). As equation (6) is static, the quarterly aggregates give consistent estimates of Λ_y, $\sigma_\varepsilon = 3/\sqrt{19}\sigma_\varepsilon^Q$ and $\mu = 1/3\mu^Q$. Similarly, $\sigma_\kappa = 3/\sqrt{19}\sigma_\kappa^Q$, where σ_κ^Q is estimated from the quarterly data.[4]

We now turn to the application of the model to interpolation and forecasting in real time. Real-time data sets typically contain missing observations at the end of the sample because of publication lags. Moreover, the amount of missing data differs across series due to the different timing of data releases. In our forecast exercise we will apply *pseudo-real-time* data sets $\mathcal{Z}_t$, which use the final data releases but take account of the timing of data releases. This is achieved by shifting the pattern of publication lags embodied in $\mathcal{Z}_T$ recursively back in time. That is, monthly observation $z_{i,t-s}, s \geq 0$ is eliminated in $\mathcal{Z}_t$, if and only if observation $z_{i,T-s}$ is missing in $\mathcal{Z}_T$. The quarterly national accounts are treated in a similar fashion.

To obtain efficient estimates and forecasts of GDP growth from unbalanced data sets, Kalman filter and smoother recursions can be applied. For state-space form,

$$\begin{aligned} z_t &= W_t \alpha_t + u_t \qquad u_t \sim \mathbb{N}(0, \Sigma_u), \\ \alpha_{t+1} &= c + T_t \alpha_t + v_t, \quad v_t \sim \mathbb{N}(0, \Sigma_v), \end{aligned} \tag{7}$$

and any unbalanced data set $\mathcal{Z}_t$, the Kalman filter and smoother provide minimum mean square error (MMSE) estimates $a_{t+h|t}$ of the state vector and their precision, $P_{t+h|t}$,

$$a_{t+h|t} = \mathbb{E}\left[\alpha_{t+h} \mid \mathcal{Z}_t \right], \tag{8}$$

$$P_{t+h|t} = \mathbb{E}\left[a_{t+h|t} - \alpha_{t+h} \right]\left[a_{t+h|t} - \alpha_{t+h} \right]', \tag{9}$$

for any $h > -t$. To handle missing observations, the rows in equation (7) corresponding to missing observations in z_t are simply skipped when applying the Kalman filter recursions (Durbin and Koopman, 2003: 92f). In the case of forecasting, $h > 0$, it is sufficient to run the Kalman filter, whereas *ex post* estimates of monthly national accounts are derived from the smoother.

Further, Bańbura and Rünstler (2010) have proposed using an algorithm by Harvey and Koopman (2003) to obtain the Kalman filter and smoother weights of individual series in forecasts and monthly estimates of national accounts. This allows expressing estimates $a_{t+h|t}$ as

$$a_{t+h|t} = \sum_{s=0}^{t-1} \omega_s(h) z_{t-s}. \tag{10}$$

As data sets $\mathcal{Z}_t$ embody fixed data release patterns, the $1 \times n$ vector of weights $\omega_s(h)$ is independent of time t, once the Kalman filter has approached its steady state (see Bańbura and Rünstler, 2010). We will consider the cumulative smoother weights $\sum_{s=0}^{t-1} \omega_{s,j}(h)$ for series j, where $\omega_{s,j}(h)$ is the jth element of $\omega_s(h)$, $j = 1,\dots,n$. The contribution of series *j* to estimate $a_{t+h|t}$ is calculated as $\sum_{s=0}^{t-1} \omega_{s,j}(h) z_{j,t-s}$.

2. Forecast evaluation

In this section we present a forecast exercise to evaluate the historical forecast performance of the dynamic factor model against various rival models, including univariate time series models and bridge equations. We consider forecasts over the period of 2000Q1 to 2006Q2. We address the following questions. First, how well can the components of GDP be forecast? While a number of studies have inspected forecasts for GDP, components have been largely neglected. Second, how does the DFM compare to benchmark models? Third, does constraint (5) help in forecasting?

2.1. Data, publication lags and forecast design

Our euro area data set ($\mathcal{Z}_T$) begins in January 1993 and was downloaded on 20 February 2007. It contains 85 monthly series including official data on economic activity, surveys and financial market data. Our choice of series is based on earlier studies by, *e.g.*, Stock and Watson (2002) and Giannone *et al*. (2008) for the United States, and Bańbura and Rünstler (2010) for the euro area.[5] The monthly series of data on euro area economic activity contains components of industrial production (17), employment and unemployment data (5), extra euro area trade values from the balance of payments (4), retail sales (1) and new passenger car registrations (1). We use 24 series from the European Commission business, consumer, retail and construction surveys. Financial data comprise 17 series including exchange rates (6), interest rates (7) and equity price indices (4). The data contain monetary aggregates and loans (5) and 11 series on the international economy including raw material prices (5) and key macroeconomic indicators for the United States (6). The series list is given in the annex, together with the data transformations we use for all models in this study. The annex also reports the publication lags of the individual series and Augmentend Dickey-Fuller (ADF) tests against unit roots in the transformed data.

The monthly data are published at different times. Surveys, financial data and raw material prices are available at the end of the respective month. By contrast, most of the official data on euro area economic activity, such as industrial production, employment and retail sales, are published with a delay of six to eight weeks after the end of the month. The same applies to the euro area monetary aggregates. This implies that in our data, surveys and financial data are available for January 2007, but most of the real activity data for December 2006 only.

Our euro area quarterly national accounts data include GDP and the major demand components, with inventories subsumed under the statistical discrepancy. In addition, we consider value added and its two major components, industry (including construction and agriculture) and services. The national accounts are published about ten weeks after the end of the respective quarter, while a flash estimate of GDP is available about one month earlier. Hence, our data contain the GDP flash estimate for 2006Q4, but first releases of the demand components and value added only for 2006Q3.

With our forecast design, we aim to replicate the real-time application of the models as closely as possible. Although we do not have real-time data sets at hand, following Rünstler and Sédillot (2003) and Giannone *et al.* (2008), we take account of publication lags in the series and use pseudo real-time data sets $\mathcal{Z}_t$ as defined in Section 1.3. In addition, we re-estimate the models at each point in time based on the available data at the time the forecast is made. Since our data were downloaded on 20 February 2007, our forecasts will replicate the data availability situation on the twentieth day of the month.

