
04.11.2020

**Statistical Production and Economic Policymaking: A View from a
Central Bank**

2020 International Conference of Statistics for Society - INE

Pablo Hernández de Cos

Governor

General considerations

Let me first congratulate the Instituto Nacional de Estadística on this anniversary. Since its creation in the very difficult times after the Spanish civil war, INE has continuously grown in stature. Today, it boasts the highest standards in terms of professionalism and credibility.

We live in a society overcrowded with data. Every decision we take and every transaction or movement we make are reflected in some or another database. But this per se has not necessarily increased our real information or our knowledge. Many of those data are not structured, lack a precise definition or do not follow a specific methodology. This is why statistics and, in particular, official statistics are so important, as they transform data into real information. And we need this to make sound decisions. Policymaking thus heavily relies on timely, exhaustive and detailed information provided by official statistics.

Unsurprisingly, statistical production plays a prominent role in central banks. Several factors are at play here, such as a long-standing tradition in producing and analysing monetary and financial information, staff who are highly experienced in these tasks and good “raw materials”, in part benefitting from the key role of the central bank as the banking regulator and supervisor. I should stress that the autonomy of central banks has also greatly helped enhance the credibility of their statistical output.

Indeed, as part of the Eurosystem, as a participant in the Single Supervisory Mechanism and given its responsibility for the proper functioning and stability of the financial system, the Banco de España regularly collects a wealth of financial and supervisory information from credit institutions, which is key to performing all our duties. This information is also necessary for strengthening the performance of other functions of the Banco de España, such as the conduct of monetary policy, the oversight of the stability of the financial system and the compilation of statistics. We are also responsible for designing financial reporting as efficiently as possible and, in this endeavour, both the Banco de España and the European authorities have some substantial challenges ahead. I shall refer to these later.

Some of the key statistics produced by the Banco de España are present in the Spanish National Statistical Plan (the PEN, as we refer to it). They include some macroeconomic statistics, such as the balance of payments and the financial accounts, which complement and have to be consistent with those compiled by INE. In this regard, I think we can be proud of the long-standing tradition of close and candid cooperation between our two institutions. Joint discussion continuously enriches our professionals, and their efforts to ensure the consistency and complementarity of the results are a guarantee of the quality of the statistical system. This is a clear example of how the quality of the overall statistical system is higher than what the mere sum of independent products would be, were such products not well-connected.

Micro data as the Holy Grail

We can all agree that statistics have been undergoing a revolution in the last few years. Important technological advances have increased the ability to process huge amounts of data and productivity in this area has surged: policymakers and analysts have more and more information and details at their fingertips and now have the possibility of applying

more sophisticated models, designing more refined economic policy responses and answering more complex questions.

Some of these changes stem from the experience of the Great Financial Crisis of the last decade. Then it was felt that only the availability of detailed information on economic agents' positions (including cross-border ones) would allow the relevant policy authorities to thoroughly analyse the exposures of an increasingly interconnected global financial system and to better understand the monetary policy transmission mechanism and the massive deleveraging that followed. To close these gaps, the Eurosystem promoted a number of projects based on granular information. These included AnaCredit (to build up a Central Credit Register at the European level), the Securities Holding Statistics Database (which contains exhaustive security-by-security information on European portfolios) and Money Markets Statistical Reporting, where daily information at the transaction level is collected from financial institutions.

In Spain, we also learned the lesson. In 2013 we overhauled the information compiled by the Central Credit Register, switching more towards a very complete granular database in order to strengthen our micro- and macro-prudential supervisory mechanisms. We are also fortunate to have, since the eighties, an exhaustive database of the annual balance sheets and the profit and loss accounts of Spanish firms (the so-called Central de Balances), whose inception was one of the landmarks as far as the set of statistics produced by Banco de España is concerned. The database also contains a very valuable sample of firms voluntarily providing quarterly information. Moreover, the balance of payments and the international investment position have been historically produced by relying on mainly micro information. And from the outset of this century we have conducted every three years (which will be every two years in the future) a wave of the Survey of Household Finances, which offers a very rich picture of some of the key economic and financial decisions taken by Spanish households. All this micro information helps the Banco de España perform our functions and allows us to address flexibly both analytical work and statistical compilation, under our multi-purpose remit.

