Integrated Structural Business Surveys – The Brazilian Experience[1]

Marcia Quintslr
IBGE, Rio de Janeiro, Brazil – marcia.quintslr@ibge.gov.br

Carlos Arieira
IBGE, Rio de Janeiro, Brazil – carlos.arieira@ibge.gov.br

Augusto Fadel
IBGE, Rio de Janeiro, Brazil – augusto.fadel@ibge.gov.br

Gisele Del Carpio
IBGE, Rio de Janeiro, Brazil – gisele.carpio@ibge.gov.br

Priscila Koeller*
IBGE, Rio de Janeiro, Brazil – priscila.koeller@ibge.gov.br

Abstract

The Brazilian Institute of Geography and Statistics (IBGE) in its effort to construct the Integrated Economic Statistics System (IEES) is working on the integration of IBGE's structural business annual surveys. The revision of the structural economic statistics is key to the development of the IEES and involves planning and development of harmonized questionnaires, sampling redesign and standardized processes for the production of annual economic information on the industry[2], construction, retail trade, and services sectors. Two issues should be previously defined for the standardized sampling design: the criteria for establishing the size of enterprises that will be included in the survey with certainty; and the choice of the basic statistical unit that will ensure adequate measurement of secondary activities of the enterprises. This paper will refer to issues like industry classification informed by enterprises, access to tax administrative records of the enterprises, the geographical coverage of data collection, treatment of non-response, calibration of data obtained in the field, and the dissemination model. Improvements are proposed in order to advance in survey's precision and efficiency, reduce response burden, and better resources allocation.

Keywords: harmonized questionnaire; certainty and sample strata; statistical unit.

1. Introduction

The revision of Brazilian economic statistics is designed to provide users with consistent and integrated statistics of economic phenomena and their` interrelationships. IBGE is developing this project based on the ongoing discussion and implementation of the Integrated Economic Statistics System (IEES), in line with the discussions in the United Nations Statistical Commission, which were

---

[1] The authors have organized this document on the Structural Economic Statistics Revision which is conducted under the coordination of the Business Statistics and Classifications Coordination and the participation of staff from the following IBGE areas: National Accounts, Statistics on Industry Sector, Statistics on Services and Trade, Methods and Quality, Business Register, and staff from the field work in the States of Amazonas, Ceará, Mato Grosso do Sul, Minas Gerais, Rio de Janeiro, Rio Grande do Sul and São Paulo. The Directorate of Data Processing is in charge of the computer systems that will be created to implement this project.

[2] Manufacturing and mining

materialized in the document "Guidelines on Integrated Economic Statistics" prepared by the United Nations Statistics Division (United Nations, 2013).

Some aspects have motivated the construction of IEES: meeting demands for consistent national and internationally comparable economic information; adoption of the System of National Accounts (SNA) as a central conceptual framework; improving the efficiency and harmonization of economic statistics produced from surveys or administrative records; and increasing the statistical use of administrative records (Quintslr, 2014). In addition, a main goal is to enhance institutional arrangements that will facilitate production, analysis, and use of integrated economic statistics. These arrangements will ensure interaction with other producers of economic information, with data providers, and users.

For IEES to be effective some guidelines are highlighted which are: adoption of international recommendations (both from the point of view of thematic coverage, as the methodological consistency of the production of information); integration of macroeconomic statistics; integration of concepts, classifications, systems, databases, data collection and data editing procedures; and reduction of the response burden who are asked to provide similar information to different surveys or government institutions (Erbisti & Quintslr, 2014).

IBGE decided to start the IEES project with the revision of the annual economic structural surveys, that is, the Annual Wholesale and Retail Trade Survey (PAC), Annual Industrial Survey (PIA), Annual Construction Industry Survey (PAIC), and the Annual Services Survey (PAS), which data for smaller enterprises are collected by sampling selection and with certainty for larger enterprises. These surveys are the main source to establish the frame for sample selection for the short-term monthly surveys and other surveys on selected topics.

The proposed revision of the structural business surveys is based on two guidelines for IEES implementation: a reduction of response burden, and the harmonization of methods, concepts and processes among all four above mentioned surveys. A harmonized questionnaire will be adopted for all four structural surveys, that is, basic variables will be answered regardless of enterprise size or economic activity. The harmonized questionnaire will be of two types: a long questionnaire for larger enterprises, and a short for smaller enterprises. A sample design for data collection is being investigated.