We inspect *eight* forecasts for growth in GDP and its components in a certain quarter. These forecasts are obtained in consecutive months. We start with forecasting in the first month of the previous quarter and stop in the second month of the subsequent (next) quarter, one month before the first estimate of national accounts is released by Eurostat. The design will be illustrated in the following section.

2.2. Forecast evaluation

All forecasts are evaluated over the period of 2000Q1 to 2006Q2, with recursive estimation starting in 1993Q1.[6] We consider the following models:

- As benchmarks, we use naive (random walk) forecasts and first-order autoregressive processes (AR(1)) for quarterly GDP and its components. The naive forecast is simply the unconditional mean of the growth rate in each quarterly series, which amounts to a random walk with drift forecast in the level of the series. Again, both forecasts are calculated recursively, *i.e.*, each forecast is based on the available data at the time the forecast is made.
- Bridge equations are widely used for the short-term forecasting of GDP and its components (*e.g.*, Baffigi *et al.*, 2004; Rünstler and Sédillot, 2003; Diron, 2008), as they

employ intraquarter information from the individual indicators. Following Kitchen and Monaco (2006), we obtain forecasts for quarterly series $y_{i,t}^{Q}$ from each indicator $x_{j,t}$, and average across those forecasts. We forecast the individual monthly indicators from monthly $AR(p)$ models, $\varphi(L)x_{j,t}^{(3)} = e_{j,t}$ over the desired horizon, where we use three-month rates $x_{j,t}^{(3)}$, as this tends to give better forecasts. We use the Bayesian information criterion (BIC) to determine lag length p. Forecasts, $\hat{x}_{j,t+h}^{(3)}$, are then aggregated to quarterly frequency, $\hat{x}_{j,t+h}^{Q}$, and the target series $y_{i,t}^{Q}$ is predicted from the "bridge" equation

$$\hat{y}_{i,t+h}^{Q,(j)} = c_{ij} + \beta_{ij}\hat{x}_{j,t+h}^{Q}.$$

We estimate parameters c_{ij} and β_{ij} by OLS. The final forecast $\hat{y}_{i,t+h}^{Q}$ is obtained as the average of the forecasts

$$\hat{y}_{i,t+h}^{Q} = n^{-1}\sum_{j=1}^{n}\hat{y}_{i,t+h}^{Q,(j)}. \tag{11}$$

- As for the DFM, we consider the multivariate model both with and without constraint (5). We apply the model to two sets of national accounts data. The first data set includes GDP and its demand components, *i.e.*, private and public consumption, gross fixed capital formation (GFCF), export and imports and the statistical discrepancy. The second version contains total value added (VAD) plus its breakdown into VAD industry and VAD services.[7]

  As to the specification of the DFM, we determine the number of static factors r from the information criterion developed by Bai and Ng (2002), which gives $r = 4$, while the number of lags in factor dynamics is found from the BIC with $p = 3$. Studies have argued that two shocks are sufficient to model economic activity, and we therefore set $q = 2$ (Giannone *et al.*, 2008). Compared to specification selection based on forecast performance, this approach has the advantage that the specification choice is independent of the target series, as we want to evaluate our model across a set of target series.

Table 1 shows the root mean square error (RMSE) from the naive quarterly forecast over the period of 2000Q1 to 2006Q2 (26 observations). As noted above, we inspect *eight* forecasts for growth in GDP and its components in a certain quarter, which are obtained in consecutive months. We start with forecasting in the first month of the previous quarter and stop in the second month of the subsequent (next) quarter, one month before the first estimate of national accounts is released by Eurostat. As an example, the table illustrates the timing of the forecasts and data releases for the second quarter of the year. We run the first forecast for the second quarter on 20 January and the final (eighth) on 20 August. Note that the last two "forecasts" are actually backcasts, whereas forecasts 4 to 6 amount to nowcasting the current quarter.

The timing of forecasts for the other quarters is equivalent. Since the naive forecast is based on the quarterly data, the RMSE shifts every three months. The timing of these shifts reflects publication dates. New observations for GDP become available in the second month of the quarter, those for components one month later.

Table 1. **RMSE of naive forecast**

Fcst	Example 2nd quarter	GDP	Priv cons	Gov cons	GFCF	Export	Import	Stat discr	VAD total	VAD ind	VAD serv
8	August	0.000	0.314	0.356	0.798	1.509	1.438	0.320	0.364	0.629	0.301
7	July	0.332	0.314	0.356	0.798	1.509	1.438	0.320	0.364	0.629	0.301
6	June	0.332	0.314	0.356	0.798	1.509	1.438	0.320	0.364	0.629	0.301
5	May	0.332	0.316	0.354	0.807	1.526	1.460	0.317	0.372	0.642	0.307
4	April	0.338	0.316	0.354	0.807	1.526	1.460	0.317	0.372	0.642	0.307
3	March	0.338	0.316	0.354	0.807	1.526	1.460	0.317	0.372	0.642	0.307
2	Feb	0.338	0.317	0.358	0.819	1.538	1.474	0.316	0.380	0.652	0.314
1	Jan	0.347	0.317	0.358	0.819	1.538	1.474	0.316	0.380	0.652	0.314

Table 2 presents the results for the AR(1) and the bridge equations in terms of the RMSE relative to the RMSE of the naive forecast. For the bridge equations, we construct the GDP forecast from a limited set of 8 series, which gives a smaller RMSE as compared to forecasts from the entire set of 87 series. The forecasts for components are based on the full set of series.[8] As regards GDP, both models improve upon the naive forecast for the very short horizons, *i.e.*, forecasts 5 to 8 (the backcasts and late nowcasts). For the one-quarter-ahead forecasts (1 to 3), only bridge equations outperform the naive forecast, but the gains remain below 10%.

For the components of VAD, the bridge equations beat the naive forecast. Among the demand components, some gains emerge for exports and imports, but these rarely exceed 10%. For private and public consumption as well as GFCF, the benchmark models give largely uninformative forecasts.