Moving towards a statistical landscape broadly relying on micro data requires deep changes in a number of processes. As automation is of the essence, new techniques need to be used to manage these data (including the necessary quality checks) and staff skills have to be upgraded accordingly.

This can prove very demanding, particularly since data quality management in micro data is a very cumbersome process. Let me be clear here: micro data per se do not ensure high-quality statistics. But it is true that if you have micro data you can check the quality of the data, pinpoint where the problems are and potentially revise them. In the case of aggregate data, you may not have the adequate means at hand to check their quality and correct them.

An important issue for ensuring data quality is the proper identification of economic agents. In a complex and globalised world (and in particular in a monetary union where domestic transactions correspond to an area larger than your own country), the identification of counterparts to maintain high-quality statistics is simply essential. Since the European National Central Banks can only properly pursue their functions by previously identifying the financial corporations under their scrutiny and the non-financial counterparts of financial

market transactions, they have created their own register: the Register of Institutions and Affiliates Database, (RIAD).

A key step forward to ensure there is adequate identification of agents at the European level would be to enhance cooperation in this field between the European Statistical System and the Eurosystem. This could be done, for example, by integrating their reference data directories (such as the ESCB's RIAD and the ESS's Euro Group Register). That would afford us better knowledge of the relations between subsidiaries and their parent companies, which are so important for understanding the effects of globalisation.

Moreover, the ESCB has contributed to the creation of a worldwide network to set up a global identifier for legal entities, the Legal Entity Identifier (or LEI). In Spain, some institutions, including INE and the Banco de España, are promoting the use of the LEI and facilitating its implementation in our country. I think there is a broad consensus that the widespread use of the LEI among institutions and firms globally would greatly favour the monitoring of financial stability and the compilation of statistics.

Globalisation also poses significant challenges for the compilation of macroeconomic statistics, on which our institutions cooperate very closely. Multinational enterprises (MNEs) organise their activity and finances globally and the complexity of allocating them correctly across countries can only be afforded by data-sharing –nationally and internationally- and perhaps even by resorting to centralised data reporting and quality checking. Obtaining information from MNEs as groups would allow reporting to be simplified, while at the same time ensuring a consistent cross-country recording of transactions, assets and liabilities.

I have mentioned before how important it is for official statistics to offer a complete and consistent picture of the whole economic situation. In this regard, it is crucial that users know the links between the different products, even when they are produced by different institutions. And this is something which should hold not only for aggregates but also for micro data.

Opening up our data for analysis / transparency

The current statistical revolution has gone hand in hand with substantial progress in empirical economics and modelling. Some years ago, macro literature often relied on representative agent models. That was a natural consequence of computational limitations. However, the aggregation of individual behaviour does not in general coincide with the behaviour of one average individual. That's why this literature is now moving towards models with heterogeneous agents (each with distinct behaviour and a distinct reaction to policies), which allows not only for a more accurate representation of the economy, but also for the analysis of distributional issues, such as inequality.

For such sophisticated analysis, detailed micro information is key. This is also true for empirical economics, and in particular for the evaluation of economic policies. Many policies target very specific groups of agents according to some individual- or location-specific characteristics. Moreover, one must take into account that the effects of policies depend on many individual features as well as on other contextual issues. Those two factors notably increase the heterogeneity of the effects of such measures. As a consequence, access to

massive data that do not limit the scope of the analysis on account of small sample biases is key for the evaluation of economic policies.

