Chapter 8

Advanced analytics

Daniel Sinnott and Rachel O'Carroll
Office of the Revenue Commissioners, Ireland

"Advanced analytics" encompasses a set of techniques to uncover insights from data to inform decisions and to test policies and interventions. From its initial use in the selection of cases for audit, the scope of advanced analytics applications has broadened significantly, as has the amount of available data. Tax administrations now use analytic techniques to inform a wide range of actions, including optimising debt-management processes, improving filing rates and quality, delivering better taxpayer service, and understanding the wider impact of policy changes. Moreover, many of these applications now support real-time (or near real-time) operational processes.

This widening of the field has given rise to a range of new organisational and technical challenges, from establishing appropriate governance and project-management structures, to building representative datasets and selecting suitable modelling techniques.

This chapter summarises how Forum on Tax Administration (FTA) administrations are using advanced analytics today and outlines the principal management and governance challenges they face.

What is advanced analytics?

Advanced analytics is the process of applying statistical and machine-learning techniques to uncover insights from data. The aim is to better inform decisions about the deployment of resources and the design of interventions and policies. Most advanced analytics projects fall into one of two categories:

- ***Predictive analytics:*** This aims to anticipate likely problems by looking for patterns in historical data. For example, it can help to identify what aspects of a tax return are most likely to be inaccurate, or find significant anomalies in a particular tax return compared to other returns from similar taxpayers. This allows administrations to consider preventive measures (for example rephrasing a question, producing better guidance or pre-filling with third-party data) and helps in the selection of individual cases for audit.
- ***Prescriptive analytics:*** The techniques used in prescriptive analytics aim to uncover causal relationships, that is whether particular actions caused or just coincided with a change in taxpayer behaviour. For example, does a particular style or type of communication make a difference to on-time filing or does it make no significant difference? Prescriptive analytics thus allows for easier testing and refining of different (and often costly) compliance interventions.

It is worth noting that prediction based on previous examples or revealed behaviours as well as causal inference are ordinary, everyday tasks carried out by administrators and policy makers using their experience and judgement. The use of advanced analytics techniques simply carries out these tasks with more reliance on data rather than human judgement (which can show bias and be highly subjective). As the size and breadth of data sets increases, this gives tax administrations the opportunity to uncover a greater range of insights which might not otherwise be apparent. This can be the case for example where anomalies in a pattern may not be very significant at an individual level, but may be very large overall; or where there are long-held and ingrained assumptions about causality driving decisions which turn out not to be supported by the data.

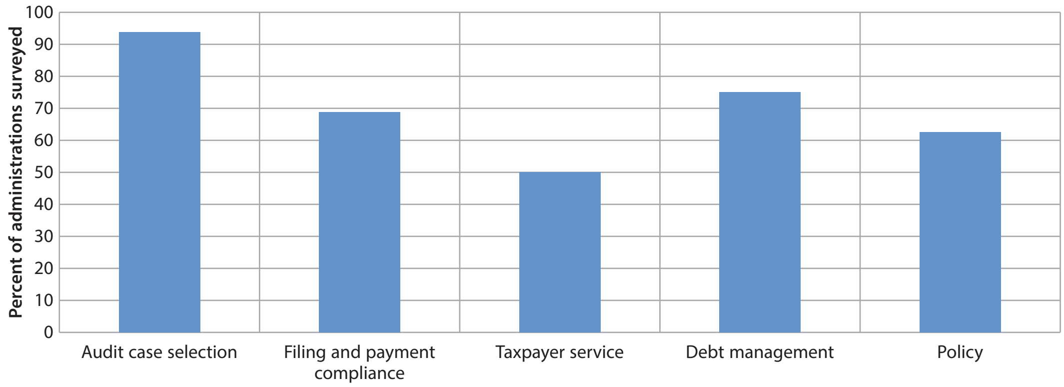
Application of advanced analytics

A key factor in determining the success of an advanced analytics project is the care taken to understand the nature both of the business problem at hand, and of the data available to address that problem. This assessment will determine the right analytical approach to be taken – whether to use supervised or unsupervised learning techniques, prescriptive modelling approaches, unstructured data, explanatory modelling, or other approaches. Administrations that take proper care over this decision will greatly improve their prospects of developing useful insights.

From its initial use by tax administrations in the selection of cases for audit, the scope of advanced analytics applications has broadened considerably. Analytic techniques are now used across all tax-administration functions and activities, with many administrations now using them to support real-time or near real-time processes.

The following figure provides an overview of how 16 FTA administrations have allocated their advanced analytics efforts across different operational areas.