The next two sections present respectively the design of the harmonized questionnaire, and the sampling design. The following topics will be discussed: criteria for establishing the cut-off values above which enterprises will be surveyed with certainty; definition of basic statistical units which will ensure adequate measurement of secondary activities of the enterprises; use of up-to-date industry classification; access to tax administrative records of businesses; geographical survey coverage; treatment of non-response; and calibration of field data. Future steps are presented in the final section.

2. Harmonization of the Questionnaires for the Annual Business Surveys

The annual business structural surveys were developed in the mid-nineties as four independent surveys (industry, construction, retail trade, and services). Nonetheless these surveys were planned to form an interconnected system based on the same assumptions, such as: following the SNA guidelines, adoption of standardized statistical business data files (Central Business Register - CEMPRE), standard sample designs, and the use of National classification of economic activities (CNAE). At the same time computer systems were developed with the same structure for all four surveys, including the use of electronic data collection.

Harmonization of these surveys will provide not only significant gains for the SNA, with the adoption of concepts and new variables according to the SNA 2008 manual, but also for respondents, with the unification of the questionnaires, and for users of economic statistics due to the increased level of consistency of data and data analyses. Additionally, because original methods and procedures have changed over time to meet specific issues related to each of the different surveys further work on standardization is needed. Thus, harmonization will update and integrate survey procedures, optimize

the use of human and material resources in the planning, data collection, analyses and dissemination processes; and also reduce the response burden.

A first step in the revision of the structural business surveys was an assessment of the surveys in terms of the degree of harmonization of the questionnaires, compatibility of concepts, survey format, and conformity to SNA requirements. The revision of business surveys follows a close alignment with the national accounts latest requirements (SNA, 2008).

In a next step, two alternative methods of harmonization of economic structural surveys were analyzed. One proposes a single survey questionnaire applicable to all enterprises (with possible exception for financial and insurance institutions). This methodology is adopted by Australia (see ABS, 2015). The other proposes separate surveys for enterprises in the following economic sectors: industrial, construction, trade, and non-financial services; with most variables common to all of them and a few specific variables for the different economic sectors. The method used for IBGE's revision was the second. A decision was not reached yet if the sample will be a unique stratified sample or four separate samples. Staff is in favor of maintaining separate data dissemination for the four economic sectors but also for all four economic sectors as a whole.

The specificities of the economic sectors are taken into account in the creation of a harmonized questionnaire. These specificities are also taken into account in the data analysis performed by experts in each of the sectors of economic activities. Also the questions are being written in the language familiar to respondents.

The harmonized questionnaire was defined according to two models: one "long or complete" which collects a larger amount of data reported by larger enterprises; and a "short or simplified", which collects a smaller amount of data from smaller enterprises. The "simplified" questionnaire collects the same variables for all economic activities, that is, all enterprises will answer the same questionnaire.

In both models, data collected will allow publication of the following statistics: number of enterprises; gross value added; number of persons employed; total revenue; revenues from each economic activity; cost of employed persons, of materials, of products, and of services; expenditure on the intermediate consumption of goods and services; use of water and sewage; of fuels and electricity; acquisition of assets; and others.

The long ("complete") questionnaire includes common variables for all sectors: industry, construction, wholesale and retail trade, and services. However for some questions there was a need for customization due to specificities of the economic activities and also in order to maintain long-term time series. The variables investigated in the complete questionnaire are listed in the IBGE Discussion Paper (IBGE, 2014). The questionnaire contains of a section with business identification variables and other sections with information related to employed persons, expenditures with personnel, revenues, costs and expenditures, inventory, taxes, assets acquisition, and others.

As mentioned, the short questionnaire will be the same for all economic activities[3]. The advantage of this approach is that if the respondent does not choose the correct economic classification of the business activity when answering the questionnaire, data will still be utilized instead of the need to ask the respondent to answer a different questionnaire (nowadays it is very common for small businesses). The harmonized questionnaire will solve this problem and only correction in the classification will be needed during the data editing process (and adjustments to the sample weights). In addition, in the harmonized questionnaire the number of questions drops considerably depending on the tax status of the company.

The decision to use this unified short questionnaire (for smaller businesses) with no customization for the economic sectors (as in the full questionnaire), demanded caution because for some variables definitions conflicted between economic sectors. This is being solved by including detailed

---

[3] In order to avoid different questions for the economic activities studies were performed on each variable of interest in each section of the questionnaire. Therefore variables with low frequency of response or with small relative value were not included.

instructions in the questionnaire. The proposed version is fully unified, with the exception of the addition of two extra questions for the construction activity. After this version is discussed within different areas of IBGE it will also be discussed with users of economic data.

