IAOS-ISI 2024
MEXICO CONFERENCE

Inter-American Statistical Institute (IASI) - IASI Session: Sampling and Official Statistics

Instructions: Click on the link to access each author's presentation.

Organiser: Mercedes Andrade Bejarano

Discussant: Pedro Luis do Nascimento Silva

Participants:

Andrés Gutiérrez: ECLAC and UNFPA approach to model populations in Latin America and the Caribbean

Leonardo Trujillo Oyola:* Estimation of proportions in small area estimation: Machine Learning approach

Denise Silva: Time series models for repeated surveys

* Work presentation not available or non-existent

IAOS-ISI 2024
MEXICO CONFERENCE

ECLAC and UNFPA approach to model populations in Latin America and the Caribbean

Andrés Gutiérrez, PhD.
Regional Adviser on Social Statistics
andres.gutierrez@un.org

INEGI

International Statistical Institute

IAOS
IMPROVING OFFICIAL STATISTICS

Why do we do the things we do?

An SDG perspective

SUSTAINABLE DEVELOPMENT GOALS

1 NO POVERTY

2 ZERO HUNGER

3 GOOD HEALTH AND WELL-BEING

4 QUALITY EDUCATION

5 GENDER EQUALITY

6 CLEAN WATER AND SANITATION

7 AFFORDABLE AND CLEAN ENERGY

8 DECENT WORK AND ECONOMIC GROWTH

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

10 REDUCED INEQUALITIES

11 SUSTAINABLE CITIES AND COMMUNITIES

12 RESPONSIBLE CONSUMPTION AND PRODUCTION

13 CLIMATE ACTION

14 LIFE BELOW WATER

15 LIFE ON LAND

16 PEACE, JUSTICE AND STRONG INSTITUTIONS

17 PARTNERSHIPS FOR THE GOALS

SUSTAINABLE DEVELOPMENT GOALS

11 SUSTAINABLE CITIES AND COMMUNITIES

Make cities and human settlements inclusive, safe, resilient and sustainable

SDG 11: Sustainable communities

- Target 11.1.: By 2030, ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums.
  - Indicator 11.1.1: Proportion of urban population living in slums, informal settlements or inadequate housing.
- Target 11.1.: By 2030, enhance inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning and management in all countries.
  - Indicator 11.3.1: Ratio of land consumption rate to population growth rate.

Censuses and recent experiences in Latin America and the Caribbean

The problem of coverage

- Censuses are massive statistical operations that try collecting data from all areas in the country in a certain period of time.
  - Some countries tried to expand the collection period to lower the under-coverage, implying a tremendous effort in resource mobilization.
  - This solution did not prove to be as effective as expected, and the lower coverage rates kept in some areas.
- The censuses should stop their collection stage after multiple extensions.
  - In several countries returning to collection in the areas of lower coverage was not an option due to limited budget.
  - Incomplete collection along the countries was a common issue.

Some challenges in population censuses

- Population censuses do not always manage to list all households and their populations throughout the country.
  - Complete omission of dwellings or misidentification of the occupancy status of the dwelling.
  - Complete or partial omission of people inside the dwellings.
  - Complete or partial omission of certain geographical areas due to problems of planning of field work, accessibility or security among others during the census enumeration.
- Most of the countries in LAC region are experiencing these kind of challenges in their censuses.

Some challenges in population censuses

- Some countries that have not made the census may face problems getting accurate and precise counts of people.
  - Obsolescence of figures based in old and outdated censuses.
  - Recent migration phenomena increased the need for up to date figures.
  - Need for prediction of counts in some districts and regions

