Promoting inclusive official statistics on migration

Prediction Alternatives for Estimating Households with Emigrated Members

January / 2023

GOBIERNO DE COLOMBIA

Content

Context about Prediction Alternatives for Estimating Households with Emigrated Members

The Colombian National Statistical Office usually produces estimations of internal migration based on the results of population censuses and social surveys, but there is a lack of disaggregated information about the main small areas of origin of the population that emigrates from Colombia.

- **Uses frequentist and Bayesian approaches based on a Fay-Herriot model**

- **Small area estimation**

- **Hierarchical Bayes prediction**.

For the first time, a SAE procedure is conducted to estimate international emigration at the municipal level.
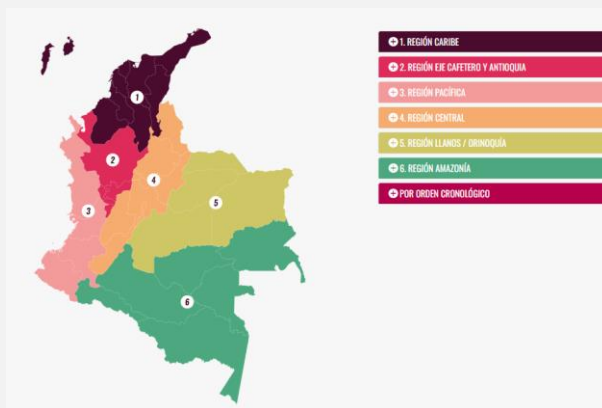
The Fay-Herriot model for small areas estimation

The goal is to explore the use of small area estimation (SAE) theory to estimate the proportion of households with at least one usual member living abroad – PHMLA in intercensal periods at the municipal level.

As the PHMLAs is a proportion, one possibility is to model the dichotomic variables indicating whether a household has emigrated members.

As the information comes from a survey, the information must be aggregate in municipalities, then the area-level model approach to SAE relies on previous calculations of direct estimators of PHMLAs and their variances. These direct estimates are entered into a Fay-Herriot (FH) model that enables the calculation of empirical best linear unbiased predictors (EBLUP) of PHMLAs at the municipal level.

The Fay-Herriot model for small areas estimation

The Fay-Herriot model is a linear mixed model that considers:

A finite population of size N, where the units are households
For each household, the variable of interest is y.

$$y_j = \begin{cases} \text{1 if at least one person who is a usual member of the household j currently lives abroad} \\ \\ \text{0 no one in the household j currently lives abroad} \end{cases}$$

The population is divided into $\boldsymbol{D}$ small areas (Colombian municipalities).

Let $\boldsymbol{S}$ be a sample extracted at random from the population according to a given sample design.
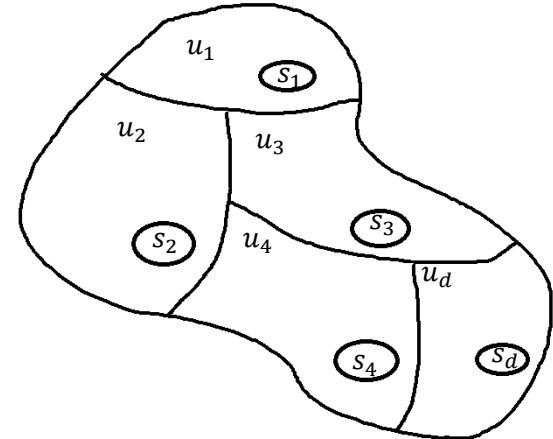
The Fay-Herriot model for small areas estimation

The population and the sample are denoted by

$$U = \bigcup_{d=1}^{D} U_d$$

and

$$s = \bigcup_{d=1}^{D} s_d$$

The objective is to estimate the PHMLAs per municipality

$$\bar{Y}_d = N_d^{-1} \sum_{j \epsilon U_d} y_j$$

The Fay-Herriot model for small areas estimation

When applying the Fay-Herriot model, the underlying idea is that the direct estimators of the proportions are not accurate, given that the municipality sample sizes are small.

For this reason, the model improves the estimates by taking auxiliary information from the explanatory variables x_d, including the variability between municipalities through the random effect u_d and considering the random sampling error e_d derived from using a direct estimator as model target variable.

In practice, the error variances σ_d^2 can be substituted by estimates $\widehat{\sigma_d^2}$

Results: Estimating International Emigration from Small Areas in Colombia

- The Fay-Herriot model is fitted to the DHS 2015 based on available census information
- Calculates the EBLUP of the selected indicator of international emigration at the municipal level
- Applies the GVF method to estimate the variances of the direct estimators of PHMLA
- The final variance estimates are the predicted values of a selected log-linear regression model

Fuente: DANE.