Table 2. **RMSE of benchmark models (relative to naive forecast)**

Quarterly AR(1)										
Fcst	GDP	Priv cons	Gov cons	GFCF	Exp	Imp	Stat discr	VAD total	VAD ind	VAD serv
8		1.04	0.97	1.10	0.95	0.91	0.91	0.89	0.94	0.96
7	0.82	1.04	0.97	1.10	0.95	0.91	0.91	0.89	0.94	0.96
6	0.82	1.04	0.97	1.10	0.95	0.91	0.91	0.89	0.94	0.96
5	0.82	1.07	0.96	1.01	1.00	1.01	1.01	0.97	1.00	0.97
4	0.98	1.07	0.96	1.01	1.00	1.01	1.01	0.97	1.00	0.97
3	0.98	1.07	0.96	1.01	1.00	1.01	1.01	0.97	1.00	0.97
2	0.98	1.07	0.99	1.03	1.01	1.05	1.01	1.02	1.01	1.03
1	1.03	1.07	0.99	1.03	1.01	1.05	1.01	1.02	1.01	1.03
Bridge equations										
Fcst	GDP	Priv cons	Gov cons	GFCF	Exp	Imp	Stat discr	VAD total	VAD ind	VAD serv
8		1.02	0.97	1.06	0.86	0.94	0.99	0.80	0.76	0.83
7	0.86	1.02	0.97	1.06	0.86	0.94	0.99	0.80	0.76	0.83
6	0.87	1.04	1.00	1.06	0.87	0.94	1.00	0.81	0.77	0.84
5	0.88	1.04	1.04	1.08	0.87	0.95	0.99	0.79	0.79	0.86
4	0.87	1.04	1.01	1.09	0.89	0.96	0.99	0.79	0.79	0.87
3	0.89	1.03	1.00	1.08	0.92	0.99	0.99	0.82	0.82	0.88
2	0.92	1.03	1.00	1.10	0.98	1.02	0.99	0.86	0.87	0.89
1	0.94	1.04	0.99	1.09	0.99	1.02	0.99	0.85	0.90	0.88

The results for the DFM version without constraint (5) are shown in the upper panel of Table 3. The model shows substantial improvements upon the alternative models for GDP and most demand and value-added components. For the short horizons, the RMSE of the GDP forecast is now 30% lower as compared with the naive forecast. For the first forecast, eight months ahead of the data release, the improvement still amounts to 20%. Similar gains occur for gross fixed capital formation, exports and imports, and the components of value added. While the absence of any gains for private and public consumption may reflect a lack of forecastability of the series *per se*, it also may be a consequence of a lack of appropriate monthly indicators in our data set. It has been argued that private consumption follows a random walk (Hall, 1988), and this also seems plausible for government consumption. However, the lack of forecastability does not necessarily preclude informative *nowcasts* of the series based on intra-quarter information.

Table 3. **RMSE of dynamic factor models (relative to naive forecast)**

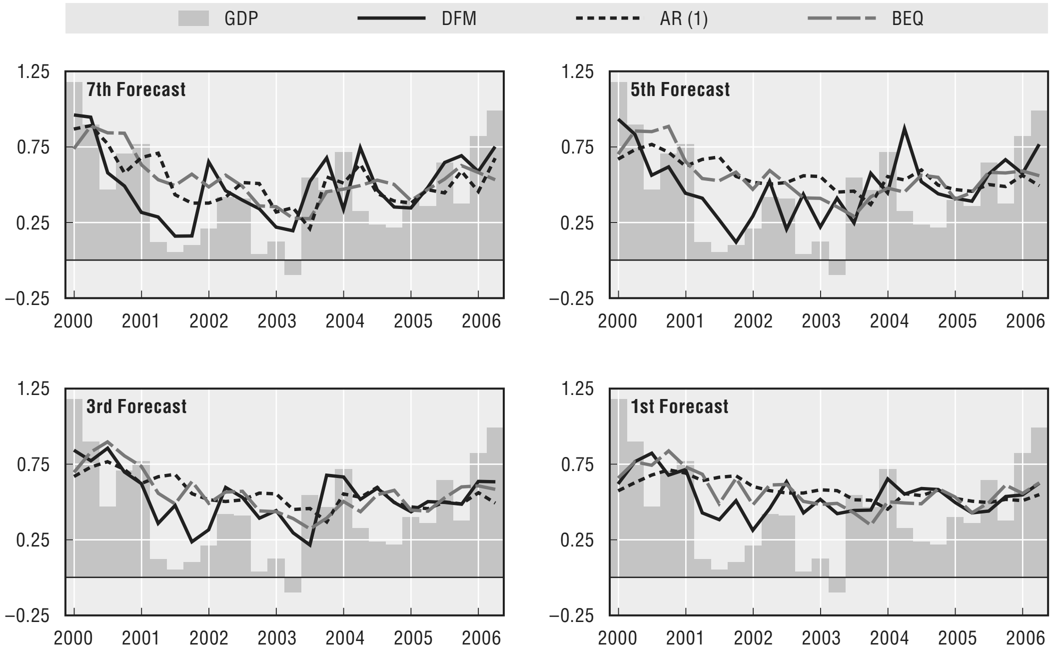
Without constraint (5)										
Fcst	GDP	Priv cons	Gov cons	GFCF	Exp	Imp	Stat discr	VAD total	VAD ind	VAD serv
8		1.01	1.13	0.81	0.77	0.69	1.13	0.72	0.64	0.85
7	0.70	0.98	1.15	0.81	0.77	0.72	1.13	0.70	0.64	0.84
6	0.72	0.98	1.20	0.84	0.77	0.70	1.11	0.75	0.70	0.86
5	0.74	0.98	1.04	0.81	0.72	0.74	0.99	0.75	0.75	0.86
4	0.73	0.97	1.02	0.82	0.70	0.68	0.98	0.77	0.76	0.87
3	0.73	0.96	1.01	0.81	0.82	0.80	0.98	0.76	0.70	0.88
2	0.80	0.96	1.01	0.82	0.86	0.86	0.99	0.76	0.77	0.85
1	0.81	0.98	1.01	0.86	0.90	0.90	0.98	0.80	0.84	0.86
Including constraint (5)										
Fcst	GDP	Priv cons	Gov cons	GFCF	Exp	Imp	Stat discr	VAD total	VAD ind	VAD serv
8		0.94	1.13	0.78	0.74	0.69	1.20	0.72	0.62	0.87
7	0.67	0.97	1.16	0.82	0.78	0.72	1.13	0.72	0.63	0.86
6	0.70	0.96	1.21	0.85	0.77	0.70	1.12	0.75	0.68	0.88
5	0.76	0.97	1.04	0.78	0.73	0.74	0.99	0.76	0.73	0.87
4	0.74	0.96	1.03	0.80	0.70	0.68	0.99	0.76	0.73	0.88
3	0.73	0.96	1.01	0.78	0.82	0.80	0.98	0.74	0.68	0.87
2	0.79	0.95	1.02	0.80	0.85	0.87	0.99	0.78	0.77	0.87
1	0.79	0.97	1.01	0.84	0.89	0.91	0.98	0.84	0.84	0.91

