

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Editing

3-7 October 2022, (virtual)

The use of administrative records for the imputation of data from the Economic Censuses of Mexico

Jose Luis Mercado Hernandez (National Institute of Statistics and Geography, Mexico)

jose.mercado@inegi.org.mx

I. Introduction

1. In Mexico, the Economic Censuses are carried out every 5 years uninterruptedly since 1930 and, since the census event of 1989, it was decided to implement a strategy that allows monitoring the establishments defined as *large*¹ throughout the different census operations, given the economic importance they have in the activities in which they participate.
2. As a complement to the strategy promoted in 1989, from 2009 the monitoring of the registered establishments was no longer carried out only to the large establishments, but the universe was expanded, which allowed to know the history and evolution of each one, in addition to giving way to other studies such as economic demography, with which it is possible to know how many economic units survived between one census and another, how many died and how many were born.
3. The information generated by the Economic Censuses has multiple applications in the public, private and academic sectors, but it is also used within the INEGI, for the updating of the System of National Accounts of Mexico (SCN), the System of National Economic Surveys (EEN) with which information on the situation is generated, as well as the Statistical Registry of Business of Mexico (RENEM). And it is precisely the latter who provides the *directory of economic units* that serves as a start for the census since it includes the updated list of establishments captured in the immediately previous census, generated from the exploitation of administrative records such as the registers of the Tax Administration System (SAT); of the entity in charge of the supply of electricity in Mexico, the Federal Electricity Commission (CFE), and information generated by the field operations of the national economic surveys (EEN).
4. In addition to the identification data, location and economic activity carried out by the economic units included in the RENEM directory, there is information on the Total Gross Production (PBT) generated by the economic units in the previous census; the income declared by the economic units to the SAT, and the size of the economic unit from the number of jobs. This information is useful to measure the impact of these economic units on the activities in which they are classified and determine which of them should be considered as priorities during the field operation.
5. An important aspect to consider are the characteristics that economic units in Mexico have in terms of their size. According to the results of the 2019 Economic Censuses (whose reference

1/ A large establishment is considered to be one that meets at least one of the following characteristics: income equal to or greater than 50 million pesos per year; 50 or more persons employed; be part of a national company, that is, share the same business name with one or more establishments located in a different state, among other characteristics.

year is 2018), which are the most recent census, 94.9% of the total of the economic units surveyed are made up of establishments called micro, which employ on average ten or fewer people in each establishment, generate 37.2% of total employment and contribute only 14.6% of total value added in the census reference year.

6. On the other hand, small and medium-sized establishments, which on average occupy between 11 and 250 employed people and are known as SMEs, account for 4.9% of establishments, generate 30.7% of employment and contribute the same proportion to the value added (30.7%).
7. Finally, large establishments (whose average occupation is 251 people or more) only represent 0.2% of the total economic units, however, they employ 32.1% of the employed personnel and contribute more than half of the total value added for that year (54.7%). These figures highlight the magnitude and importance that the establishments corresponding to this segment have in the national economy and highlight the need for a special strategy to obtain their data and ensure that they are part of the census results.
8. These characteristics, together with the diversity of activities under study (990 different activity codes according to the North American Industrial Classification System, SCIAN) require differentiated strategies for obtaining information.
9. In this way, various operational groups responsible for the retrieval of information are formed, which differs not only by the strategies for the collection of information, but by type of interview (direct or indirect), type of questionnaire (sectoral or basic), and means of capture (internet, paper or electronic device), among other aspects.

Strategy for capturing information in Large Establishments

10. In the segment of large establishments, a strategy is defined in which the year prior to the census, the directories available to INEGI are verified (as mentioned above, those are provided by RENEM)
11. After reviewing the directories, a visit is paid to the parent companies of the multi-establishment enterprises (made up of more than one establishment with which they share the same corporate name) and the single establishments are used to update the data related to identification, location, economic activity and form of organization, as well as data on the personnel employed, among other variables, including company branches, which updates the formation of the company and plans to capture information in the census year.
12. Part of the strategy for this segment of establishments is to establish with the parent or the single establishment the dates and ways of capturing the information and methods are established for the completion of the questionnaire, such as advice on which variables correspond to each category, proper use of the system for filling out the questionnaire on the Internet, or even the capture of the information itself, among other supports.
13. This strategy has yielded good results, given that the non-response rate is insignificant. However, there are cases in which the informants do not deliver the information and it is here that statistical tools are used for the imputation of the information.

II. Data imputation as an alternative to non-response

A. Imputation of information for large establishments

14. When a large establishment does not provide its information, it is first verified if it is part of the sampling framework of the National Economic Surveys (EEN); also, if its information is available in the SAT, and if the financial statements of the company in question are published on the Internet.
15. Thanks to the verification carried out the year before the census, there is available information on how the company is formed, that is, how many establishments make it up, the economic activity in which each one, and how many and which only provide support services to the producing establishments.
16. In addition, it is verified if the company is included in the sampling framework of the ESNs, and if so, the data collected both in the monthly surveys of the reference year, and the data of the annual survey, of the immediately previous period, are reviewed.