As I have stressed many times recently, and in particular over the last few months during the COVID crisis, the key to understanding the degree of success of policies or the occurrence of potential unexpected consequences is to proceed with proper and timely evaluations based on sufficient background information. This is the only way to ascertain to what extent policies are working or need adjustment in terms of their timing, magnitude and focus. Global authorities are now in uncharted waters in terms of the dimension of the crisis and its multiple ramifications, and it is very difficult to calibrate the response of economic policy. Only good and timely data can allow us to design an optimal policy reaction in such a complex context.

The present crisis has also stressed the importance of having data in real time: in the last few months, information on the use of payment cards, traffic on roads and at airports, electricity consumption, mobility, Google searches and so on has been fundamental for assessing the magnitude of the economic slump and the subsequent timid recovery. These indicators will hardly join the compact and methodologically sound realm of official statistics; but I think we should also pay attention and invest some resources in understanding them, familiarising ourselves with their use and incorporating them into our forecasting models. Probably, some of these indicators could be temporarily considered “experimental statistics”, a label statistical institutions may use to explore new fields and to convey to users the as yet not fully sound nature of these data. At the same time, I think we should improve the timeliness of official statistics as much as possible. In this connection, the recent shorter time lag in the production of INE’s CPI and GDP flash estimates are good examples.

In the case of the Banco de España, we have also been able to obtain some real-time granular information that is very useful for assessing the current difficult conditions. As regards information from banks, the data provided by the Central Credit Register is now more necessary than ever for assessing the risks of granting credit and the level of aggregated risk. For us and other economic authorities, this information has also helped in the design and assessment of the effectiveness of the measures adopted during this period (moratoria on loans, public guarantees on lending, etc.). Last April we also made use of the Central Balance Sheet Data Office database to conduct a quick survey of a sample of non-financial companies in order to capture their initial reactions to the lockdown. This information proved very helpful in guiding our assessment of the initial impact of the pandemic on Spanish firms.

And, of course, the challenge posed by the implications of climate change for the economy and the financial system also requires that new databases be built and that the key signals to keep the authorities’ policies focussed be extracted from those we already have.

In all these tasks I believe that the role of external analysts and researchers can be pivotal. In this regard, I see only advantages from making information produced by authorities public to the scientific community. Evaluation of public policies should not be the sole preserve of analysts from policymaking institutions. Data availability paves the way for evaluations by independent researchers with a clear potential for enriching our knowledge on the impact

of public policies. This in turn increases the number of cross-checks to ensure more robust results by looking at the same question from different angles. My impression is that the more ambitious we are in this field, the better the assessment by researchers we are going to have. Indeed, it is a fact that countries with more open data typically enjoy larger than average high-quality economic evaluations.

That's why we have created a data laboratory at the Banco de España (BELab) and intend to gradually open up our data to the scientific community. We have started offering the annual financial statements of about one million companies per year, made available thanks to the agreement signed with the Mercantile Register Association (Colegio de Registradores Mercantiles). As part of our medium-term strategic plan, we are currently analysing the legal and technical conditions to broaden the databases disseminated at this laboratory and to ease access to the data. Currently, it is only possible to access the data lab using the "data-room" located on our premises. But we are preparing the tools to make such access also possible by a remote and secure means.

Naturally, a dilemma arises between the advantages of opening up our micro data to society (in particular for scientific purposes and for improving the design of policies) and the statistical confidentiality that reporting agents wish to keep. I think this is something that in the past has been biased on the side of excessively restrictive access to data. Given modern anonymisation techniques, my impression is that the picture should now become more balanced in favour of opening up our institutions and data to society.

INE's own laboratory is a good example of this movement. And I can in fact affirm that some Banco de España analysts are already enjoying access to INE's data lab. Their experience shows the potential advantages and benefits of cooperation between institutions and researchers.

Once public institutions provide the appropriate technology ensuring data confidentiality, analysts can undertake relevant policy evaluations and analyses that until now were not available for policymakers. This opens the way towards improving the policymaking process by conferring a greater weight on evidence and evaluation techniques. In this connection, the role of policymakers is to decide on public policy, but this role can be conducted more effectively the greater the information they have on economic conditions and on the consequences of each choice made. Useful evidence-based inputs to policymaking crucially require good and accessible datasets.