Results: Estimating International Emigration from Small Areas in Colombia

Fuente: DANE.
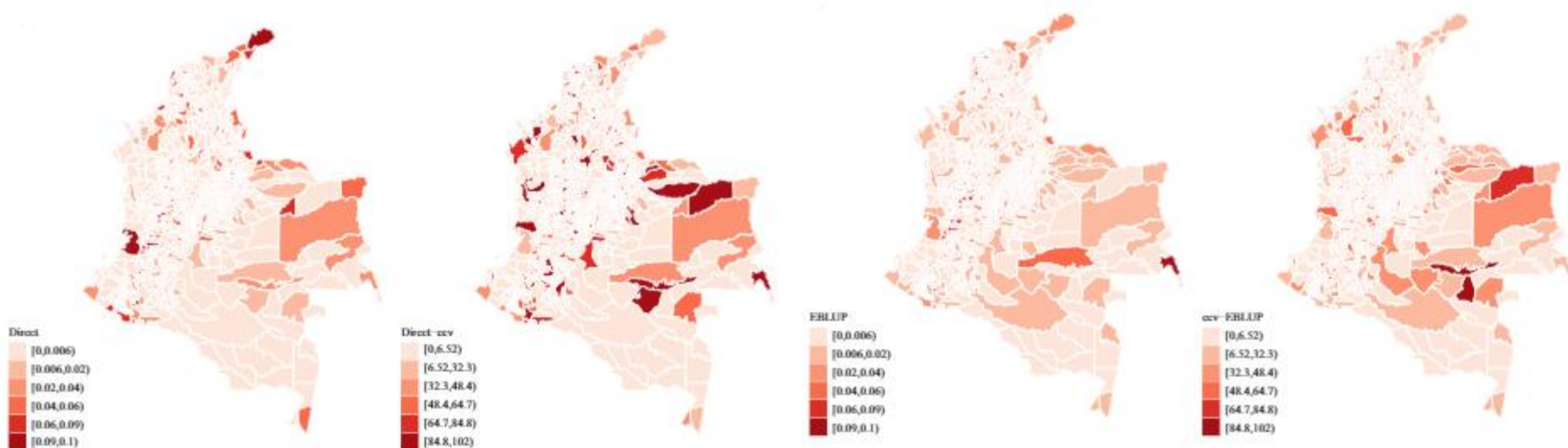
Colombian results
Small areas estimation for Households with Emigrated Members

Estimation for people at the municipal level
Households with Emigrated Members

Municipality	DIR	ecv-DIR	EBLUP	ecv-EBLUP	HB	ecv-HB
Bogotá D.C.	0.04359	14.75	0.03469	13.22	0.03453	15.07
Cali	0.08919	14.11	0.06040	10.53	0.06067	9.82
Armenia	0.08568	14.23	0.07397	9.57	0.07345	9.60
Cartagena de Indias	0.07348	16.80	0.03001	14.47	0.03078	14.27
Soledad	0.04368	24.79	0.02408	16.74	0.02437	17.32
Chía	0.04563	75.30	0.03563	13.87	0.03542	16.08
Villa del Rosario	0.04623	48.87	0.03362	9.82	0.03359	12.36
Manaure	0.04320	50.62	0.02751	17.29	0.02773	18.54
Valle del Guamuez	0.04872	67.01	0.02579	16.69	0.02609	18.50
San andrés de Tumaco	0.03600	45.94	0.01888	20.08	0.01922	22.29

Fuente: DANE – Dirección de Censos y Demografía.

Conclusions

- The main Colombian cities, which are Bogota, Cali, Medellin, Barranquilla and Cartagena, present some of the highest estimations in the proportions of households that have experienced international migration (with household members abroad). This is interesting inasmuch as these cities are the main areas where services, education and work are concentrated.
- Moreover, high indicators can also be seen in municipalities of the coffee-growing region, which doubtless have a series of migration determinants and are different to those of the capital cities. This is a topic that merits further investigation in greater detail in order to better identify and characterize these differences.

Fuente: DANE.

Conclusions

- It must be highlighted that events that occur in specific geographical locations, in this case capital cities, often have consequences on their direct neighbors and the adjoining municipalities. In this sense, what was mentioned earlier can be reconfirmed: migratory phenomena do not occur homogeneously within the country. Instead, there is a selection filter meaning that there is a greater international emigration in some populations; for example, educational level and employment on an individual scale, but also at the scale of small areas. In fact, it is for these very same reasons that migration is affected by the economic cycle of countries and regions.

Fuente: DANE.

Conclusions

- In Colombia, high-quality disaggregated indicators on a municipal level are only available after each population census, that is, every 10 or 15 years. In the intervening years, it is normal for changes to occur inmigration patterns and in the intensity of the phenomenon. For this reason, sample surveys provide data that makes it possible to obtain estimates for interim periods (intercensal estimations) and for periods following censuses, whereby survey results are available (postcensal periods).

- The proposed statistical methodology can be easily applied to developing countries, where sufficient data sources are not available in quality and quantity.