The lower panel of Table 3 shows the results for the DFM using constraint (5). For the demand components, the inclusion of the constraint tends to improve the RMSE for the eighth forecast, but leaves it unchanged otherwise. This appears to be related to the fact that the flash estimate of quarterly GDP is already available for the eighth forecast, as it is published about four weeks before the full national accounts. In this situation, the information contained in the flash estimate contributes to forecasting the demand components.

Figure 1 shows forecasts from the DFM and the benchmarks. The graphs visualize the higher precision of the DFM forecasts compared to the AR(1) and the bridge equations. They indicate a better performance of the DFM compared to the other models, in particular for the 2001–2003 period. As shown in Table A.2 in the annex, however, the DFM also improves upon the alternative models in subsequent years, although to a smaller extent.

Finally, Table A.3 in the annex reports forecasts encompassing and Diebold-Mariano tests of the DFM using constraint (5) against the bridge equations. They indicate that, for

Figure 1. **GDP forecasts**



GDP and most of the components, forecasts from bridge equations do not add information to those from the DFM and are significantly less precise over short horizons. More precisely, we run the encompassing regression

$$y^{Q}_{i,t+h} = \lambda \hat{y}^{Q,DFM}_{i,t+h} + (1-\lambda)\hat{y}^{Q,BE}_{i,t+h} + u_t \tag{12}$$

to explain observations $y^{Q}_{i,t+h}$ by the two forecasts (Clements and Hendry, 1998: 228ff). A value of $\lambda = 1$ indicates that forecasts $\hat{y}^{Q,BE}_{i,t+h}$ from bridge equations do not add information to forecasts $\hat{y}^{Q,DFM}_{i,t+h}$ from the DFM, and the same holds in an opposite manner for a value of $\lambda = 0$.

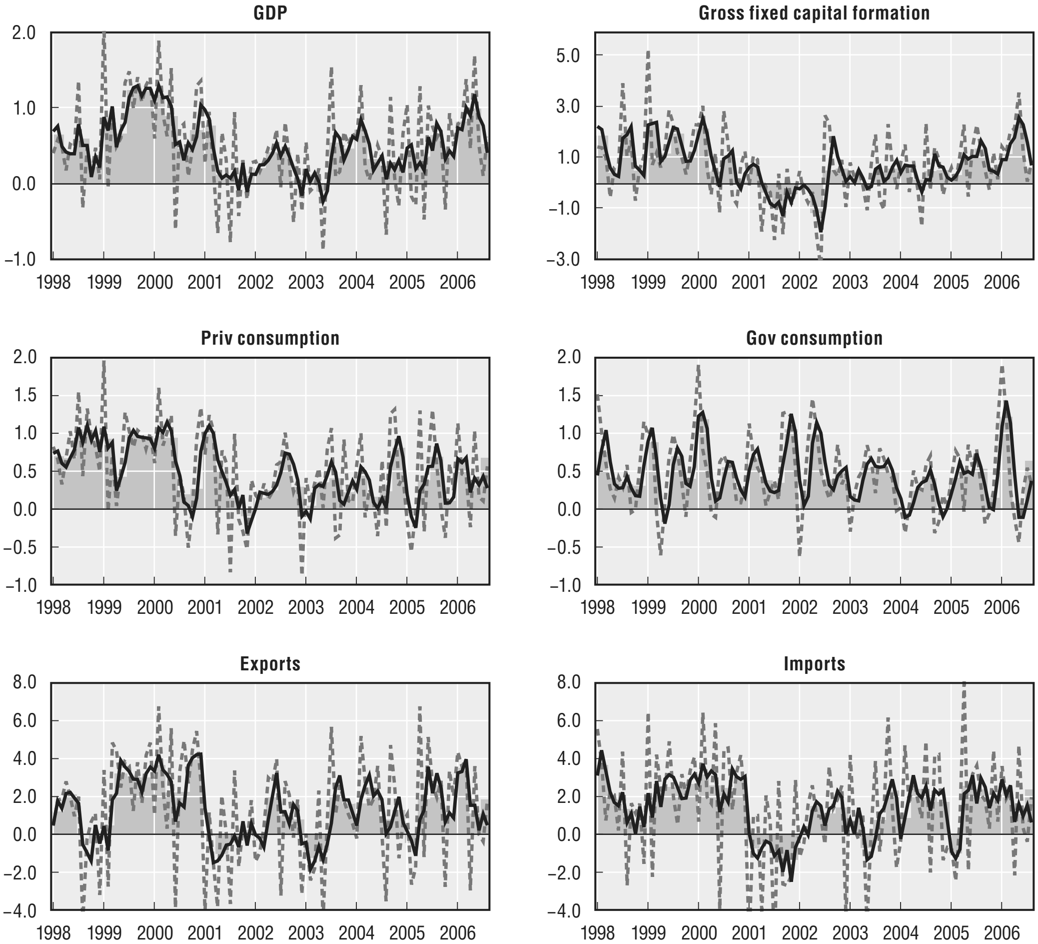
Table A.3 shows that, with the exception of government consumption, estimates of λ are generally close to one or at least higher than 0.5. The Diebold-Mariano tests against the null hypothesis of equal forecast efficiency find the efficiency gains from the DFM over part of the horizon significant at the 10% level for GDP, consumption, exports, imports and VAD services.

3. Estimates of monthly national accounts

The smoothed estimates of growth in monthly GDP and its components from the DFM using constraint (5) are shown in Figure 2. The graph contains estimates of both three-month and month-on-month rates, multiplied by three, together with the observed quarterly rates. Note that these estimates are obtained from the Kalman smoother based on the entire data set $\mathcal{Z}_T$.

