

Micro foundations of macro data: increasing data quality and exploiting the “true wealth” of National statistical institutes

Enrico Giovannini

President of the Italian National Institute of Statistics (Istat)

PRELIMINARY VERSION

Paper prepared for the Sixth ECB Statistics Conference

*“Central bank statistics as a servant of two separate mandates:
price stability and mitigation of systemic risk”*

17-18 April 2012

1. Introduction

For decades, the main outputs of National statistical offices (NSIs) have been macrodata. Economic and social indicators, national accounts data, census data, etc. have been disseminated in statistical tables, disaggregated by sector of economic activity, sex, territory, etc. Users have used them both to carry out analyses using different socio-economic conceptual frameworks, as well as statistical and econometric tools, and to evaluate policies and make forecasts.

Of course, NSIs calculated macrodata starting from microdata collected through statistical surveys or derived from administrative processes. For example, national accounts are based on data sets coming from a very wide range of sources: the dream of every national accountant is to produce final data just summing up microdata (and from a conceptual point of view this is what the accounts should do), but in practice they have to use imputation and other statistical methods to transform the raw data into a coherent set of macroeconomic aggregates.

Maybe, this is why for a long time microdata were just considered as raw materials useful to produce the final results, more than the “true wealth” of NSIs. Until not many years ago, when computers were not readily available or powerful enough, managing microdata was complicated for statisticians, and especially for national accountants. In a similar fashion, the attention of

analysts was focused on macrodata, since statistical and econometric tools for exploiting microdata were not well developed. Finally, policy makers were mainly interested in macroeconomic policies and tools, so that there was less interest in pushing NSIs to make microdata available.

Over the last two decades the world has dramatically changed also from this point of view. The power of computers and of statistical and econometric tools make the management of huge data sets very easy. Micro-datasets are daily generated by all kinds of economic and social activities and can be exploited for commercial and analytical purposes. We live in the “information society” and the opening of new frontiers are daily announced by experts of data generation and management. In this new perspective, NSIs (and National central banks – NCBs) have seen that the demand for microdata is growing very quickly and have tried to develop new ways of exploiting and disseminating them to users.

In this paper we will discuss some new frontiers in the use of microdata to generate macro economic and social statistical information, as well as the challenges that NSIs are facing in this area. In particular, the second section is devoted to national accounts and some new ways of using microdata in this context, while the third section deals with similar opportunities in the field of short-term economic statistics. The third and the fourth sections deal with the use of microdata in structural business statistics and in the development of statistical registers. The fifth section discusses some opportunities and risks in the use of microdata for statistical and research purposes, especially in the context of the European Statistical System (ESS). Some conclusions are then drawn in the final section.

2. National accounts

As already mentioned, national accounts (NA) departments within NSIs have the mission of estimating data which go far beyond the information contained in the collections of their micro-level data ancestors. However, rooting the production process in microdata themselves provide significant enhancements in the quality and the wealth of information produced. A good example in this area is the process of “data integration” that Istat has developed almost 25 years ago for the production of Economic Accounts, especially to produce “exhaustive” estimates of GDP taking into account the large share of underground economy existing in Italy. The “Italian approach” to

estimate national accounts, based on the integration of different sources at micro-level, has become a recommended practice by Eurostat and OECD¹.

During the general revision of national accounts carried out in the 1980s, Istat moved from estimation models of value added by sector, based on the macro-results from primary sources (at that time, mostly censuses and Structural Business Statistics - SBS surveys), to a methodology based on micro-level information and on a wide range of sources. The analysis of microdata became fundamental, especially for small enterprises, to estimate underreporting and overcome inconsistencies between different sources (for example concerning employment figures), which were used to estimate irregular working positions.

More recently, with the development of European regulations on structural business statistics (SBS), international trade, etc. direct use of microdata is necessary because estimation domains used by national accounts are different from those established by legislation concerning SBS. This in turn implies using specific data control and correction methods, due to the difference that also exists in the concept of quality between SBS surveys and National Accounts estimates: the process used to improve the quality of the NA estimates from survey data complements that of the survey itself². For example, error and coherence checks are performed in NA, with corrections for larger enterprises and re-estimation based on integration with other sources for SMEs. This process results in:

- higher accuracy in economic accounts, getting to finer estimation domains with respect to the original ones;
- quality improvements for both NA and revised SBS data, as a feedback is provided to the department in charge of SBS estimates;
- enhancement of the information produced, with estimates on the irregular component within employment and of the contribution from the Non Observed Economy to value added.

Of course, maximizing the benefits of this process requires a strict interaction between statisticians in charge of NA and SBS statistics. Also, improving the accessibility and quality for statistical purposes of additional sources – in particular administrative records – is of paramount importance.