The new questionnaires will only have an electronic version (not paper) and will be submitted to testing and validation by internal specialized teams in economic data collection with the experience on respondents´ ability to understand the meaning of the questions.

3. Sample design for implementation of the harmonized questionnaire - exploratory studies

The development of the sample design for implementation of the harmonized questionnaire requires the definition of two strata. The first one (for the complete questionnaire) should comprise of large enterprises which are selected with certainty ("certainty stratum") after the definition of a cut-off value based on some measure of enterprise´s size. The other stratum (for the simplified questionnaire) should be designed for smaller enterprises which are selected using a random sample ("sample stratum"). The criteria and variables for definition of the cut-off value is the first topic that will be discussed in this section.

Another issue is due to the diversification of businesses activities. When businesses have secondary activities data may be underestimated because the statistical unit is, nowadays, the enterprise for most of business surveys (except for the Annual Industrial Survey, which is the enterprise and local unit), and therefore the concept of the principal activity is utilized as recommended by *International Standard Industrial Classification of all Economic Activities* - ISIC (United Nations, 2008). The revision project will define the appropriate statistical unit for data collection of secondary economic activities. Other three issues under evaluation are the expansion of data collection coverage of the North Region of Brazil; non-response treatment; and calibration.

Some countries' experiences have shown, in general, that their strata are constructed based on economic activities, number of employees, and revenue (when available). When revenue information is available from administrative records, countries create the certainty stratum by defining a cut-off percentage of total revenue. Among the institutions that do not have access to administrative data on revenue and therefore only use number of employees, there is no consensus regarding the definition of the boundary between certainty and sample strata. Basically, each institute adopts its own methodology. Moreover, it was found that in other countries the relationship between the sizes and representativeness of the strata show a more balanced ratio than in the Brazilian case (where the proportion of certainty strata is much higher than the sample strata). The revision of the relative sizes of the two strata is part of the IEES project.

As part of the exploratory studies it was noticed that throughout the years due to the increased number of large enterprises in Brazil the criterion to construct the strata based on the number of employees caused a disproportionate increase of the number of enterprises allocated to the certainty strata. This is placing a huge increase of the field work. In particular, this increase is largely seen in the services and wholesale and retail trade annual surveys.

The current cut-off value for the industry and the construction surveys is 30 employees, and for the trade and services surveys is 20 employees. These values were set in late 1990's and need to be revised due to the previously mentioned increasing size of the enterprises since then. Also the number of employees varies widely depending on the characteristics of the enterprise's economic activities and location (State and Region). Economic activity and location must therefore be also considered in the new sample design, allowing for optimization and precision of sample size.

The studies suggested that varying the strata cut-off values by economic activity contribute greatly to the efficiency of the sampling design and to a most consistent representation of each economic activity, since the number of employees of an enterprise depends on its economic activity. Following the same reasoning, the cut-off values based on gross revenue should not be the same for all States as it can be seen in international practices and recommendations.

The motivation for the analytical studies were to evaluate the issues mentioned above, and especially evaluate the 'small' size of the sample stratum in relation to the certainty stratum, which results in very high weights for some enterprises and in some cases results in strata with very few enterprises, increasing the instability of the estimates.

To evaluate these issues studies are being carried out based on the algorithm proposed by Lavallée and Hidiroglou (1988) and a generalization of this work, proposed by Baillargeon and Rivest (2009). Such evaluation made by IBGE's staff consisted of trying several alternatives, that is, changing parameters associated with economic activities and location. This resulted in sample sizes much smaller than those now being practiced and a more balanced ratio between the size of the certainty and the sample strata. In all cases there was a significant reduction of the final sample size and a significant reduction of the certainty stratum compared to the sample stratum. This solves the problem of rarefied sampling strata and high weights for the enterprises in them. Consequently this will decrease the problem of non-response due to too few enterprises in the sample strata.

Other two issues are part of the staff's concerns. Enterprises indicated as active on the Central Register (they are part of the frame) but during data collection inform they are not active anymore and with no employees demand more intensive treatment of non-response. On the other hand, these enterprises should not be excluded from the sample because, although with no employees, they may have expressive revenue (this would underestimate economic activity).