Conceivably, we could think about a further step forward, linking and federating our data labs. Or we could think about the possibility of establishing "safe hubs" of granular information where data from different institutions are shared with every guarantee of retaining statistical secrecy and confidentiality. By way of example, in some Nordic countries there is a consortium of public institutions which share their data not only with their own analysts but also with the scientific community.

It would also be worth strengthening the possibilities of having access to administrative registers of other institutions and to allow the matching of different databases. The value of micro datasets increases exponentially when they can be combined, including the granular information of one dataset in the observations of another. The amount of research

possibilities here is really huge. Therefore, we should make every effort to achieve these types of dataset combinations, through the above-mentioned “safe hubs”, through a consortium or by any other means.

Another related issue is the importance of providing all the observations making up a given register. A large amount of observations is by itself important, as it allows the study of specific groups of agents with sufficient accuracy. But it is also crucial when it comes to merging different datasets. If an administrative register is reduced to some sub-set of observations, then the chances of finding the same agents in a different dataset decrease dramatically. This is of particular importance in the case of surveys containing important information to be merged with some registers. Surveys usually have a much smaller sample size. Therefore, the number of observations in common with an artificially reduced register could be very low or even zero.

The path to striking a balance between giving access to the research community to the more interesting databases while ensuring confidentiality and anonymisation could much benefit from the experience of other institutions. To this end, a few years ago several central banks set in motion the so-called INEXDA network. At present this network comprises central banks, including the participation of the ECB Statistical Directorate, and other statistical institutes, like Eurostat. In this forum participants from the statistical areas and datalabs exchange experiences in the management of micro data, the aim being to guarantee the harmonisation of certain procedures and share best practices in this domain.

Access to the information in the hands of private sources is also a very interesting field where we can make further progress. INE had a very good experience this year in exploiting data from mobile phones to analyse citizens’ mobility in different cities and regions, easing and lessening the cost of developing such statistics, with the help of new technologies. The need to better understand and organise the transition from customer data, privately used by service corporations, to citizens’ micro data, of interest for public analysis, has to be studied thoroughly. That is why I welcome and congratulate INE on recently launching a Business to Government Working Group (B2G WG), under the name of “WG about the role of official statistics on data stewardship”. In this WG, key players will reflect on ways to provide access to private data while maintaining and promoting ethical principles.

In fact, I consider there is a need to break the quasi-monopoly which some big tech-companies have on personal data, something which requires a co-ordinated international effort. The European data strategy announced by the European Commission a few months ago, aimed at public/private-sector collaboration in data-sharing, is a good example in this respect.

It is also important to recall the usefulness of statistical information for public opinion. Central banks’ information is quite often of a very technical nature. But, at the same time, some variables –such as interest rates- have a strong impact on household finances and SMEs, whereas information about recent developments on key economic and financial variables are also very relevant for public opinion. Therefore, accessible (transparent, clear) statistics are also needed.

In our case, we are continuously striving to improve how statistics are conveyed to society and to adapt our communication to different types of users. A couple of years ago we conducted a survey of the users of our statistical information. The results offered a useful guidance about some key issues. Following the input from this survey, we have in recent years tried to exploit new technologies to display data, including more charts, interactive menus, dashboards, etc. And just a few days ago we launched an application which will allow users to have our statistical information readily available on their mobile devices.

The quest for efficiency in statistical reporting and compilation

As we have seen, data - whether micro data or aggregate variables - are an essential ingredient for analysing and designing economic policies. But at the same time, reporting data are not free. This always means a burden on reporting agents. And I think that the responsibility of all of us as authorities is to minimise that burden and establish the mechanisms which ensure data collection is as efficient as possible and smooth the ensuing interactions with agents to check data quality.