(a) In case data's company has been captured by the National Economic Surveys

17. The data obtained during the monthly operations for the census reference year are taken up, which are the personnel employed in their different categories; the details of remuneration; the value and detail of the production; the value and detail of the sales of the products produced; the other income derived from the activity and expenses included in the monthly questionnaire.

Monthly Manufacturing Industry Survey Questionnaire

VII. CONSUMO DE BIENES Y SERVICIOS Anote el valor de los bienes y servicios consumidos por este establecimiento en el mes de referencia, considerando todas las erogaciones relacionadas con estos bienes, como los fletes y seguros; deduciendo los descuentos, rebajas y concesiones recibidas. Incluya el valor de los bienes y servicios que recibió en transferencia para su consumo o transformación. NO INCLUYA EL IVA	MERCADO NACIONAL (Miles de pesos)	MERCADO EXTRANJERO (Miles de pesos)
	B	C
K₃₁₆ Materias primas consumidas propias Si consumió materias primas de su propiedad, no olvide anotar en la variable M ₃₁₀ las ventas totales de productos elaborados con materias primas propias.		
K₃₁₇ Materias primas consumidas propiedad de terceros En la columna "C" incluya aquellas que están amparadas por un pedimento de importación temporal, considerándose en todos los casos como operaciones en el mercado extranjero.		
K₁₀₀ Costo de las mercancías revendidas		
K₃₁₀ Envases y empaques		
K₅₀₀ Pagos por alquiler de bienes muebles e inmuebles		
K₈₁₀ Pago por suministro de personal Si tuvo pagos por suministro de personal; no olvide anotar en la variable I ₀₀₀ el total del personal suministrado por otra razón social.		
K₉₆₁ Fletes de productos vendidos		
K₇₁₀ Pagos por maquila		
K₈₃₀ Servicios profesionales. Incluye trámites aduanales		
K₉₉₉ Otro consumo de bienes y servicios		
K₀₀₀ TOTAL del consumo de bienes y servicios		

Where:

K310 Value of own consumed raw materials

K317 Value of consumed raw materials owned by third parties

K100 Cost of resold goods

- K910 Payments for the supply of personnel
- K961 Freight of products sold
- K710 Payments by maquila
- K630 Professional services, includes customs procedures
- K999 Other consumption of goods and services
- K000 Total consumption of goods and services.

18. To obtain the value of the variables detailed in this questionnaire, and that are equivalent to the census variables, the annualized data is taken up and it is settled on the corresponding census variables. Those that are not requested explicitly disaggregated, are disaggregated from the *other expenses* variable, where expenses such as electricity are included, for example, which is not captured in the monthly survey but it is explicitly requested in the census questionnaire.
19. With this information, the data of the annual survey of the year prior to the census, captured by the EEN are taken up and, with this source, the proportion that each of the expenses captured in this operation represents with respect to the total of the expenses is calculated, and the result is multiplied by the item of other expenses for the consumption of goods and services, which is where the informant included all those expenses that he did not report specifically. In this way, the chapters of employed personnel, remunerations, expenses and income are completed.

Questionnaire of the Annual Industrial Survey

VI. CONSUMO DE BIENES Y SERVICIOS (Del 1 de enero al 31 de diciembre de 2021) Indique el valor de los bienes y servicios consumidos por este establecimiento durante 2021, considerando todas las erogaciones relacionadas con estos bienes, deduciendo los descuentos, rebajas y concesiones recibidas. Incluya: el valor de los bienes y servicios que recibió para su consumo o transformación. NO INCLUYA EL IVA, LAS COMPRAS NI LA DEPRECIACIÓN DE ACTIVOS FIJOS	MERCADO (Miles de pesos)	
	NACIONAL B	EXTRANJERO C
K₃₁₀ Materias primas y auxiliares consumidas, propias Incluya: las materias primas y auxiliares enviadas a maquilar, las consumidas que provienen de otros establecimientos de la misma empresa, y el agua cuando se utiliza como materia prima. Excluya: las materias primas y auxiliares propiedad de terceros que recibió para maquilar, las materias primas y auxiliares enviadas a otros establecimientos en el mismo estado en que se adquirieron, así como los materiales para el mantenimiento o la construcción de activos fijos para uso propio. <i>Cuando anote materias primas y auxiliares consumidas propias, no olvide reportar en la variable M310 las ventas netas de productos elaborados con materias primas propias.</i>		
K₂₁₀ Consumo de envases y empaques Incluya: las cajas, bolsas y otros materiales para proteger, transportar, presentar y contener las mercancías. Excluya: los que sean propiedad de terceros.		
K₄₁₁ Consumo de combustibles y lubricantes Incluya: el consumo de gasolina, diesel, gas, grasas, aceites y otros combustibles y lubricantes.		
K₄₁₂ Consumo de energía eléctrica		
K₇₁₀ Pagos por maquila Indique los pagos efectuados a otras razones sociales por los servicios industriales de transformación parcial o total de sus materias primas.		
K₁₀₀ Costo de las mercancías revendidas Anote el costo de las mercancías vendidas en el mismo estado en que fueron adquiridas. Incluya: el valor de las materias primas vendidas o transferidas en el mismo estado en que se compraron a otros establecimientos tanto de la misma empresa, como de otras empresas. Excluya: el valor de las mercancías recibidas en consignación y el pago de impuestos al comercio exterior. <i>Cuando anote el costo de las mercancías revendidas, no olvide reportar en la variable M100 los ingresos por la compra-venta sin transformación de mercancías.</i>		
K₂₂₀ Pagos por servicios de comunicación Indique los pagos efectuados por los servicios de telefonía, internet, radiocomunicaciones, etcétera.		

