

Influence of Governance Issues on the Quality of Official Statistics

Edvard Outrata

Member of European Statistics Governance Advisory Board

Prague, Czech Republic, e.outrata@centrum.cz

Abstract

The constant growth of the scope and depth of evidence-based decision making in government throughout the 19th and 20th centuries has led countries to develop the concept of statistical independence of political and other vested interests. This independence is the most important condition necessary for maintaining and increasing data quality, yet the institutional position of statistical offices as part of government makes this hard to achieve. The institutions and governance of statistics in many countries, including developed ones, still leave much to be desired. In spite of this, history of the last several decades suggests there is systematic improvement in the governance of statistics within countries, with legislation and e.g. the acceptance of codes of practice ensuring this independence, etc. Recent developments have, however, introduced a relatively new danger in the rapid development of automatism, where evidence-based decision making is being replaced by automatic decisions that virtually, or even fully, leave the responsible human decision maker out of the process. This development, which seems to be a welcome gradation of evidence-based decision making, is in fact a grave danger for the independence and quality of statistics, because it re-orientates the target of lobbying from the decision makers (who are no longer in the loop) to the data sources, thus putting the quality of official statistics under new powerful pressures. In this context it is more necessary now than ever before to protect the statistical system from these influences by ensuring full independence in the governance of statistics.

Keywords: evidence-based decision making; automatism; statistical independence; governance of statistics.

1. Evidence-based decision making.

The period of enlightenment in the 18th century brought forward the idea that important decisions should be based on facts and provable theory rather than on hunches, creed and ideology. This major paradigm shift affected all aspects of human life and did so gradually throughout the 19th and 20th centuries. While generally considered axiomatic by most people, its practical application is still rather patchy and uncertain. This is a paradox caused mainly by human nature as it developed over millions of years, where the success of the human race in cutting out a living depended on decision-making mechanisms that did not spend time and energy on gathering evidence and had to rely on other ways of reaching a result.

In particular, this is apparent in the area of governance of public affairs. When the importance of finding and considering evidence increased due to the paradigm shift described above, many of the other more traditional mechanisms remained. Among these were formal consultation, formalized delays to allow better consideration where possible, etc.; and, most importantly, accountability. Nevertheless, the relative weight and cost of gathering and processing data in government structures grew rapidly.

During the 19th century it became obvious that governments needed systematic gathering and processing of evidence about the numbers, wealth and economy of the people they governed in order to produce credible and effective decisions. Official statistics became institutionalised in one country after another, mostly in the form of national statistical offices. These were then responsible for providing

government decision makers with a credible up-to-date picture of the country, its people, its wealth and a growing number of other social and economic aspects as these became relevant for governing the nation. As the growing wealth and strength of our civilisation was being attributed to rational, therefore evidence-based, decision making, nations were willing to devote more resources to the provision of reliable data, and so to the growth of official statistics.

The model which thus developed along similar (but not identical) lines in developed nations always contained a built-in contradiction. On the one hand it was built to provide objective information to feed into government (and other) decision making processes, thus helping governments to make decisions that were objectively of better quality. Therefore, it had to be both objective and professional, i.e. not influenced a priori by the decision making process and up-to-date on the optimal methods science could provide at the time. On the other hand, by being part of the government structure it was always suspect of providing a biased government rationale for pre-ordained decisions or justifying ex-post unpopular actions. For a long time (and often to this day) formal structures of government have maintained in existence channels through which government might influence what evidence the model produces. This is reflected in a paradox in the view most citizens have of their official statistics: in practice and if it suits them they firmly believe in their objectivity, but if asked about their credibility, they reply with doubts. This differs, of course, from country to country and time to time, but mostly is well reflected by Churchill's old quip that he believed only those statistics he had biased himself.

As the complexity grew of both governance and statistics, and also with the strengthening of representative democracy in the world in the 19th and early 20th century, the need to protect the objectivity of the data and analysis prevailed over other considerations, and independence of statistical offices became the declared standard.