¹ See OECD (2002) “Measuring the Non-Observed Economy: A Handbook”, available at <http://www.oecd.org/dataoecd/9/20/1963116.pdf>.

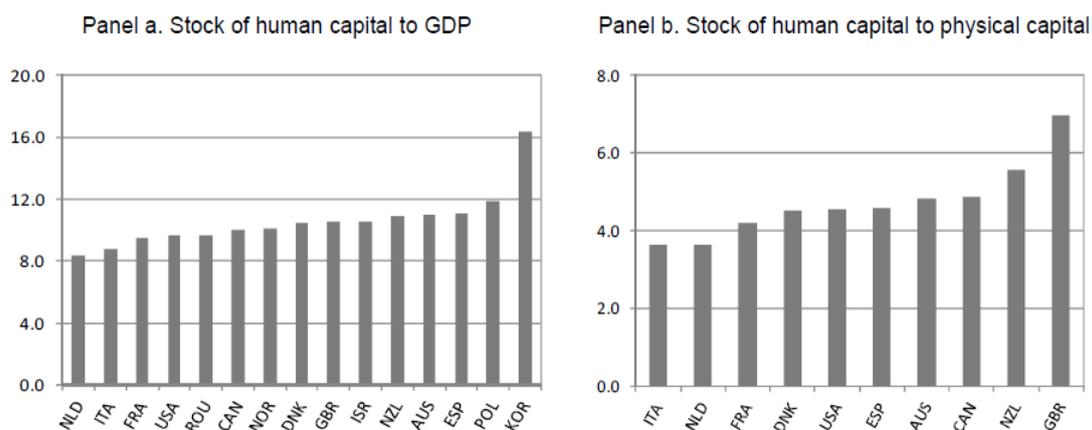
² Of course, the two aspects tend, though partially, to overlap each other (see, *inter alia*, Puggioni, 2000). Controls on micro data are both preliminary and integrated in the estimation of NA aggregates and in the analysis of their accuracy as, generally, they provide good indications on the entity of errors that influence the estimates. In particular, the integration of the sources by the NA allows identifying possible errors in micro data, in addition to evaluating the components of non-sampling error (coverage errors, measurement errors, non-response errors).

Few more recent examples of the power of integration and ongoing enhancements in this field can be mentioned. In the area of Household accounts, after the publication of the Report of the “Stiglitz Commission”³, Eurostat and the OECD have recently set up an Expert Group on *Disparities in a National Accounts framework*, to produce estimates of disparities between groups of households in the framework of National Accounts, based on existing micro data. This work is expected to bring to the publication of household accounts split by main source of income (into 5 subgroups), by adjusted income quintiles (5 subgroups) and by type according to household size, the presence of children and the age of adult members (7 subgroups).

Similar but more ambitious is the work for the calculation of *Social Accounting Matrices*, based on the full use of National Account Matrices (NAMs). To this respect, a pilot-experience was carried out in Europe at the beginning of the last decade, but remained isolated. Only the Netherlands is currently publishing SAM on a regular basis⁴.

Another experience, carried out by several NSIs under the auspices of the OECD⁵, relates to the income-based measurement of the stock of human capital. Individual records on education, gender and age were integrated to achieve estimations of human capital remuneration trajectories at the macro level. An estimate of the stock of human capital in 2006 was produced, using the lifetime labour income approach developed by Jorgenson and Fraumeni, by age, gender and educational level, computing ratios to GDP and physical capital (see Figure 1).

Figure 1 – Micro-level based estimates of human capital stock ratios to GDP and physical capital



Source: Liu (2011)

³ See http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf

⁴ See <http://www.cbs.nl/NR/rdonlyres/F5AA0D6D-1257-48B1-9B07-72D6979FAB68/0/2011socialaccountingmatricesandlabouraccounts.pdf>

⁵ See <http://www.iariw.org/papers/2010/5Mira.pdf>

A significant enhancement of micro-level estimates is now underway at Istat, by using more detailed data on educational attainments and further integrating individual records with micro-level information on sectors of activity. This would overcome some of the shortcomings of the Jorgenson-Fraumeni approach and would lead to the production of satellite accounts on human capital within the national accounting system, comparable with other aggregates as well as at the international level.

3. Short-term economic statistics

Integration of microdata is also increasingly used in other fields of economic statistics. This activity aims at adding value to the information produced, but also reflects the changing priorities of users, both at policy and research levels. More and more policy makers require information about how policies or other economic and social changes influence the behaviour of enterprises and consumers and to do that it is necessary to go beyond macro-aggregates, investigating micro data-sets. On the other hand, analysts are not satisfied anymore with the cross-tabulations performed by NSIs, but want to explore new ways of aggregating microdata, following experimental classifications not yet included in the standard classifications used by NSIs.