Figure 8.1. **Use of advanced analytics**



StatLink http://dx.doi.org/10.1787/888933546811

Source: OECD (2016), *Advanced Analytics for Better Tax Administration: Putting Data to Work*, p. 20, Table 2.1, http://dx.doi.org/10.1787/9789264256453-en.

Audit case selection

For most tax administrations, the first use of advanced analytics was in assisting with the selection of cases for audit or verification. Standard advanced analytics techniques and approaches are well suited to this type of activity as they enable administrations to learn directly from the outcomes of past interventions. For example analysis of previous audits which have resulted in large adjustments might show that there was a pattern of particularly high or low values in some components of the tax return. This might then provide a basis for future audit selection (with further testing to check the validity of those insights). The set of techniques which allow administrations to identify such characteristics fall under the heading of supervised learning.

Where administrations have used such techniques with a broadly representative sample of taxpayers, this has proved to be a highly effective approach. However, where the sample used does not adequately reflect the wider population, supervised learning techniques may give unreliable results since the model developed can only learn about a particular segment of cases. In such scenarios, administrations have begun to apply techniques that identify anomalous taxpayers or returns by comparing outcomes across relevant peer groups (as opposed to learning from past interventions). While these models will not always target risk accurately (since some anomalies may be perfectly innocent), they are capable of uncovering wholly new insights into non-compliance. Such approaches generally fall into the category of unsupervised learning.

Filing & payment compliance and debt management

While not used as widely as in audit case selection, advanced analytics techniques are increasingly deployed to improve filing and payment compliance and the settlement of arrears. The analytical approaches used in these areas tend to be quite different from those used in pure case selection. In these, the task is not to predict how a taxpayer will behave in the absence of intervention, but how they might behave in response to a particular intervention. To tackle this problem administrations are beginning to build models that learn from the outcomes of controlled experiments in order to identify which cases should be subject to intervention, and which specific interventions should be carried out. This category of analysis usually referred to as “prescriptive analytics”.

Box 8.1. **Data mining models for non-filer programmes**

In ***Canada*** the Canada Revenue Agency (CRA) continues to refine its predictive models developed to assist in managing its non-filer programmes, which undertake a range of actions to obtain overdue returns. The models improve selection and prioritisation of cases, allow better workload management and improve business information and reporting. In its first year in operation, one of the non-filer models was responsible for the assessment of CAD 127.6 million of additional taxes. The CRA has also developed several other models to improve programme effectiveness and enhance taxpayer services by predicting self-resolution (i.e. which taxpayers will file without intervention) and responsiveness to a specific compliance action.

In addition to predictive techniques, CRA applies *prescriptive analytics* (methods which focus on finding the best course of action for a given situation) to support improved strategic and operational programme delivery. Prescriptive analytics is used to enrich the CRA's understanding of the non-filer population, optimise operational processes, and direct the application of compliance activities, allowing for more fact-based decisions. Complementing the use of predictive models, the non-filer programme is expanding its use of behavioural economics through nudge experiments to influence taxpayer compliance behaviour. "Nudges" are carefully designed interventions intended to steer people towards better decisions by altering the way that different choices are presented.

Source: OECD (2016), *Advanced Analytics for Better Tax Administration: Putting Data to Work*, p. 25, Box 2.1, http://dx.doi.org/10.1787/9789264256453-en.

Taxpayer service

The use of pro-active messaging, calling, and other interventions in anticipation of potential non-compliance has paved the way for administrations to look more closely at how advanced analytics can improve service delivery for taxpayers. Such uses are set to become of greater importance to tax administrations in the coming years as compliance and verification moves upstream. Wider use of "unstructured" data (e.g. customer emails, call transcripts, etc.) can help significantly in these efforts. This can uncover areas of common confusion or lack of knowledge, for example of filing deadlines or of how the tax rules work in particular areas. This then allows tax administrations to consider and test different interventions with different audiences in mind, for example proactive contacts with particular groups of taxpayers, different types of media (which may be appropriate to particular age groups) or the production of guidance among other things. Using advanced analytics in this way can also help in the design of machine rules, for example in telephone menus, or identifying the most effective way to present information on web pages.

Box 8.2. **Text mining of inbound emails**

In ***Singapore***, the Inland Revenue Authority of Singapore (IRAS) in 2014 began using text-mining techniques to analyse the content of emails received from taxpayers. The objectives of this project were to identify the nature of taxpayer inquiries and highlight important changes and trends that might require response. Text data from taxpayer correspondence was extracted, cleansed, and structured to derive patterns and insights. Close collaboration between the analysts and business users during the course of the project enabled findings to be contextualised and thereby improved the text mining process.