As mentioned another ongoing study refers to the appropriate statistical unit for data collection due to regionalization and secondary activities for enterprises located in multiple sites and with various economic activities. Also data for all important economic activities in a region should be collected and this refers back to the definition of statistical unit.

Currently the statistical unit is the enterprise (the exception is the industrial sector where local units are included in the sample). This may underestimate enterprises' secondary activities. However, to collect data in all establishments is inefficient or infeasible (especially for services and wholesale and retail trade sectors), and it would also increase the response burden by dramatically increasing the number of questionnaires.

As mentioned not having data collected in local units (mainly in wholesale and retail trade and services enterprises), may be causing underestimation of secondary economic activity. Some National Statistical Offices adopt the profiling procedure to address this issue, which briefly consists in identifying the enterprise's relevant secondary activities and defining along with the enterprise the statistical units for data collection process.

Due to limited staff IBGE will only be able in the short term to adopt the profiling procedure for all enterprises with relevant secondary activities in the industrial sector (due to the smaller number of enterprises). For this sector, IBGE is suggesting data collection for all local establishments with economic activity or in locations other than the headquarter. This procedure will not be possible for establishments in service and trade sectors. Other studies for the best solution to this issue will continue.

Another study under way is expansion of data collection in the wholesale and retail trade and services surveys in the North region of Brazil. Studies show underestimation in these economic activities data in this region.

Another major investigation is the greater use of administrative records. This, and the use of a harmonized questionnaire, will minimize the problem of non-response. With regard to imputation, harmonized criteria and methods are being defined for all annual structural surveys.

Staff is also working on the issue of sample calibration. This involves the choice of variables to be used, for example, the amount of salary paid by the enterprises has not been a very good variable for calibration. Nevertheless, IBGE may have to keep using this variable because it is the only economic data available in the Central Business Register (CEMPRE).

In CEMPRE the enterprises´ economic activity classification may be different than that obtained during data collection, especially for the service enterprises. This problem should also be addressed in the calibration process. This raises the issue about the level of aggregation at which the calibration process should be performed. Studies and simulations are underway in search of solutions for calibration issues and for others addressed in this paper.

4. Final observations

The revision of IBGE's structural surveys with regard to the harmonized questionnaire and to future sampling design must include procedures to ensure the comparability of new information with those of existing time series. The revision should also consider the SNA needs, as the central conceptual framework of IEES, and be reconciled with the revision calendar of national accounts.

The development and implementation of IEES, due to their complexity should be carried out step by step without the loss of view of the whole system. For example, decisions on the harmonized questionnaire and on the sample design should take into account the impact on short-term monthly surveys and on special surveys and also on future greater statistical use of administrative records (Quintslr, 2014). In the future a needed revision of the survey on products' data will be developed by IBGE´s staff.

Finally, access to tax administrative records would alleviate some of the identified problems, diminish the response burden, and contribute decisively to the accuracy and efficiency of IBGE's annual structural surveys and of IEES as a whole.

References

ABS (2015). Australian Bureau of Statistics,
http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8155.0Explanatory%20Notes12010-11?OpenDocument# (Retrieved February 3, 2015).

Baillargeon, S. & Rivest, L. (2009). A General algorithm for univariate stratification. *International Statistical Review*, 77 (3), 331-334.

Erbisti, B. & Quintslr, M. (2014). Registros de Empresas e Banco de Dados Integrados nos Sistemas de Estatísticas Econômicas Integradas. *Seminário de Metodologia do IBGE   (SMI)*, 3.

IBGE (2014). Sistema de Estatísticas Econômicas Integradas - SEEI: O Projeto de reformulação das estatísticas econômicas. Rio de Janeiro. Comitê de Governança das Estatísticas Econômicas. *Textos para Discussão*. Diretoria de Pesquisas, n. 54.

Lavallée, P. & Hidiroglou, M. (1988). On the stratification of skewed populations. *Survey Methodology (Statistics Canada)*, 14 (1), 33-43.

Quintslr, M. (2014). Registros Administrativos no contexto dos Sistemas de Estatísticas Econômicas Integradas. *Seminário de Metodologia do IBGE (SMI)*, 3.

United Nations (2008). International Standard Industrial Classification of All Economic Activities - ISIC. Rev. 4. New York:, Department of Economic and Social Affairs, Statistics Division, 2008. (Statistical papers. Series M, n. 4, rev. 4).

United Nations (2013). Guidelines on Integrated Economic Statistics. *Department of Economic and Social Affairs, Statistics Division, Studies in Methods, Series F*, No. 108, New York.