The use of record-linkage to produce integrated sets of annual indicators is quite common in most NSIs and international organizations: for example, more than 15 years ago Istat published data about exports disaggregated by size and other characteristics of enterprises, produced linking international trade data and the business register data, an approach that has subsequently become a recommended practice at international level⁶. Several activities could be mentioned here: some of them require the availability of a fully-fledged business register, others can be carried out without it. In this section two recent examples developed at Istat will be mentioned, just to illustrate the opportunities available to NSIs to satisfy new users' requirements in the area of short-term economic analysis. The first one does not require a business register, while for the second one such a register is needed.

The first examples are related to the analysis of inflation, using data concerning the Consumer price index (CPI) to assess the impact of an exogenous shock, such as the recent increase of one percentage point in the standard VAT that took place in Italy in September 2011, affecting about half the value of goods and services included in the CPI basket. For this purpose, Istat carried out a statistical analysis of frequency and intensity of changes concerning about 240.000 price observations per month, disentangling the effects of seasonality and other exogenous factors at the

⁶ See <http://www.oecd.org/dataoecd/19/14/47946268.pdf>

product level. This exercise allowed to ascertain that the translation to consumers of the VAT rise from 20% to 21% was only partial, while a macro based analysis would have led to the opposite result the monthly increase in CPI was higher than the one due to VAT – or would have been inconclusive. Besides, the pass on of the VAT rise was diluted along time and not uniform across products.

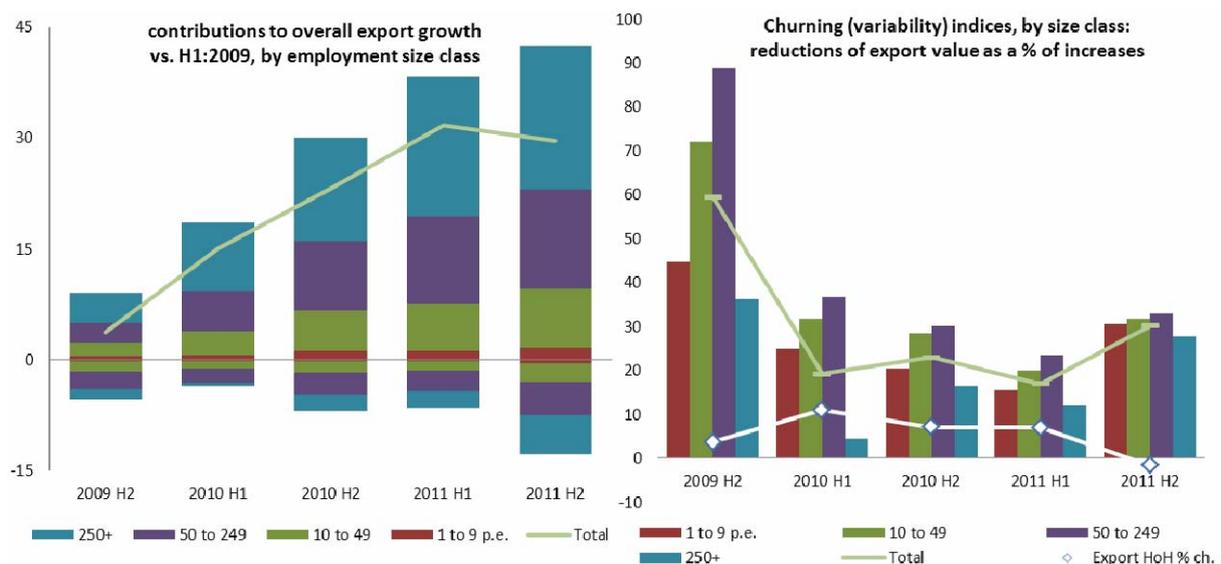
Indeed, the analysis showed that the VAT increase had pushed up, for the products involved, the frequency of the elementary quotes with a positive change with respect to the previous month, from 5.2% in October 2010 to 17.7% in October 2011, then this went down again to 6.7% in the following month (in November 2010 it was 3.8%), highlighting a reduction of the tensions due to the VAT increase, and its only partial pass on prices. In both months, the distribution of price rises was concentrated around the amount corresponding to the VAT shock (i.e. +0.8 percentage points), with increases affecting in the first place dearer and subsequently lower priced products, while the price of cheapest products (where only rounding effects could take place) in most cases did not increase during the observation period. A similar analysis, addressing price changes in the cereals supply chain in response to raw materials price shocks and volatility, confirmed that the behavior of downstream agents (producers and retailers of bread and pasta products) was characterized by a relevant degree of inertia, joint to the extensive use of discounting policies for packaged items. These examples suggest how analyses based on micro-level observations on prices can be useful for policy purposes, as they can help to understand and model behavioural patterns across sectors, along the supply chain and at different points in the business cycle. Moreover, the use of microdata to perform non-standard calculations is particularly important to overcome the criticisms that NSIs have faced about the reliability of CPIs after the euro changeover. At that time in several European countries the public opinion strongly criticized the official inflation data, comparing them with the so-called “perceived inflation”. Actually, as several analyses have demonstrated, perceptions are closely linked to changes in prices of items most frequently purchased: this is why the Italian statistical office includes in its official monthly press release not only data for the most common breakdown of the CPI, but also those concerning items purchased at different frequency.