2. Automatism

The growth of complexity of governance combined with the increasing sophistication of statistical theory in the second half of the 20th century led to symptoms of information overload in society. While on the one hand more and more relevant information was becoming available on the state and development of society, people responsible for making decisions found themselves more and more often unable to absorb all the relevant information necessary for evidence-based decision making. The obvious solution, namely having specialised staff prepare comprehensible alternatives and choosing from among them after broader discussions, started to fail under information overload. Firstly, the large amount of information available required screening of data, while this screening in turn was by nature subjective and prone to prejudice. Secondly, choosing clear-cut meaningful and comprehensible alternatives became impossible, so that in the broader discussions new alternatives were being brought up that could no longer be easily either identified with one of the originally proposed ones, or falsified. Thus, thirdly, adherence to a certain set of alternatives started competing with another set of alternatives without a rational solution being possible within the given framework. Discussion instead of leading to solutions resulted with increasing regularity in a dead end. Then, fourthly, time pressure and the need to move ahead at any cost prevailed and decision makers reverted to pushing through one of the alternatives simply on the basis of a power advantage. That, fifthly, brought us full circle: however sophisticated and professional the evidence, decisions were being based again more on a hunch than on evidence (albeit coached in the form of erudite evidence-based processes).

The development described above was, of course, further enhanced by secondary influences. Decision makers, working under the described overload, are mostly politicians, whose claim to fame is not meticulous weighing of pros and cons in an office full of only partially comprehensible data, but rather the ability to simplify the complex and gain support of the uninitiated. If evidence was leading us

nowhere, while hunches gave easier answers, politicians were not unhappy. The world was coming back to where their profession started from and where their training implicitly helped them best. The ideal of enlightenment, viz. evidence-based decision making, was becoming compromised, because the only reason for sustaining this rather expensive contribution was the guaranteed quality of the eventual decision.

A clear answer to this development, however, appeared in the form of generalisation of decisions, a trend that started in the 1940's and developed massively in the second half of the 20th century. In some instances repetitive decisions had to be made on the basis of regularly published statistics and the result of these decisions was directly related to published figures and, therefore, fully predictable. In such a case, there did not seem to be a reason to repeat the decision making process each time (and there was often a reason not to do so in that every repetition of the process allowed unwelcome influences to enter the fray). So, a general decision was made to apply a certain formula to the regularly published statistic and use the result instead of a separate decision by the decision makers. The decision makers, of course, could still change the formula if necessary, but this would be unusual, as it would require the decision makers to reopen all the discussions and negotiations that the general decision was brought in to avoid in the first place.

Probably the oldest, and certainly the best known, such general decision is indexing payments by a measure of inflation (usually by CPI), e.g. salaries indexed by CPI, pensions indexed by CPI, rents indexed by property index or, more usually, by CPI, etc. In all these cases what used to be a considered decision made from time to time over certain evidence became an automatic effect of a formula. Its big strength is the simplicity of its application, which does not require extra work each time it is performed. Its weakness is its abstraction: indexation is performed in a pre-judged way, using often an imprecise measure that in practice may be insensitive to changed circumstances and variations of real performance. What seemed to be a fitting model, does not necessarily work in all circumstances, as e.g. when CPI is used for indexing instead of more precise measures of inflation, ignoring the fact that different social groups have different patterns of consumption. Of course, CPI indexing has become so much a part of everyday life that these disadvantages are rarely mentioned even in times like now, when they are significant because food prices rise while other consumer prices fall, so that using a common measure of inflation means poor pensioners get hurt more than the general population.

There is an important difference between general decisions over national statistics (like CPI indexing) and general decisions over separate individual decisions of a third party (like indexing interest rates to a central bank interest rate). In the first case, which we are interested in here, no individual decisions are made. CPI is not a decision of a third party, but supposed to be an objective measure. In the second case, there is a third party (the central bank), which is pursuing separate objectives and which is accountable for its actions. The provision is, therefore, more a delegation of responsibility to this third party. Indexing on LIBOR, of course, today seems to have migrated from the first category to the second, so this distinction is not as clear as it should be. The experience with the recent LIBOR debacle suggests that it might be interesting to investigate, how many of the "objective facts" we happily work with are actually not objective at all.

The introduction of indexing and other general decision making, and particularly its extension to many sensitive areas, has altered the role of the statistical office. It is no longer a simple provider of best up-to-date information to decision-makers, who in turn remain fully accountable for their decisions. Instead, it is slowly becoming a lonely arbiter between conflicting interests, applying as scrupulously as possible pre-arranged methodological rules whose principles are clear only to a small class of initiates (and whose detail is clear only to a very small number indeed) to a vast array of issues. The specifics of these affected issues are varied and mostly unknown. This is an awesome unsought responsibility which multiplies manyfold the requirement for pristine quality and discipline. This has not been widely

appreciated outside the statistical community, and it goes very much to the credit of statistical offices and statisticians that they have by and large lived up to this responsibility and have overall throughout maintained their credibility.

