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

Expert Meeting on Statistical Data Editing

José Luis Mercado Hernández

3-7 October 2022
Since 1930, the Economic Censuses in Mexico are conducted every five years.

More than 90 years of economic history.

The data collected correspond to the economic activity carried out from January 1 to December 31 of the year prior to the year of the survey; in the Economic Censuses 2024 it will correspond to the year 2023.
Observation Unit

The observation unit is the establishment, which allows the information to be referred to the geographical space where the activities are carried out.

Installations not anchored or not permanently subject to the ground are not registered.
Basic information for the updating of the infrastructure in economic matters

- 100% upgrade
- National Economic Surveys
- System of National Accounts
- Price indices
- Geostatistical Framework Update
- Population and Housing Census
According to the Economic Censuses 2019, the establishments are predominantly Microbusinesses; employ 4 out of 10 Employed Persons and generate 15 out of every 100 pesos of the Census Value Added.
From the EC 1989 onwards, each of the Large Establishments has been monitored on time.

A “large establishment” is the one that complies with at least one of the next characteristics: equal income or greater than $2.5 million annually; 50 or more persons employed; be part of a national company, i.e. sharing the same business name with one or more establishments located in an entity different federative, among other characteristics.

From the EC 2009 onwards, the other establishments have been added to the monitoring.
### Operational strategy for data collection

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Collection Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit via directory to large establishments (GPN) of Manufacturing, Commerce, Services</td>
<td>Sectoral questionnaires with the specific theme of each sector</td>
</tr>
<tr>
<td>Visit via directory to establishments in all other sectors</td>
<td></td>
</tr>
<tr>
<td>Comprehensive “sweep” in all urban areas (massive operation)</td>
<td>Basic questionnaire</td>
</tr>
<tr>
<td>Comprehensive “sweep” of rural AGEBs selected by sampling for Manufacturing, Trade and Services SMEs</td>
<td>Basic questionnaire</td>
</tr>
<tr>
<td>More detailed questions on some topics, to establishments selected by sampling: Non-profit institutions Environment Other</td>
<td>Additional modules to census questionnaires</td>
</tr>
</tbody>
</table>
Means for data collection

All informants may opt for these modalities:

- Mobile Computing Device (new DCM).
- Printed questionnaire
- Online Questionnaire

97% of the questionnaires
50% of the value of census production
Procedure for the imputation of information of large missing establishments

- When the establishment is part of the sampling framework of the National Economic Surveys (EEN)

- When the establishment is not part of the Sample Framework of the EEN and there is information from the Tax Administration Service and/or the Financial Statements:
  - With information from the immediately preceding census
  - No immediately preceding census information
Procedure for the imputation of information of large missing establishments

When the establishment is part of the universe of the EEN

For variables that are equivalent, the annualized data of the ESNs is taken and recorded in the corresponding variable of the EC questionnaire.
Procedure for the imputation of information of large missing establishments

When the establishment is part of the universe of the EEN

➢ For the grouped variables, it is calculated, either from the annual questionnaire of the EEN captured one year before, the census reference period or, from the previous census, of the combined value of both variables, which proportion corresponds to rentals of movable property and which to real estate.
Procedure for the imputation of information of large missing establishments

When the establishment is part of the universe of the EEN

In cases where the information is grouped into a single concept, it is calculated, either from the annual questionnaire of the year prior to the census reference or from the previous Census, how much these concepts weigh in the total of the expenses and the result obtained is multiplied by the total of the expenses (K000A) reported by the monthly Survey. The data obtained is based on the corresponding variable and is subtracted from the concept that was captured grouped by the monthly operation of the EEN. Once the chapter is concluded, the sums are reviewed and the following chapters are continued.
Procedure for the imputation of information of large missing establishments

When the establishment is part of the universe of the EEN

The value of the fixed assets reported by the EAIM for the year prior to the EC reference year is applied the variation of the National Producer Price Index by type of asset.
The tax data are presented by company, so it is reviewed, from the verification carried out the year prior to the census, by how many establishments are formed and type of activity carried out by each of them.

If information is available from the previous Census, it is calculated how much each establishment contributes to the total of the company and the proportions obtained are replicated with the data reported in its annual declaration.

If there is no information from the previous Census, the distribution reported by the nearest neighbor for each establishment that makes up the company is resumed.
Procedure for the imputation of information of Micro, small and medium-sized establishments

➢ Once the questionnaires captured with information have been reviewed and refined

For all variables, the averages are calculated by economic activity, size of the establishments (measured by the number of people employed on average per economic unit) and federal entity and municipality. These averages are recorded in the missing questionnaires.
Conclusions

The differentiated operational strategies designed according to the sizes and activities of the establishments allow a longitudinal follow-up for practically all of them, which contributes, along with the information available in the administrative records, to greater precision when estimating the missing information and editing the one that presents atypicalityes.

Advantages of data imputation

➢ The imputed data, in the case of large establishments, reflect what is actually exercised by them.

➢ The level of accuracy in the estimation can be done with high levels of geographical and sectoral disaggregation.

➢ The results obtained for the imputation of SME Establishments over time and the low levels of non-response recorded reduce the risk of giving bias to the information.

Disadvantages of data imputation

➢ Applying the structures of past events to the current data does not allow to reflect the evolution presented by specific variables, which can impact the calculations of the derived statistics.

➢ With the method applied to SME Establishments, very homogeneous behaviors are presented at very detailed disaggregation levels, with the possible biases that the calculation of the averages entails.
THANK YOU