More generally, this experience shows the need to use microdata to describe the variability of economic phenomena, going beyond what averages show. This capacity is vital for NSIs, as they operate in a world in which “special cases” can be presented by media (especially new media) as “general cases”, thus influencing public opinion well beyond the reality described by statistics. Of course, this approach to news makes the life of NSIs very difficult, because when there is a discrepancy between perceptions and statistical reality people may tend to trust the former, and not the latter.

The second example is the extension to infra-annual data of the data linkage between export and business register data. The analytical question that led Istat to perform such extension was concerning the capacity of Italian manufacturing enterprises to respond to the recovery of international trade after the 2008-2009 crisis, when exports fell down in a dramatic way. To this end, it was performed an analysis on the panel of manufacturing enterprises exporting throughout the period 2009 to 2011, and consisting of 43.000 units with 2,2 million persons employed (nearly 50% of total employment in manufacturing). This panel provides an extremely effective representativeness, in as it covers about 96% of the value of exports from manufacturing firms.

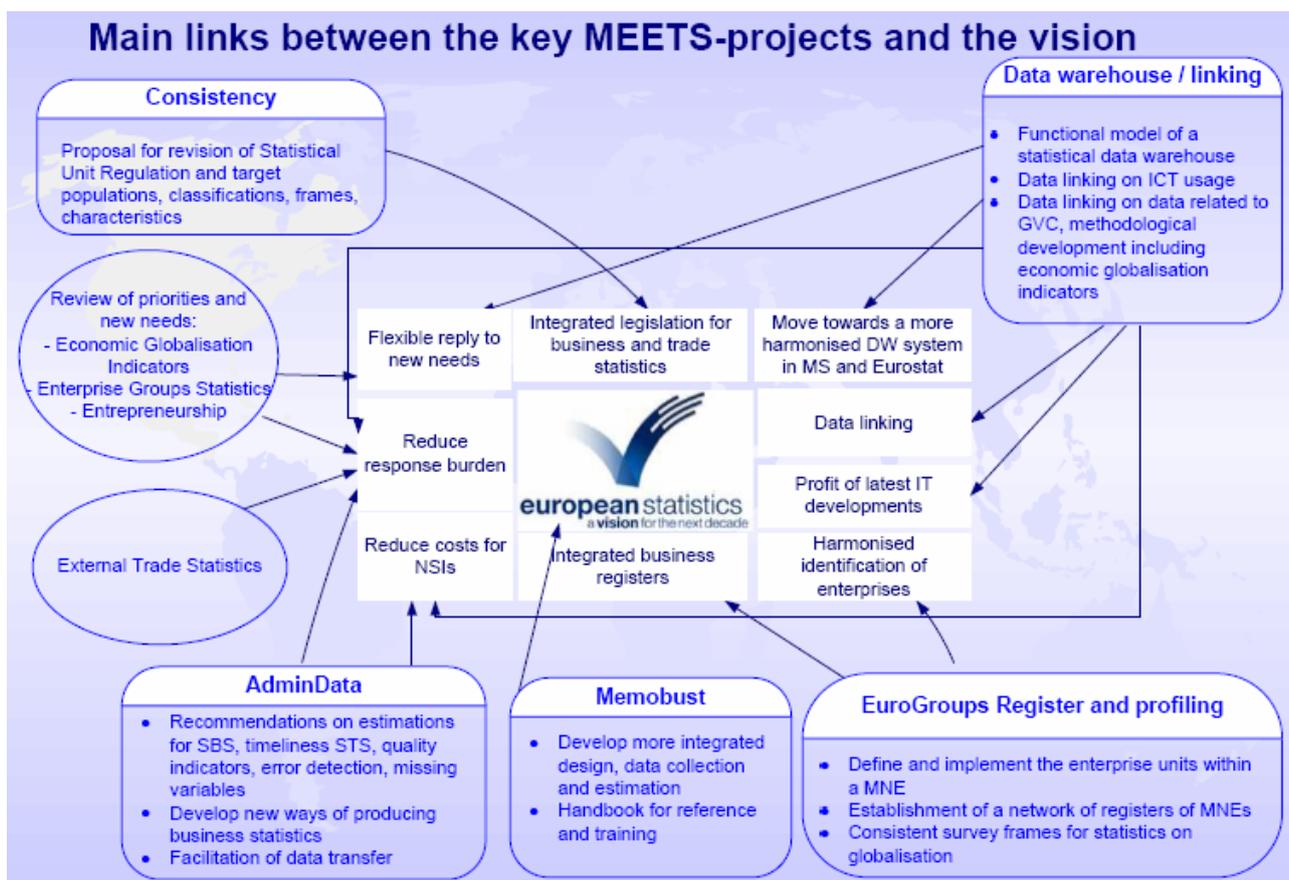
This analysis shows that in the second half of 2011, about 66% of firms in the panel had increased export levels with respect to the first half of 2009 (the lowest level), with larger firms heading with respect to smaller ones (72% vs. 64% at the two ends of the employment size distribution). The overall increase of 29.7% in export value reflected a contribution of 42.5 percentage points from enterprises with rising exports and a drainage of 12.8 points from enterprises with shrinking exports, with a different pattern along time leading to “net” export increase across size classes (see Figure 1 below). This type of analysis can be – and indeed was – performed also to a much finer level, considering patterns by products and markets, together with firms’ characteristics, in order to monitor business cycle changes on business on an extremely timely basis.