Angelini *et al.* (2006) compare factor-based interpolation methods with the traditional method by Chow and Lin (1971), and conclude that both methods fare well. We therefore inspect estimates of monthly GDP growth from applying the Chow-Lin method to a single equation. Following existing studies on estimating euro area monthly GDP (Mitchell *et al.*,

Figure 2. **Estimates of monthly growth in GDP and demand components**



Notes: The grey bars show quarterly growth in the component, while the bold and thin lines show estimates of three-month and month-on-month growth rates, respectively. The latter are multiplied by three.

2005a, 2005b; Proietti and Frale, 2007), we choose euro area industrial production in manufacturing, total employment, the business confidence indicator and retail sales as explanatory variables.[9]

Figure 3 demonstrates high correspondence among estimates of monthly growth rates from the two methods, with a contemporaneous correlation of 0.86 among the monthly series over the period of 1998M1 to 2006M6. This reflects the fact that the Kalman smoother attaches high weights to items of industrial production and, to a lesser extent, business surveys, when backcasting monthly growth rates. In this case, the DFM effectively uses information similar to what has been chosen in the aforementioned studies.

Bańura and Rünstler (2010) have shown, however, that the weights of individual series in quarterly GDP forecasts may change considerably with the forecast horizon. From contribution analysis, as in equation (10), it can be shown that the same applies to estimates and forecasts of monthly growth. Table 4 presents the mean absolute contributions (MACs) of individual data groups to the forecasts of monthly GDP growth. Sample contributions have been estimated from the same pseudo real-time forecast design over the period 2000Q1 to 2006Q2 as used in Section 2.

Figure 3. **Estimates of monthly GDP growth**

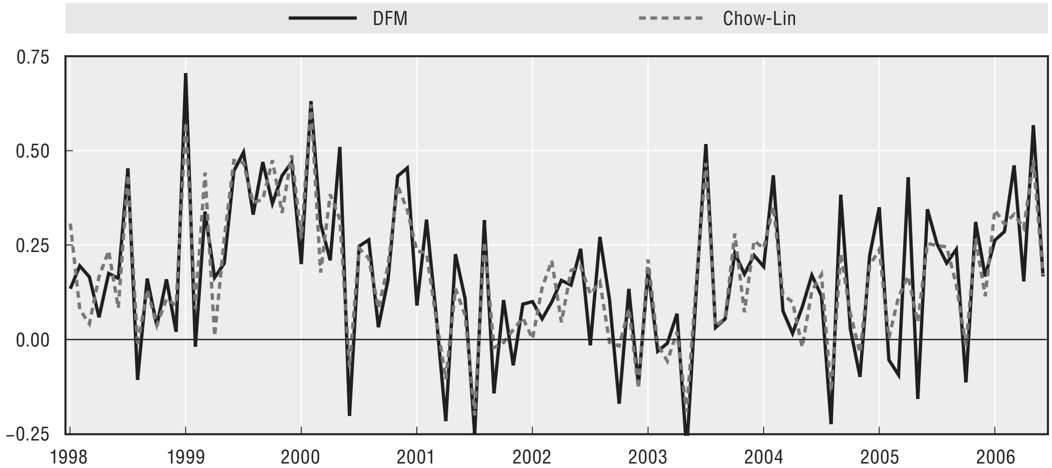


Table 4. **MAC to forecasts of monthly GDP growth**

Fcst	Industr Prod	Surveys	Financial	Int'l	Labour	Money	GDP
9	0.85	0.18	0.18	0.02	0.01	0.01	0.18
8	0.90	0.18	0.19	0.02	0.00	0.00	0.18
7	0.90	0.22	0.30	0.05	0.01	0.01	0.11
6	0.44	0.63	0.35	0.06	0.01	0.01	0.11
5	0.21	0.32	0.33	0.10	0.01	0.01	0.11
4	0.10	0.21	0.20	0.04	0.00	0.00	0.00
3	0.05	0.14	0.21	0.03	0.00	0.00	0.00
2	0.06	0.07	0.12	0.05	0.00	0.00	0.00
1	0.04	0.11	0.10	0.02	0.00	0.00	0.00

Notes: MACs are calculated as the sum of contributions from individual series belonging to a given group (see Table A.1). They are divided by the mean absolute deviation of monthly GDP growth. The GDP column shows the MAC of quarterly GDP growth. The sum of MACs across data groups exceeds one because in some periods contributions are of conflicting sign.

Table 4 shows the mean absolute values of the contributions of data groups as defined in Table A.1. The table demonstrates the shifts in the contributions of the individual data groups over the forecast horizon. Forecasts 8 and 9 are actually estimates of monthly GDP, where quarterly GDP is already known. In this case, the model attaches very high weights to industrial production data. Note, also, that the relative weight of quarterly GDP growth itself is small, even though temporal aggregation constraints (3) and (4) are relevant. As the horizon increases, survey and financial data gain more weight relative to industrial production and (quarterly) GDP growth. For nowcasts of monthly GDP in the current quarter (*i.e.*, forecasts 4 to 6), the weight of survey data already exceeds that of industrial production series.

Overall, these findings parallel those of Bańbura and Rünstler (2010) for quarterly GDP forecasts. They indicate that equations designed to estimate historical monthly growth in GDP, and therefore that rely heavily on industrial production data, are not necessarily optimal for the purpose of assessing the economic stance in real time.

4. Conclusions

The paper has combined the dynamic factor model of Doz *et al.* (2005) with equations to obtain monthly estimates and short-term forecasts of quarterly growth in the national accounts. The model contains the necessary temporal aggregation and national accounting identities. Hence, monthly estimates and short-term forecasts of the quarterly accounts are mutually consistent, which has advantages when the model is used for monitoring economic developments in real time.

For GDP and a number of components, the model beats forecasts from quarterly models-based time series and from bridge equations using selected indicators. One exception is public and private consumption, for which all forecasts remain largely uninformative.

Ex post estimates of monthly GDP growth are similar to those derived from single equation methods that employ monthly indicators, such as industrial production and confidence indicators. Our findings suggest, however, that equations that have been designed to estimate historical monthly growth in GDP are not necessarily optimal for assessing the economic status in real time.