Recent technological developments particularly in the IT area have allowed this trend toward general decision making to expand further. General decisions are supposed to work automatically, leaving out of the loop not only the decision-maker, but actually any human middleman. This has not yet been "achieved" in statistics (where after all official figures are always approved by humans for technical correctness), but has reached dangerous levels in finance. The role of rating agencies is undergoing a similar, but much more radical change. This institution was created to provide summary information to investors to help them make evidence-based decisions and served them well as long as the final decision was made individually by the investor or his accountable agent. But once these decisions were generalized and rules were introduced for automatic transactions to be initiated as a result of a rating change, control of the process was lost and the way was open for serious financial swings. Soon these started to threaten the world with financial and economic crises.

Automatisms of this kind, whether in finance or in statistics, have significantly altered the underlying system of accountability in the complex decision-making system in government (using statistics) and in finance. In government the division of roles among elected politicians, appointed politicians and the professional civil service has been affected. In particular, politicians have relieved themselves from the responsibility for individual decisions, and the complex system is often left to be run as if on autopilot. This has in its turn put growing responsibility on the producers of information, namely the statisticians, to preserve the quality of data in a changing situation without anybody realising what was happening. The automatic character of the resulting process in my opinion significantly reduces the effectiveness of an important check in the system of governance, and is therefore dangerous for the smooth functioning of democracy. But this is still not the full scope of the problem.

3. Goodhart's Law and the Quality of Statistics

In the 1970's it became obvious that indicators that were being used as controls in the developing automatisms (as described above) were losing their usefulness. This process was already notorious in the Communist East. Thus, for example, gross material product figures were generally used there to justify quarterly premiums for employees and this practice resulted in making the GMP quarterly figures irrelevant as employees learnt to manipulate internal procedures to achieve the result they wanted. In the free market world this effect became first obvious in financing and economic modelling and led to the formulation of Goodhart's Law in 1975 (Goodhart, 1975), which for our purposes has best been expressed by Professor Marilyn Strathern as: "When a measure becomes a target, it ceases to be a good measure" (Strathern, 1997).

The recent rapid growth in the number of general decisions is best visible in Europe where there is a significant difference over time. If the process of accession to the EU, designed basically in the 1970's, is compared to that of the accession to the eurozone, designed in the 1990's, we notice a striking difference in the manner statistics (and data in general) are treated.

The standard process for accession to the EU was constructed as individual for every country (albeit requiring the acceptance of the *acquis* and managed according to common principles). There was no significant distortion of statistics in this process. Exceptions were negotiated and accountability was clear.

On the other hand, accession to the euro was formalised in the Maastricht criteria, which turned several basic statistics into controls. Here Goodhart's Law went into full force: in a race not to be left behind most countries strove to meet the criteria by means fair or foul, ranging from interpretation through using definition loopholes to reach the criteria up to outright fraud. While the criteria were meant to ensure that countries entering the eurozone were closely converging economies, it turned out that in practice they allowed all old members to enter (if they wanted). Later fiscal pacts were agreed only to be broken by the originators. Newly, further automatisms are being designed, where breaches of the fiscal pact are supposed to lead to automatic sanctions.

The EU has drawn two lessons from this history. Firstly it has strengthened somewhat central oversight of European statistics by adding some necessary powers to Eurostat and by creating the European Statistics Governance Advisory Board to monitor the European Statistical System and advise Parliament and the Council on governance issues in the European Statistical System. This is a step forward, but statistics still remain the responsibility of member states and European oversight remains relatively weak. Secondly, however, it seems to be increasing reliance on the automatic entry into force of sanctions for breaking rules controlled by basic statistics. This would significantly increase the danger to the quality of the statistics involved by triggering the impact of Goodhart's Law.

I have chosen my example from the EU both because I am its citizen and a member of ESGAB, and because the example in my view explains the problem better than other instances, due to the institutional lack of preparedness in the EU. The process of introducing automatisms, however, is a general one and appears everywhere, providing an alibi to politicians for shirking from their responsibilities and turning these responsibilities over to the "autopilot" and to data providers.