Figure 2 – Export growth and contribution by employment size class of enterprises



4. Structural economic analysis

Recent years have witnessed several advances in quality, availability and usage of microdata for structural economic analysis in the European Statistical System, and other are in progress. Some of them stem from the spread of best practices developed by individual NSIs, others (actually, the majority) are the result of joint efforts within the ESS, to which contributed both the EU 7FP in R&D and Eurostat Programme for the Modernisation of European Enterprise and Trade Statistic – MEETS (See Chart below), which also promotes few collaborative Networks (ESSNET) on this topic.



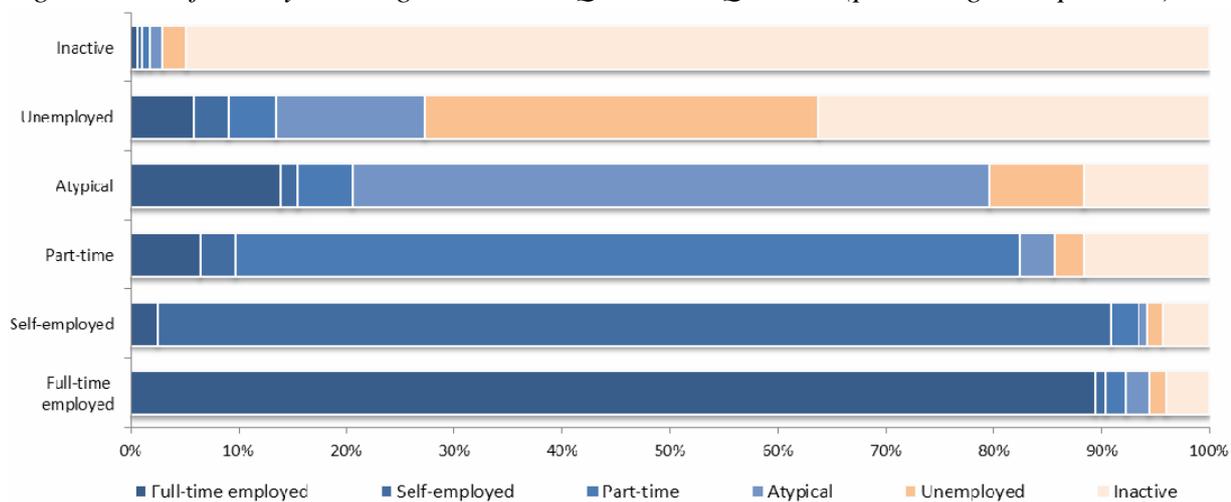
Source: Eurostat

One large collaborative project involving NSIs and Research centres is the BLU-ETS project, which addresses research and dissemination of improvements on business data concerning collection, accessibility and re-use across the EU. It also explores and promotes ways to reduce response burden on enterprises (as a way to increase response rates and quality), to enhance the usability of administrative data for statistical purposes, and to develop methodological improvements to increase the added value of existing data – such as small-area estimation in sample surveys – and disclosure techniques, including remote access.