Our forecast exercise seeks to replicate the real-time application of the models as closely as possible. However, we did not have a real-time data set at hand. Clearly, the properties of short-term forecasts based on real-time data are important for future research. The findings of Diron (2008) from forecasting euro area GDP with a small real-time data set indicate that this may increase the relevance of survey and financial data.

Notes

1. In the dynamic representation of a DFM, series x_t are related to dynamic factors g_t from $x_t = \lambda(L)g_t + \xi_t$ with some finite lag polynomial $\lambda(L)$. Defining $f_t^{'} = (g_t^{'}, g_{t-1,...}^{'})$ as the stacked vector of current and lagged values of g_t gives equation (1) (see Stock and Watson 2005).
2. Giannone *et al.* (2008) use a two-step approach to forecast quarterly GDP growth from the factor model. In a first step, they obtain forecasts of the latent factors, as from the state-space model given by equations (1) and (2). In a second step, quarterly GDP is predicted from quarterly aggregates of forecasts by means of a static regression.
3. This notation assumes that $t = 1$ represents the first month of a certain quarter.
4. With ε_t being white noise, ε_t^Q follows an MA(1) process with coefficient 4/19. This does not affect the consistency of estimates from the quarterly version of equation (6). Doz *et al.* (2006b) present an expectation-maximisation (EM) algorithm to obtain maximum likelihood estimates, but report few gains in forecasting performance.
5. Boivin and Ng (2006) have argued that using smaller data sets may improve the forecasts. However, series selection methods for unbalanced data sets have not been established thus far. Since our aim is to forecast not only GDP but also demand components, we opt for a reasonably large data set.
6. We have value added data beginning only in 1995Q1. Hence, estimates of equation (6) for these series start in this period.
7. The two model versions, hence, contain complete breakdowns of GDP and total value added, respectively, and κ_t in equation (5) reflects only approximation errors. The model without the constraint is almost equivalent to running models that contain a single quarterly series (the case of $m = 1$).
8. The series used for GDP are: industrial production in manufacturing, retail sales, new car registrations, the unemployment rate, and the European Commission business, consumer, building and retail confidence indices. They are used in bridge equations for euro area GDP

proposed by Rünstler and Sédillot (2003) and Diron (2008). We have also experimented with some specific equations proposed in these studies. They did not outperform our approach.

9. The aforementioned studies actually use more series as they derive monthly GDP as the sum of interpolates of its value-added components. For the most part, estimates use sectoral equivalents of our series. Other equations give similar results, as long as major industrial production items are included.

References

Angelini, E. J. Henry, and M. Marcellino, 2006, "Interpolation and backdating with a large information set", *Journal of Economic Dynamics and Control* 30(12), 2693-2724.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin and G. Rünstler, 2008, "Short-term forecasts of euro area GDP growth", *ECB working paper* 949.

Baffigi, A., R. Golinelli, and G. Parigi, 2004, "Bridge models to forecast the euro area GDP", International *Journal of Forecasting* 20(3), 447-460.

Bai, J. and S. Ng, 2002, "Determining the number of factors in approximate factor models", *Econometrica*, 70(1), 191-221.

Bańbura , M. and G. Rünstler, 2010, "A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP", *International Journal of Forecasting*, forthcoming.

Banerjee, A., M. Marcellino, and I. Masten, 2005, "Leading indicators for euro-area inflation and GDP growth", *Oxford Bulletin of Economics and Statistics*, 67, 785-813.

Boivin, J. and S. Ng, 2006, "Are more data always better for factor analysis?", *Journal of Econometrics* 132(1), 169-194.

Breitung, J. and C. Schumacher, 2008, "Real-time forecasting of GDP based on a large factor model with monthly and quarterly data", *International Journal of Forecasting* 24, 368-398.

Clements, M. and D. Hendry, 1998, "Forecasting Economic Time Series", Cambridge University Press: Cambridge.

Chow, G. and A. Lin, 1971, "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series", *Review of Economics and Statistics* 53, 372-375.

Diron, M., 2008, "Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data", *Journal of Forecasting*, 27, 371-390.

Doz, C., D. Giannone, and L. Reichlin, 2006a, "A two-step estimator for large approximate dynamic factor models based on Kalman filtering", THEMA Working Papers 2006-23, Université de Cergy-Pontoise.

Doz, C., D. Giannone, and L. Reichlin, 2006b, "A quasi maximum likelihood approach for large approximate dynamic factor models", CEPR Discussion Paper No. 5724.

Durbin, J. and S.J. Koopman, 2001, "Time Series Analysis By State Space Methods", Oxford University Press: Oxford.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin, 2000, "The generalized dynamic factor model: identifi cation and estimation", *The Review of Economics and Statistics*, 82, 540-554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin, 2005, "The generalized factor model: one-sided estimation and forecasting", *Journal of the American Statistical Association* 100, 830-40.

Giannone, D., L. Reichlin, and D. Small, 2008, "Nowcasting: the real-time informational content of macroeconomic data", Journal of Monetary Economics, 55(4), 665-676.

Hall, R. 1988, "Intertemporal substitution in consumption", Journal of the Political Economy 96(2): 339-352.

Harvey, A.C., 1989, "Forecasting, Structural Time Series Analysis, and the Kalman Filter", Cambridge University Press: Cambridge.

Harvey, A.C. and S.J. Koopman, 2003, "Computing observation weights for signal extraction and filtering", *Journal of Economic Dynamics & Control*, 27, 1317-1333.

Harvey, D., Leybournce, S. and P. Newbold, 1997, "Testing the equality of prediction mean squared errors", *International Journal of Forecasting* 13, 281-291.

Kitchen, J. and R. Monaco, 2003, “Real-time forecasting in practice: The U.S. Treasury staff’s real-time GDP forecast system”, *Business Economics* 38(4), 10-19.

Mariano, R.S., and Y. Murasawa, 2003, “A new coincident index of business cycles based on monthly and quarterly series”, *Journal of Applied Econometrics*, 18, 427-443.

McKinnon, J.G., 1991, “Critical values for co-integration tests”, in: R.F. Engle and C.W.J. Granger (eds.), Long-run Economic Relationships Oxford University Press: Oxford.

Mitchell, J., R. Smith, M. Weale, S. Wright and E. Salazar, 2005a, “An indicator of monthly GDP and an early estimate of quarterly GDP Growth”, *Economic Journal* 115, 108-129.