4. Danger to the Statistical System

As the number of automatisms grows, Goodhart's Law suggests that the indicators involved become less and less informative, whatever further controls are imposed on their formal quality, thus in a way emptying the information content of the data. This happens surreptitiously: from day to day nobody need notice, until at the moment of a crisis the underlying model on which government operates ceases to yield results. Whenever this is recognised, statisticians will seek to develop new indicators to provide better measurements of the phenomenon under scrutiny. In such an environment, however, even where such indicators are available (which is not very often), they will soon suffer the same fate as their discredited ancestor.

There are many instances of this development going on under our eyes. Gross domestic product and its growth used to be a good general measure of prosperity and its development in spite of its recognised imprecision. However, as GDP growth became an issue on which elections were fought, governments discovered first how to manipulate its periodicity (viz. how to postpone the fall until after the election). Once growth became the main indicator on which political success was judged, governments were nudged toward the *fata morgana* of permanent growth bought by growing debt. Much work was done in order to make the indicator more precise and timely, but as it still remained a target, it was no longer a "good measure". In the end we are now searching for a better indicator for the original phenomenon, namely prosperity and happiness, and not really succeeding. Of course, GDP is only one example, there are many more and nearly all the "sensitive" indicators today have become targets and are, therefore, affected by Goodhart's Law.

While statisticians are trying to save the indicators through quality improvement, governments (and institutions in general) are tightening up the rules. Earlier times, when countries could get away with breaching the rules and suffered at best only formal reprimands, are on their way out and there is a cry for

ensuring strict adherence. This again is most clearly visible in the European Union, where e.g. automatic sanctions, hard to bypass, are being proposed for breaching the Fiscal Pact. Consultants are being reproached for having helped some of the countries bypass rules in the past, so that such advice is no longer as easily available above the board. Indeed, a setup is sought, where countries will not be able to avoid consequences if they break the rules. This in turn makes living with the results and consequences of true statistical information much harder than it has been to this day.

This development at first sight seems to be positive: ideally countries will truly have to conform to the real meaning of the rules and (as long as the original model is correct) will not be able to disrupt the system by their irresponsible behaviour. There are, however, at least two caveats that are important in this context.

Firstly, the model may not be correct and valid for all instances. It is a corollary of Goodhart's Law that once an indicator is made a target, its relevance shifts, so that it is highly probable that the model no longer reflects the conditions that led to its creation. Even if this were not relevant, conditions change in time and the model may become inadequate in the new situation. Thus, the rigidity of the hard-coded rules may force developments that are neither optimal nor most acceptable, thereby hurting the system instead of ensuring its function.

Secondly, and more perniciously, by not leaving a relatively easy way of "changing the rules" on the fly, the system forces those who for whatever reason will not really comply to spend their energies not on modifying the specific decision (which route is no longer open as a result of the general decision codified in the automatism), but rather on affecting the data. And this can only be done by high-level political pressure on the statisticians.

The statistical system should, therefore, expect a significant increase in the amount of political pressure against the quality and relevance of statistics in step with these developments. It will require strengthening of ethics in the statistical system, and not least a review and subsequent tightening of institutional guarantees of statistical independence. In Europe at long last the danger of not acting to improve the level of statistical governance has been highlighted by the development of the continuing economic crisis and the role statistics played in causing and reacting to it. Eurostat has received certain further powers allowing it to intervene with countries that do not act according to the Code of Practice. In addition, EU member states are being requested to sign a Commitment for Confidence, in which governments (and not just statistical offices) commit their countries to implementing the Code of Practice, which includes strong guarantees of the independence of statistical offices.

Whether this is going to help significantly in protecting the quality of statistics in the future is yet to be seen. To this date, member countries have shied away from making this commitment. Whether further pressure from the Commission or Parliament will nudge countries to make this commitment is yet to be seen. Even if it does, however, this still does not guarantee political practice. Official statisticians will be left with much of the burden of protecting the profession and the quality of its product in these uneasy times.

References

- Goodhart, Charles (1975) *Monetary Theory and Practice*, p. 96
Goodhart, Charles (1981) *Problems of Monetary Management, The U.K. Experience*, p. 116
Strathern, Marilyn (1997) *Improving Ratings: Audit in the British University System*, *European Review* 5, 305-321