Another example in the ESSNET strand is the ESSLIMIT project, participated by 15 European NSIs, which gathers information from different sources (ICT, CIS, SBS, foreign trade surveys, balance sheets, Business register, employer-employee register where available), with the aim of exploring the relationships between ICT usage, innovation patterns and enterprise dynamics, and producing multi-source indicators for Eurostat on these issues⁷. Aggregated microdata are shared within the consortium for common research, and it is envisaged to set up of a repository for micro-aggregated information open to external researchers.

Changes in collection techniques, together with improvements in archives, with a more extensive use of administrative sources, increasingly allow for longitudinal analyses. In several EU countries, already now Labor Force Surveys permit tracking patterns in labor market dynamics at the micro-level, by confronting the occupational condition of single units over time. For the case of Italy, the Labor Force Survey includes a longitudinal structure based on a rotation system of households within quarterly samples. Half of the households included in a sample is interviewed again after three and twelve months, and a quarter of them after fifteen months. This makes it possible, therefore, to build transition matrices and to implicitly estimate the probability of future occupational status given the initial condition. As an example, 92% (Figure 3) of people employed

Figure 3. Exit flows by working condition – Q1:2009 – Q1:2010 (percentage composition)



Source: Istat

in Italy in the first quarter of 2009 (near the minimum of the business cycle) were still employed one year later, but this result varies from 94.4% for permanent full-time employed workers to 79.6% for workers with atypical contracts. Little more than one fourth of the unemployed (27.3%)

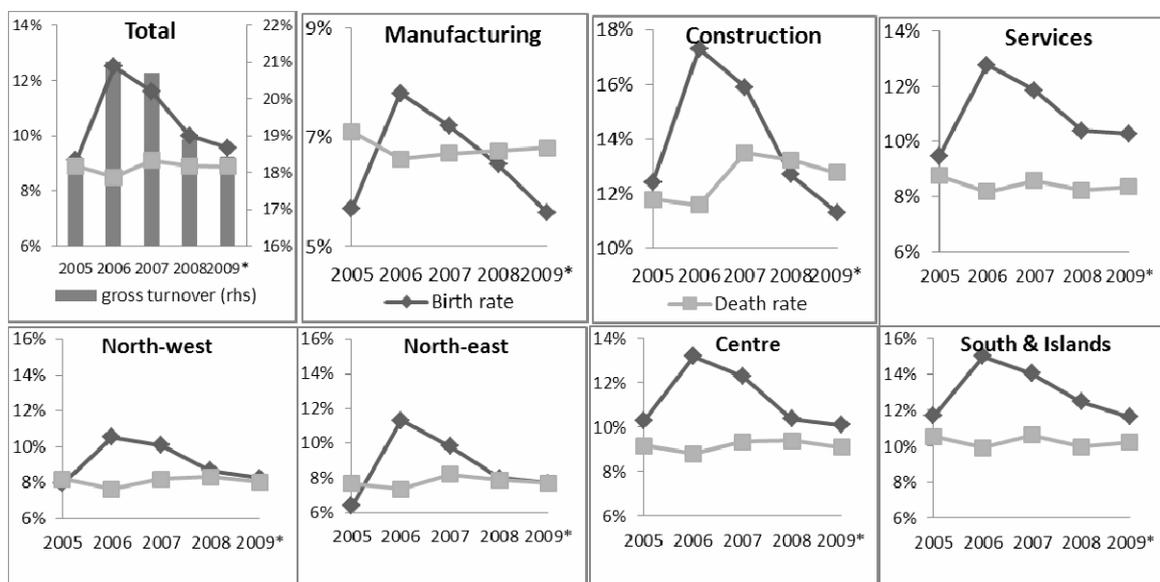
⁷ See <http://www.esslimit.eu>

had managed to find a job, and in half of the cases this was with an atypical contract. After one year, 36,5% of the unemployed were still looking for a job, while the remaining 36.2% had slipped into inactivity.