Box 8.2. **Text mining of inbound emails** *(continued)*

As a result, the IRAS was able to uncover insights, otherwise locked in textual data, on issues pertinent to taxpayers, for example in one project, text-mining helped to identify the common queries taxpayers had after an existing tax policy was changed. Based on this analysis, IRAS was able to launch timely and targeted campaigns, provide more guidance on its website, and proactively initiate updates to taxpayers impacted by the changes. These approaches reduced the need for taxpayers to contact the IRAS.

Ongoing tracking of the nature of email enquiries combined with the IRAS' existing analysis of structured data has helped the IRAS to identify trends on particular topics and to pre-empt or reduce contacts, thereby improving service delivery for taxpayers. Text mining has now replaced the manual tracking of email enquiries, which has saved time and improved staff productivity. It has also enabled the IRAS to track the nature of enquiries more objectively, avoiding the inconsistencies of interpretation typical of manual tracking.

Source: Singapore – Inland Revenue Authority of Singapore (2017).

Policy evaluation

Although most analytics work is carried out to support operational activity and decision-making, tax administrations are increasingly using analytics to assist in decision-making in relation to strategy and policy. The most common analytic applications in this field aim to evaluate or anticipate the impact of changes in tax policy. In general, the techniques used in this area come under the heading of explanatory modelling, since the main objective is to understand and explain the relationship between specific variables, rather than to predict the overall outcome. While it is often difficult to gauge the accuracy of such models, the approach can be effective where used to test a clearly articulated theory of behaviour.

An example of the use advanced analytics for policy assessment is China's creation of a general-equilibrium model to measure the economic and social impact of the introduction of a value added tax (VAT) in 2012. This model played a key role in the policy reform process by enabling the Chinese tax administration to trace the effects of the VAT changes on tax revenue, industry structure, social welfare, and a wide variety of other economic indicators.

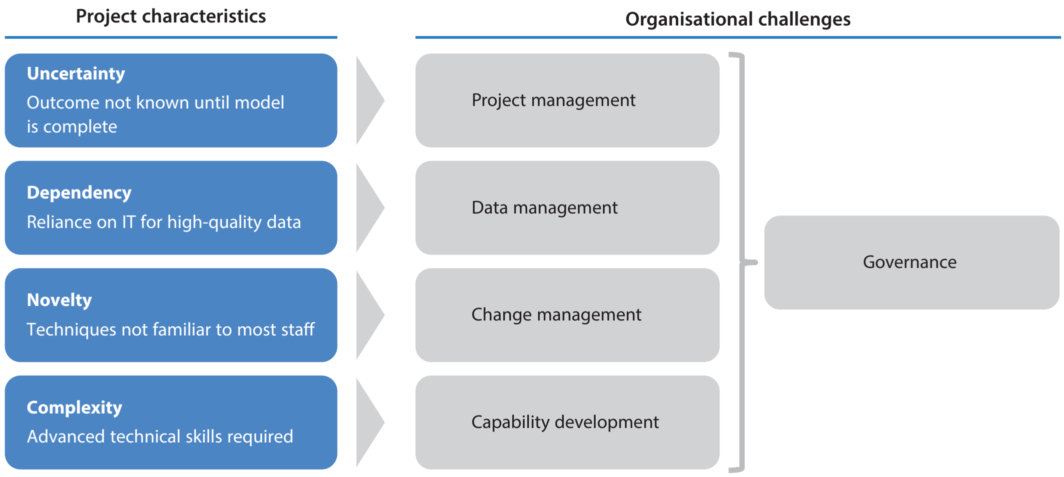
Organisational management of advanced analytics projects

Figure 8.2 illustrates how four key characteristics of advanced analytics projects give rise to major organisational and governance challenges.

Governance

Analytics governance requires a strong focus on integrating the business, information technology (IT), and analytics perspectives, and managing the uncertainty inherent in most advanced analytics projects. Many tax administrations have established integrated governance bodies to prioritise, resource, and oversee analytics projects. By consolidating analytics governance in a single, permanent body, administrations can begin to build expertise and experience across multiple projects.

Figure 8.2. **Key characteristics of advanced analytics projects**



Box 8.3. **Centralised governance of advanced analytics**

In ***Ireland*** in 2015, the Office of the Revenue Commissioners established a senior management group to prioritise and oversee all advanced analytics initiatives across the organisation. Prior to this, governance was organised on a project-by-project basis. While a number of effective models were introduced under this system, the absence of a centralised, permanent governance structure made it difficult to build organisational momentum behind analytics, and to maintain and upgrade the models that had been built.