Mitchell, J. and M. Weale, 2005b, “A monthly indicator of GDP for the euro area”, paper presented to the Workshop on frontiers in benchmarking techniques and their application to offical statistics, Eurostat.

Proietti, T. and C. Frale, 2007, “A monthly indicator of euro area GDP”, *Econometrics Journal* 9(3), 357-372.

Proietti, T., 2008, “Estimation of Common Factors under Cross-Sectional and Temporal Aggregation Constraints: Nowcasting Monthly GDP and its Main Components.”, mimeo.

Rünstler, G. and F. Sédillot, 2003, “Short-term estimates of GDP by means of monthly data”, ECB Working Paper No. 176.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. den Reijer, A. Jakataine, P. Jelonek, A. Rua, K. Ruth, C. van Nieuwenhuyze, 2010, “Short.term forecasting of GDP using large data sets”, *Journal of Forecasting* 28(7), 595-61.

Stock, J. H., and M. W. Watson, 2002, “Macroeconomic forecasting using di¤usion indexes”, *Journal of Business and Economics Statistics*, 20, 147-162.

Stock, J. H. and M. W. Watson, 2005, “Implications of dynamic factor models for VAR analysis”, Princeton University, mimeo.

ANNEX

Table A.1. **Data**

No.	Series	Group	Publication lag (months)	Transformation code	Dickey-Fuller test
1	IP-Total industry	IndProd	3	2	−2.930
2	IP-Total Industry (excl construction)	IndProd	2	2	−2.911
3	IP-Manufacturing	IndProd	2	2	−2.962
4	IP-Construction	IndProd	3	2	−5.343
5	IP-Total Industry excl construction and MIG Energy	IndProd	2	2	−2.880
6	IP-Energy	IndProd	2	2	−6.274
7	IP-MIG Capital Goods Industry	IndProd	2	2	−2.985
8	IP-MIG Durable Consumer Goods Industry	IndProd	2	2	−3.657
9	IP-MIG Energy	IndProd	3	2	−7.054
10	IP-MIG Intermediate Goods Industry	IndProd	2	2	−3.246
11	IP-MIG Non-durable Consumer Goods Industry	IndProd	2	2	−4.942
12	IP-Manufacture of basic metals	IndProd	2	2	−3.192
13	IP-Manufacture of chemicals and chemical products	IndProd	2	2	−4.555
14	IP-Manufacture of electrical machinery and apparatus	IndProd	2	2	−3.207
15	IP-Manufacture of machinery and equipment	IndProd	2	2	−3.137
16	IP-Manufacture of pulp, paper and paper products	IndProd	2	2	−4.476
17	IP-Manufacture of rubber and plastic products	IndProd	2	2	−3.591
18	Retail trade, except of motor vehicles and motorcycles	IndProd	2	2	−3.712
19	New passenger car registrations	IndProd	1	2	−4.893

Notes: Transformation code: 1 = 3-month difference, 2 = 3-month growth rate.
Augmented Dickey-Fuller tests: 5% and 10% critical values for the t-statistics are −2.86 and −2.56, respectively (McKinnon, 1991).

Table A.1. **Data** *(cont.)*

No.	Series	Group	Publication lag (months)	Transformation code	Dickey-Fuller test
20	Unemployment rate, total	Labour	2	3	−4.551
21	Index of Employment, Construction	Labour	3	2	−2.299
22	Index of Employment, Manufacturing	Labour	3	2	−1.946
23	Index of Employment, Total Industry	Labour	3	2	−3.422
24	Index of Employment, Total Industry (excluding construction)	Labour	3	2	−1.995
25	Industry Survey: Industrial Confidence Indicator	Surveys	0	1	−3.414
26	Industry Survey: Production trend observed in recent months	Surveys	0	1	−3.501
27	Industry Survey: Assessment of order-book levels	Surveys	0	1	−3.461
28	Industry Survey: Assessment of export order-book levels	Surveys	0	1	−3.637
29	Industry Survey: Assessment of stocks of finished products	Surveys	0	1	−3.940
30	Industry Survey: Production expectations for the months ahead	Surveys	0	1	−3.558
31	Industry Survey: Employment expectations for the months ahead	Surveys	0	1	−3.526
32	Industry Survey: Selling price expectations for the months ahead	Surveys	0	1	−3.059
33	Consumer Survey: Consumer Confidence Indicator	Surveys	0	1	−3.363
34	Consumer Survey: General economic situation over last 12 months	Surveys	0	1	−2.680
35	Consumer Survey: General economic situation over next 12 months	Surveys	0	1	−3.688
36	Consumer Survey: Price trends over last 12 months	Surveys	0	1	−3.094
37	Consumer Survey: Price trends over next 12 months	Surveys	0	1	−3.485

Notes: Transformation code: 1 = 3-month difference, 2 = 3-month growth rate.
Augmented Dickey-Fuller tests: 5% and 10% critical values for the t-statistics are −2.86 and −2.56, respectively (McKinnon, 1991).

Table A.1. **Data** *(cont.)*

No.	Series	Group	Publication lag (months)	Transformation code	Dickey-Fuller test
38	Consumer Survey: Unemployment expectations over next 12 months	Surveys	0	1	−3.317
39	Construction Survey: Construction Confidence Indicator	Surveys	0	1	−3.191
40	Construction Survey: Trend of activity compared with preceding months	Surveys	0	1	−4.188
41	Construction Survey: Assessment of order books	Surveys	0	1	−3.013
42	Construction Survey: Employment expectations for the months ahead	Surveys	0	1	−4.033
43	Construction Survey: Selling price expectations for the months ahead	Surveys	0	1	−3.877
44	Retail Trade Survey: Retail Confidence Indicator	Surveys	0	1	−3.487
45	Retail Trade Survey: Present business situation	Surveys	0	1	−3.635
46	Retail Trade Survey: Assessment of stocks	Surveys	0	1	−6.208
47	Retail Trade Survey: Expected business situation	Surveys	0	1	−3.777
48	Retail Trade Survey: Employment expectations	Surveys	0	1	−4.257
49	Total trade - Intra Euro 12 trade, Export Value	Int'l	2	2	−2.708
50	Total trade - Extra Euro 12 trade, Export Value	Int'l	2	2	−3.527
51	Total trade - Intra Euro 12 trade, Import Value	Int'l	2	2	−3.130
52	Total trade - Extra Euro 12 trade, Import Value	Int'l	2	2	−3.043
53	US, Unemployment rate	Int'l	1	1	−2.748
54	US, IP total excl construction	Int'l	1	2	−2.621
55	US, Employment, civilian	Int'l	1	2	−3.482
56	US, Retail trade	Int'l	1	2	−3.351
57	US, Production expectations in manufacturing	Int'l	0	1	−5.429

Notes: Transformation code: 1 = 3-month difference, 2 = 3-month growth rate.
Augmented Dickey-Fuller tests: 5% and 10% critical values for the t-statistics are −2.86 and −2.56, respectively (McKinnon, 1991).