5. Statistical business registers

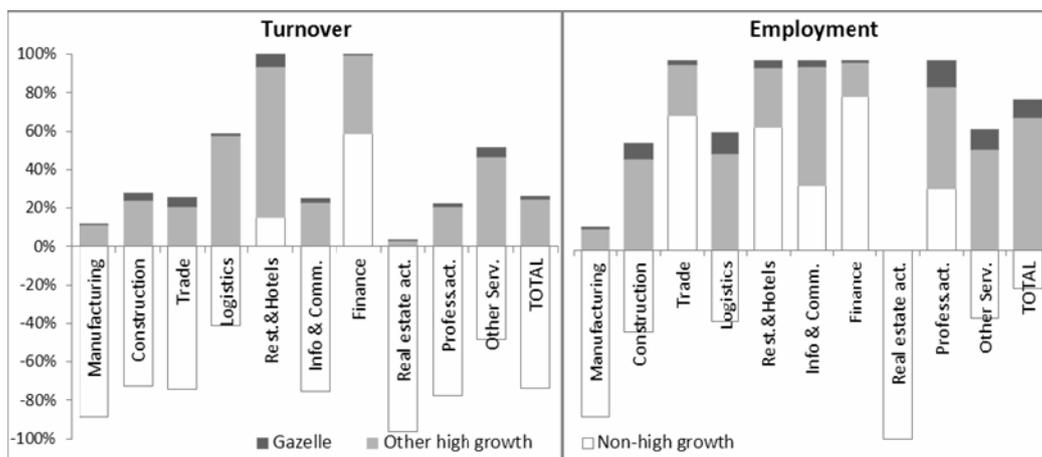
Business registers form the backbone of Business Statistics- In the last decade, registers underwent major improvements in most EU countries, increasing coverage and reliability of information, in particular for the statistical treatment of administrative events. In many EU countries, underlying microdata can thus be used for mapping the enterprise system structure, and to portray business demographic and entrepreneurship features and dynamics, at any level of detail. As an example, for the case of Italy, the Active Enterprises Statistical Archive - ASIA allows to ascertain that the crisis impacted especially on enterprises birth rates, which shrunk from 12.5% in 2006 to 9.6% in 2009, while death rates stayed roughly stable just below 9%. In 2009 alone, when the crisis hit the hardest, net enterprise creation in the economy as a whole stayed positive, with a growth of 0.7% in the number of firms with employees, although gross birth rates in manufacturing and construction shrunk of 0.8 and 1.4 percentage points, respectively, with a net contraction in the number of active firms. Differences in demographic patterns could also be found geographically: indeed, the 2009 increase in the number of active firms stemmed from central and southern regions, which structurally show higher than average net birth and turnover rates (see below, Figure 4).

Figure 4 – Birth and death rates of Italian enterprises with employees by sector and location of HQ in the years 2005-2009 – main sectors of activity and NUTS1 areas



Business register information allows to analyse firm demographics also with respect to other characteristics; for instance, about half of new-born firms in 2009 were established with a sole proprietorship status, and female represented about 28% of owners in this group, with a strong concentration in service activities. Enterprise creation and turnover rates (by sector and/or geographical area can be also put in relationship with survival rates (showing a mild inverse correlation when sectors are taken into account), or with size growth patterns, looking for high growth and gazelle enterprises (Figure 5)

Figure 5 – Percentage of firms with turnover and employment growth by group: high growth (including gazelles) and other active firms; years 2006 to 2009; total and main industries



A remarkable prospective improvement in the area of Business Registers is expected by the Euro Groups Register (EGR) project promoted by Eurostat, which is creating a network of national business registers focused on multinational enterprise groups. The EGR contains structural economic information on enterprises part of multinational groups with an interest in Europe, and is foreseen to become the coordination frame for all European statistical authorities, National Statistical Institutes and National Central Banks for sampling populations in the production of statistics on globalization, starting from statistics on foreign affiliates (FATS) and on foreign direct investments. It will offer statistical compilers access to integrated and up-to-date register data on enterprise groups with relevant transnational operations in at least one of the European countries.

6. Opportunities and risks for future developments in the context of the European Statistical System (ESS)

Over the last decade NSIs have made a special effort to increase the access to microdata for research purposes. For example, Istat currently delivers microdata for research purposes in the form

of anonymised standard-files and more detailed and customized “research files”. External researchers can also have access to original microdata upon inter-institutional research agreements for joint projects with Istat, but can also directly access its “Elementary data analysis laboratory” – ADELE. In this data research center (now accessible from all Istat regional offices) researchers can work on microdata in a secure environment, bringing home the output after this is checked for confidentiality by Istat personnel.

Moreover, Istat, the Bank of Italy and the Bruno Kessler Foundation are developing a project to establish a national data archive to collect and make available to researchers microdata produced by public and private institutions for research purposes. Istat and Bank of Italy already provide access to microdata for research purposes, but this is not the case for other public institutions (including universities) who receive funds from national or international research programmes. The aim of the initiative is to facilitate the access to this huge amount of data through the same technological infrastructure, via Internet or via safe centres. According to the proposed division of labour, Istat would take care of data made available by agencies participating in the National Statistical System, while the Kessler Foundation would do the same for data provided by academic and other research institutions. The Ministry for Education and Research has also been involved in the project, to include in the national legislation the obligation for institutions who receive financial support from public programmes to provide their data to the data archive.

