SAE motivation

INEGI has been investing in the development of new technical and methodological alternatives, enhancing the country’s statistical infrastructure with data to enable informed decisions for the design and evaluation of public policies. The growing demand for timeliness and reliable data, particularly for local governments, and limited resources to enlarge and/or spread samples, prompted the incorporation of SAE methods to produce official statistics (Orozco et al., 2021; INEGI, 2020).

The INEGI has already published experimental small area statistics for disease prevalence at the municipal level and state level labor indicators. In addition, municipal level labor indicators have been produced for 2017-2020.

Experimental labor indicators for municipalities – 2017-2020

The main indicators produced are:

- The ratio of the Economically Active Population to the Population aged 15 and over.
- The ratio of the Employed Population to the Economically Active Population.
- The ratio of the Informally Employed Population to the Employed Population.

Input data

The Mexican National Survey of Occupation and Employment (Encuesta Nacional de Ocupación y Empleo - ENOE) data (referring to the first quarter of every year) is the source of direct estimates whereas the Population and Housing Census, administrative registers of social security, as well as information published from various public institutions in Mexico, provided auxiliary data (for details see Table 1 from Orozco at al., 2021).

The ENOE is the main source of information on the Mexican labor market and publishes monthly and quarterly estimates of employment, unemployment, labor informality, and underemployment. It provides, besides national figures, estimates for four groups of cities defined according to population sizes, as well as statistics for each of the 32 Mexican states and for 39 cities (INEGI, 2019), however, the 2457 municipalities (under the current geostatistical framework) constitute unplanned domains.

Building the SAE model/ Model Building

The small area estimation method is based on an area-level model that takes into account spatial dependence between areas (municipalities). The model incorporates spatially correlated random area effects and is implemented using a Spatial Empirical Best Linear Unbiased Predictor (SEBLUP) that borrows strength from correlated random area effects between neighbor municipalities. The domains of interest correspond to 2457 Mexican municipalities.

A mixed linear model was built for each variable: Economically Active Population, Employed Population, and Informal Employed Population. The response variables are direct estimates produced by ENOE for the first quarter of every one of the given years (2017-2020). The auxiliary variables were selected according to the response indicators: population affiliated to different social security systems compared to the population aged 15 years old and over; enrolled in preschool, elementary and junior high school population compared to the population aged 15 years old and over; female population aged 18 to 29 years old compared to the population aged 15 to 29 years old; male population aged 15 to 44 years old compared to the population aged 15 and over; population aged 45 years old and over compared to the population aged 15 years old and over; and population under 15 years old and 65 years old and over compared to the population of 15 to 64 years. In addition, measures of the distance between the municipal capitals were also incorporated into the model's spatial component.

Information about the SAE models, variable selection procedures, verification of model assumptions, as well as evaluation of model-based estimates, are reported by Orozco et al. (2021), and the data is disseminated by INEGI as experimental statistics.

Benchmarking/data validation

Traditional quality assessment of model-based estimates was carried out, e.g. comparison of final model results using high-quality direct estimates of the ENOE or municipal percentage structures of the Census applied to ENOE. The latter was used when the direct ENOE estimates were low-quality estimates. Coefficients of variation model-based were also assessed.

Update of SAE methods and small area estimates

INEGI is working to provide 2021 labor figures in the first half of 2022, focussing on improving the relevance and timeliness of these experimental statistics. Furthermore, new approaches based on multinominal and Bayesian models are being explored to improve the estimates of interest.
SAE work within the organization

There is a work team dedicated exclusively to SAE within INEGI. As SAE techniques are useful in statistics, the INEGI is looking to provide not just labor figures but also other types of sociodemographic figures. The institute is motivating the team to keep their tasks in this field.

Future work on SAE

At this moment INEGI is working to produce income figures for Mexico’s municipalities, so the publication of labor figures for Mexico’s municipalities constitutes the beginning of additional implementations on the subject.

Challenges

There are two main challenges to producing SAE estimates:

The first one is related to the stakeholders. The SAE work team within the INEGI had to convince the authorities in turn the statistical office of the enormous utility of SAE in terms of expanding the offer of statistical information from sample surveys and the considerable savings in financial resources.

The second one is about the auxiliary variables. High-quality auxiliary variables are required to produce SAE estimates, but it is sometimes not easy to meet the established quality standards in administrative records.

Sources/References


Information provided by

Edgar Vielma Orozco
Jorge Ochoa Setzer
José Vences Rivera
Gustavo Aguilar Mata
Josué Roberto Esquivel Balseca
Alfonso Millán Licona
César Bistrain Coronado
José Rodolfo Campos Luévano
Marco Antonio Vázquez Andrade
José Daniel Loera Serna
Luis Francisco Salazar López