Table A.1. **Data** *(cont.)*

No.	Series	Group	Publication lag (months)	Transformation code	Dickey-Fuller test
58	US, Consumer expectations index	Int'l	0	1	−5.082
59	World market prices of raw materials in Euro, total, HWWA	Int'l	0	2	−3.574
60	World market prices of raw materials in Euro, total, excl energy, HWWA	Int'l	0	2	−3.503
61	World market prices, crude oil, USD, HWWA	Int'l	1	2	−4.064
62	Gold price, USD, fine ounce	Int'l	0	2	−2.596
63	Brent Crude, 1 month fwd, USD/BBL converted in euro	Int'l	0	2	−4.356
64	ECB Nominal effective exch. rate	Financial	0	2	−3.117
65	ECB Real effective exch. rate CPI deflated	Financial	0	2	−3.083
66	ECB Real effective exch. rate producer prices deflated	Financial	0	2	−3.151
67	Exch. Rate: USD/EUR	Financial	0	2	−3.348
68	Exch. Rate: GBP/EUR	Financial	0	2	−3.306
69	Exch. Rate: YEN/EUR	Financial	0	2	−3.088
70	Eurostoxx 500	Financial	0	2	−3.138
71	Eurostoxx 325	Financial	0	2	−3.226
72	US S&P 500 composite index	Financial	0	2	−2.782
73	US, Dow Jones, industrial average	Financial	0	2	−3.383
74	US, Treasury Bill rate, 3-month	Financial	0	1	−2.072
75	US Treasury notes & bonds yield, 10 years	Financial	0	1	−3.527
76	10-year government bond yield	Financial	0	1	−3.537
77	3-month interest rate, Euribor	Financial	0	1	−3.020
78	1-year government bond yield	Financial	0	1	−2.665
79	2-year government bond yield	Financial	0	1	−2.864
80	5-year government bond yield	Financial	0	1	−3.286
81	Index of notional stock - Money M1	Money	2	2	−3.013
82	Index of notional stock - Money M2	Money	2	2	−2.834
83	Index of notional stock - Money M3	Money	2	2	−2.636
84	Index of Loans	Money	2	2	−2.031
85	Money M2 in the U.S.	Money	2	2	−2.087

Notes: Transformation code: 1 = 3-month difference, 2 = 3-month growth rate.
Augmented Dickey-Fuller tests: 5% and 10% critical values for the t-statistics are −2.86 and −2.56, respectively (McKinnon, 1991).

Table A.2. **RMSE for GDP for subsamples (relative to the naive forecast)**

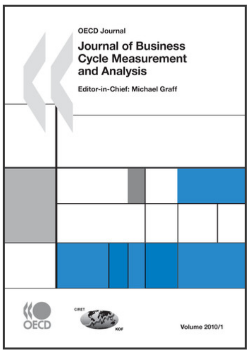
	2000Q1 - 2003Q3					2003Q4 - 2006Q2				
Fcst	Naive	AR(1)	QVAR	BEQ	DFM	Naive	AR(1)	QVAR	BEQ	DFM
8										
7	0.375	0.77	0.74	0.89	0.59	0.282	0.91	0.90	0.99	0.80
6	0.375	0.77	0.74	0.91	0.60	0.282	0.91	0.90	1.04	0.80
5	0.375	0.77	0.75	0.94	0.60	0.282	0.91	0.90	1.05	0.97
4	0.385	0.99	0.98	0.94	0.60	0.284	0.96	1.00	1.04	0.88
3	0.385	0.99	0.98	0.96	0.64	0.284	0.96	1.00	1.07	0.85
2	0.385	0.99	0.98	1.00	0.71	0.284	0.96	1.00	1.06	0.84
1	0.392	1.03	1.04	1.01	0.74	0.287	1.05	1.11	1.06	0.85

Note: The table shows the RMSE of the naive forecast and, for the remaining models, the RMSE relative to the naive forecast.

Table A.3. **Encompassing and Diebold-Mariano tests**

(2000Q1 - 2006Q4)									
Fcst	GDP	Priv cons	Gov cons	GFCF	Exp	Imp	VAD total	VAD Ind	VAD serv
Encompassing tests									
7	0.939	0.899	0.055	0.630	0.683	0.922	0.961	0.612	1.094
4	0.952	0.151	−0.237	0.783	0.945	0.978	0.908	0.558	1.254
1	0.892	0.716	−0.111	0.720	0.808	0.827	0.813	0.512	1.186
Diebold-Mariano tests									
7	* 1.538	** 2.327	−1.113	0.309	* 1.446	1.080	1.216	0.527	0.957
4	0.952	0.505	−0.628	0.534	** 2.767	* 1.583	1.211	0.281	* 1.282
1	0.259	0.582	−0.442	0.262	0.932	0.526	1.204	0.511	* 1.281

Notes: The upper panel of the table shows coefficient λ from encompassing regression (13). The lower panel shows the Diebold-Mariano statistics using the small-sample correction by Harvey et al. (1997).
* Significant at the 5% level.
** Significant at the 10% level.



From:

OECD Journal: Journal of Business Cycle Measurement and Analysis

Access the journal at:
https://doi.org/10.1787/19952899

Please cite this article as:

Angelini, Elena, Marta Banbura and Gerhard Rünstler (2010), "Estimating and forecasting the euro area monthly national accounts from a dynamic factor model", *OECD Journal: Journal of Business Cycle Measurement and Analysis*, Vol. 2010/1.

DOI: https://doi.org/10.1787/jbcma-2010-5kmmsxgf2qbs