The effort made by NSIs to make microdata available to researchers has been very much appreciated by them. A similar effort has been carried out by Eurostat⁸ and a European project is now underway to develop remote access facilities to microdatasets stored in individual NSIs⁹. It would be important to accelerate the activities of this project, in order to develop one of the first truly Europe-wide statistical infrastructures, that would represent a great achievement for the European Statistical System, avoiding duplications of efforts at the national level.

In the opposite direction, a potential threat on the future availability of micro data and, even, to their use in statistical production, is represented by the EC proposal for a General Data Protection Regulation (COM(2012) 11/4 draft). This Regulation is aimed at simplifying and unifying the current legislation setting and at the same time reinforcing protection for citizens, but unwillingly creates some serious concern for data producers – NSIs in the first place, but also Central Banks – which affect policy owners and the research community as well. Indeed, the draft proposal might

⁸ See

http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/statistical_confidentiality/confidential_data/anonymised_microdata_files

⁹ See <http://www.essnet-portal.eu/essnet-decentralised-and-remote-access-confidential-data-ess-dara>

lead to limit the access to micro-data for users and to some impacts on the usability and preservation of microdata themselves, as published statistics could be considered as “personal data” when the data controller still keeps the underlying individual records, while no attention is paid to research needs in the regulatory setting.

To avoid these threats, European NSIs deem that a “general exception” for statistical data should be included in the definition of the material scope of the Regulation (Art.2), and a new restriction be added (in Art. 21) for statistics, and statistical, historical and scientific research purposes as areas of public interest, in a similar fashion to the directive on data protection (95/46/EC) currently in force.

More generally, from an institutional point of view, there are two serious obstacles to a better exploitation of the potential of existing microdata for statistical and research purposes. The first one is due to the separation between the ESS and the European System of Central Banks (ESCB). Due to this separation, notwithstanding the Regulation (EC) 223/2009, the exchange of microdata between NSIs and NCBs can still be complicated, especially for data that are not relevant for the production of European statistics. In some countries the exchange of microdata is left to agreements established at national level, if the legislation allows for this possibility. In Italy, for example, this is the case for non-anonymous data and this limitation has a clear impact on the efficiency of statistical activities and on the quality of results. Therefore, it would be important to address this issue, eventually with the occasion of the forthcoming revision of the Regulation (EC) 223/2009, and to make more open the legislation in this field.

A second serious limitation to the improvement to European statistics is the absence of a legislation that allows the free circulation of statistical microdata between NSIs participating in the European Statistical System. This is a true nonsense in a globalised world, which blocks the possibility of improving statistics in several fields (international trade, business registers, structural business statistics, etc.). Also in this case a revision in the existing legislation is urgent and such a change would be extremely beneficial for all parties involved, and for final users.

More generally, these limitations mainly come from the weakness of the institutional framework in which European NSIs and Eurostat work. Since 1999, I have been arguing that the only way to build a strong European Statistical System is to mirror the functioning of the ESCB. Transforming the current ESS into a European System of Statistical Offices (including Eurostat) is the only way to overcome these and many other existing difficulties. This change has been made even more urgent by the strengthening of the European governance after the “Greek case”. ECB and the participants ESCB should sustain this approach as a way to improve the quality of statistical outputs and the independence of NSIs from political interferences.

7. Conclusions

In this paper we have mentioned several initiatives undertaken by NSIs and Eurostat concerning the development and the use of statistical microdata. They represent just a sample of what is possible to do, also thanks to the development of ICT tools, and what is being done in Europe and in other statistically developed countries. From this point of view, the existing instruments developed by the ESS in order to stimulate new research projects and to promote good practices have to be strengthened and improved.

Nowadays, it is possible to say that microdata represent the “true wealth” of NSIs, especially taking into account that the competition between NSIs and other statistical producers is tougher than ever. Thanks to the development of ICT and statistical methodologies, for private research institutes and marketing companies it is much easier than before to carry out specific statistical surveys, also in areas that have been covered for years by NSIs. But it is quite clear that the competition is much lower (almost close to zero) about the use of microdata. NSIs have a special power in this respect given to them by legislation. Therefore, they can strengthen their competitiveness vis-à-vis other statistical producers, exploiting and integrating microdata coming from different sources, developing longitudinal studies and analyzing economic agents’ behavior. Moreover, by integrating micro and macrodata NSIs can disaggregate key variables estimated within the framework of national accounts: for instance, data by sex, income levels, education, etc. for the case of individuals, or split the economic structure in unconventional ways, and look at the behavioral pattern of enterprises. These reconstituted data would indeed allow a better evaluation of the impact of macroeconomic policies on different population groups, especially important when trade-offs of various policy alternatives have to be quantified.