20. As mentioned, the specific weight of the missing variable is calculated by dividing the variable *K412 electricity consumption* by *K000 Total consumption of goods and services*, and the result is multiplied by the value of K000, obtained from the monthly operation; the result is settled on the corresponding variable of the census questionnaire. This same procedure is performed with the other missing variables. Once the missing values are obtained, the data of *K999 other expenses* is adjusted in such a way that the sum of all the variables is equal to the total recorded in K000.
21. Regarding stocks, it is adjusted by taking up what was reported in the annual operation of the EEN of the year immediately prior to the census from the final stocks, plus what was produced, minus sales. The issue of fixed assets is adjusted from what was reported in the annual operation by applying the variation of the National Producer Price Index by type of activity.
22. Finally, regarding the qualitative part required in the census questionnaires, and that are not part of the coverage of the EEN, such as training topics, age groups of employed personnel, problems for the performance of activities, among other topics, what was reported by the nearest neighbor (data from an establishment of similar characteristics is taken to impute data for another one) is taken up. This is done for all the establishments that make up the company.

(b) In case the company has not been captured by the National Economic Surveys

23. When the establishment is not part of the sampling framework of the EEN, the data reported by the economic unit to the SAT or the income statements that are publicly accessible via the Internet are taken up and the information is imputed from the following two scenarios:
 - If information is available from the immediately preceding census, the percentages of each variable are calculated, with respect to the totals of expenses and income (as appropriate), by dividing each of the partials by the total, and these percentages are multiplied by the data obtained from the SAT or the financial statements to complete the corresponding chapters. The part of fixed assets is updated with the National Producer Price Indices.
 - If there is no prior information, it is reviewed if in the databases of the corresponding activity there is information from an establishment with similar or close characteristics and the percentages obtained from dividing the partials by the total of each chapter are applied to the values obtained from the alternate source.

B. Imputation of information for Micro, Small and Medium establishments

24. For Micro, Small and Medium Establishments (MIPyMES) the strategy followed by the field operation at the time of obtaining the census information consists of the verification of the directory generated by the RENEM, where the location, identification and economic activity data are reviewed, while the registrations or newly created establishments between one census and another are incorporated. The records that were no longer located during the route called "sweep" (where all urban localities with 2500 inhabitants or more are visited) are canceled and even doors are knocked to identify the economic activities carried out inside a house.
25. The interview for the collection of information is carried out through a mobile computing device (DCM), which has loaded the corresponding questionnaire, and begins with the verification, that is, the name of the establishment, its location and the activity carried out, or, requiring those

same data to the newly created establishments and to those who reported any change in the data recorded. The interview continues with the request of the employed personnel to then request the part related to remuneration, expenses and income; it is from these chapters of the questionnaires where some informants refuse to provide data. In these cases, the method used to impute the missing data is that of *averages*, that is, to divide the total value of each variable required in the questionnaire by the number of establishments registered with information, considering the size of the economic unit, the economic activity and the geographical space where it is located.

26. This is done once the review and editing of the information of all the records received is completed. The average of each variable by size of establishment (measured through the employed personnel) is obtained for each of the classes of activity involved and with a geographical cut at the municipality level.
27. The missing information of the establishments that only provided us with information up to the chapter of occupied personnel, are filled with the averages obtained in each variable.

Note: The questionnaires shown here can be consulted at the following addresses:

Annual Manufacturing Industry Survey Questionnaire

https://www.inegi.org.mx/contenidos/programas/eaim/2013/doc/CEAIM_2022.pdf

Monthly Manufacturing Industry Survey Questionnaire

https://www.inegi.org.mx/contenidos/programas/emim/2013/doc/c_311110.pdf

Questionnaire for the Manufacturing Industry of the Economic Censuses 2019

https://www.inegi.org.mx/contenidos/programas/ce/2019/doc/cuestionarios/IOS_2019.pdf