In a sense, collecting micro data may help reduce the reporting burden on economic agents (such as banks, investment funds, non-financial companies, etc.) if they are able to replace other aggregate requirements imposed by the authorities. Moreover, these requirements may sometimes change over time, thus demanding a continuous and costly adaptation of the agents' reporting schemes, including complex IT developments.

But, in any case, data collection imposes short-term costs on agents. Public institutions should be sensitive to their demands to establish an efficient system to collect information, avoiding overlaps and harnessing synergies.

In order to tackle this issue there are already some data-sharing mechanisms in place across statistical institutions. Indeed, the special relationship between INE and the Banco de España in the national accounts and balance of payments fields is a good example. Yet more can probably be done both at the national and international level.

On the international front there are good practices we can explore. A case in point is that in Portugal, where non-financial companies have a sort of single window to report the data required by the main Portuguese authorities (Tax Agency, National Statistical Institute, Mercantile Register, Central Bank). In particular, our neighbour has been able to build a single data platform to which every authority can subsequently resort to obtain the information they are entitled to access. I have seen that in the Spanish National Statistical Plan (PEN) for next year there is a reference to a project setting up a "Sistema Integrado de Información de las Empresas", which I think is along the same lines. It could then contribute to significantly reducing firms' reporting burden in the future.

As concerns financial information, we should also be open to more efficient financial reporting and statistical production at the Banco de España and the ESCB level. The Banco de España is very much committed to this goal. In fact, we already have a strategy to simplify and rationalise regulatory reporting that is starting to bear results. The main aim is to alleviate the reporting burden and to improve data management by institutions and by the Banco de España itself.

On top of that, various initiatives have been taken as part of the Banco de España's Strategic Plan (which runs to 2024) dealing with data management and transparency. The goal here is to have a better knowledge of the overall datasets available at the institution (which may point to some synergies, overlapping and gaps), and to establish good data governance procedures and a technically sound platform to host and exploit all that information.

At the European level, several initiatives are in motion as well. First, the European Commission has asked the European Banking Authority (EBA) to conduct a feasibility study to explore the possibilities for statistical, supervisory and resolution reporting to converge, thus reducing the reporting costs of financial institutions. The ESCB proposal to the EBA feasibility report –published last September- is based on three main avenues of work: a) a common standard data dictionary; b) the harmonisation of reporting formats, the removal of duplications and improved data-sharing; and c) increased cooperation between authorities and reporting agents.

Linked to this project, the ESCB is developing an Integrated Reporting Framework whose aim is to reduce the reporting burden for banks and to ensure the effort made by any credit institution does not depend on the particular euro area country in which a bank is resident. To this end, the purpose is to integrate the existing ESCB statistical data requirements for banks into a single, standardised reporting framework, making it homogeneous across the EU.

There are also initiatives to explore the role played by national central banks as producers of statistics at the euro area level. For many statistical products and for different stages of the statistical process, central banks mainly rely on their own capabilities, running different IT systems and developing autonomous projects in each country. Here, some synergies can probably be harnessed and costs reduced if collaboration between central banks is further strengthened in that statistical process.

Concluding remarks

Allow me to conclude. Modern technologies offer a wide range of possibilities for obtaining and assessing information at much lower costs than in the past. Here we have a technological challenge, but another we face is to recruit our staff and to provide them with the skills needed to address these new issues. This may be a long process but we need to tackle it now. As I have tried to explain today, the potential benefits for economic policymaking are very significant.

At the same time, there are other obstacles in the way of improving our knowledge of the rapidly changing world. These have more to do with institutional factors and even some cultural traits. They are not easy to remove, either; but we cannot deny we have a major role to play in gradually changing our mind-set and overcoming the barriers which today hamper, for example, data-sharing among institutions and open data for scientific purposes. My impression is that we need to be very ambitious if we do not want to miss that train. We also have more possibilities of succeeding if we join forces with other institutions, both at the national and the international level.