The new senior management group – the Revenue Analytics Group (RAG) – is led by the Chairman of the Revenue Commissioners, and consists of representatives from the business, analytics, and IT functions. The RAG also has direct links into the key operational and IT governance bodies, the latter of which oversees the advanced analytics budget. Business intelligence initiatives are governed through a separate but linked structure.

This structure provides cohesive governance of all of Revenue's advanced analytics work: it aligns analytics projects to organisational priorities; it ensures that the analytics function works within the appropriate infrastructure; and it co-ordinates the activities of multiple units to ensure that analytics initiatives deliver a strong return on investment.

Source: OECD (2016), Advanced Analytics for Better Tax Administration: Putting Data to Work, p. 39, Box 4.1, http://dx.doi.org/10.1787/9789264256453-en.

Project management

The nature of advanced analytics projects (which are essentially an attempt to find a pattern that may or may not exist) creates significant uncertainty in relation to benefits and timelines. In many ways, advanced analytics initiatives are closer to research and development work than to ordinary IT or business projects. For analytics functions to deliver value, administrations must find ways to manage this uncertainty. Most administrations address this problem by using iterative, "test-and-learn" approaches in order to gather regular feedback and deliver incremental improvements. Many have taken an exploratory approach to project prioritisation and management, tending to begin work on a wide range of areas, and narrowing their focus only as it becomes clear that a particular project is likely to yield results.

Data management

To get the most out of advanced analytics, it is essential that administrations recognise that "big data" may not be "useful data". Advanced analytics models can only learn from the data they are applied to. If this data is inaccurate, or incomplete, or subject to selection bias, then the value of any resulting model will be severely limited, regardless of the volume of data available. To address this issue, administrations should develop strategic approaches to data collection and management. Instead of seeing data simply as the residue of operational processes, administrations must treat it as an asset to be actively managed and developed. To this end, administrations may wish to consider sampling programmes, randomised controlled trials, and similar data-gathering exercises. In addition, administrations should invest in the development of data dictionaries to ensure that analysts and business users can fully understand the information they are working with. Finally, administrations should actively look to domestic third-party sources and data acquired through new sources, for example automatic exchange of information initiatives, to develop a more rounded picture of taxpayer characteristics and behaviour.

Change management

Administrations are deploying a variety of approaches to ensure that advanced analytics models are successfully brought out of the laboratory and into the field. These include demand-side measures such as training operational staff in understanding analytical principles, and supply-side measures such as establishing specialist change-management units dedicated to analytics implementation. Perhaps the most promising approaches are those that combine demand and supply-side elements. In Norway, for instance, an analytical project will proceed only if the prospective business "client" is willing to second a member of staff to act as project manager. This ensures close collaboration between analytical and operational staff, and also imposes a useful check on the project prioritisation process.

Building capability

Finally, the inherent complexity of the modelling process creates challenges in relation to capability development. Tax administrations are working hard to secure and retain resources with the skills to assemble, clean, transform, and fit models to large datasets. These skills are scarce, and very much in-demand both in the private sector and academia. The need for highly-skilled technical staff is especially acute for administrations that wish to take advantage of the flexibility and low cost of open-source statistical programming languages (as distinct from commercial software packages). While competition to recruit analysts is intense, tax administrations do have certain competitive advantages: they can offer large and varied datasets, a diverse set of interesting analytical problems, and the opportunity to use advanced analytics to serve a wider public interest.

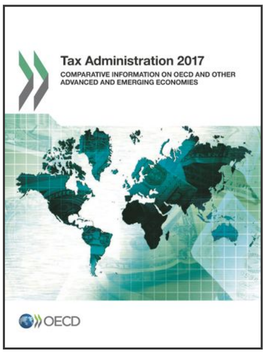
Next steps

Advanced analytics is now well established as a core capability in the decision-making and work-management approaches of advanced tax administrations. Its importance will only continue to increase as new data sources become available, including specific tax data such as that received through enhanced automatic exchange of information as well as data from non-tax sources, including other parts of government. It is also important that tax administrations keep abreast of how advanced analytics is being used across public administration and business and consider what insights or lessons this may offer them for their own work.

The nature of advance analytics – essentially the automated search for useful patterns in large data sets – means that success is not guaranteed and many seemingly promising initiatives may turn out to be dead-ends. Care is therefore needed in selecting projects, designing analytical approaches, curating data, and evaluating results. Administrations must be prepared to make a substantial investment of time and effort if they wish to turn raw data and computing power into practical, actionable insights.

References

OECD (2016), *Advanced Analytics for Better Tax Administration: Putting Data to Work*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264256453-en.



DOI: https://doi.org/10.1787/tax_admin-2017-13-en