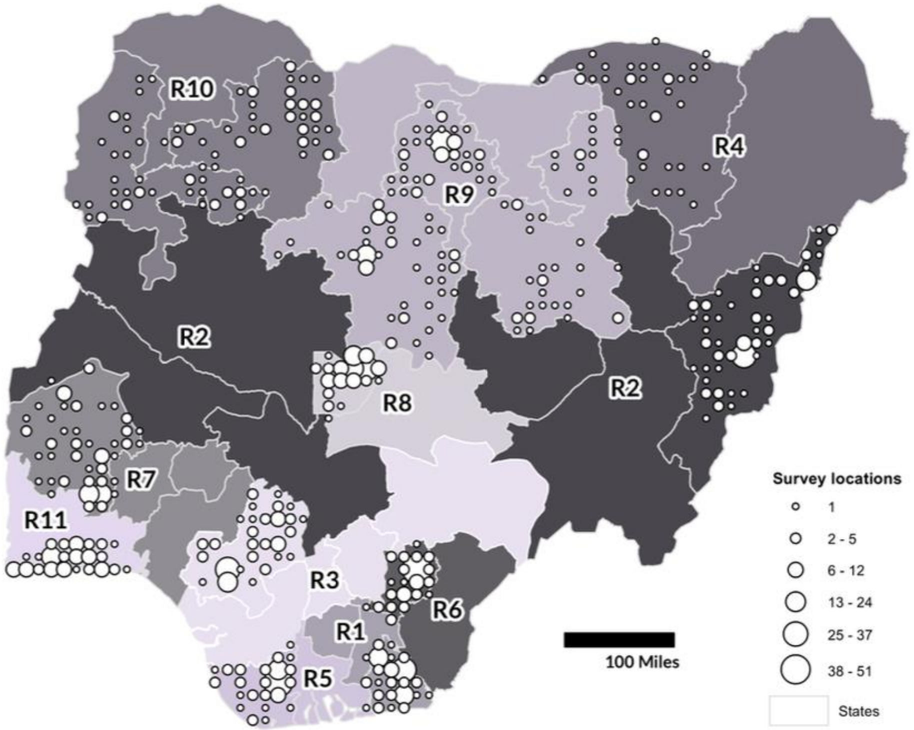
Parsimonious solution

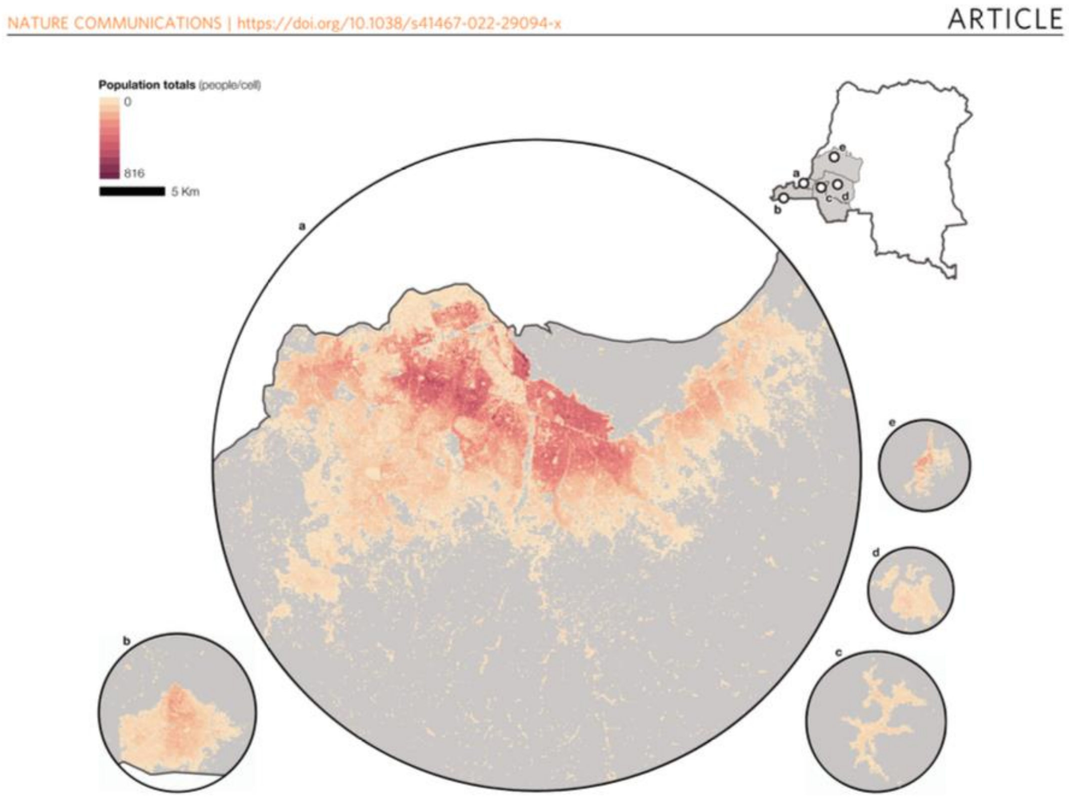
- When incomplete enumeration of areas is a problem in the census, we can rely on statistical models to predict counts of people (along with their demographic structure: age and sex).
  - Model-based estimates of counts represents a new approach to the problem of complete or partial omission.
  - The rationale behind these kind of models is borrowing strength from complete areas.
  - This approach uses remote sensing covariates that should be available for all of the areas in the country.
- In the literature we find a lot of experiences with similar models:
  - Boo, et. al (2022), Leasure, et. al (2020), Berg (2023)
  - ECLAC and UNFPA join venture in Latin America and the Caribbean

Population models

The approach of ECLAC and UNFPA

Models based on enumeration surveys



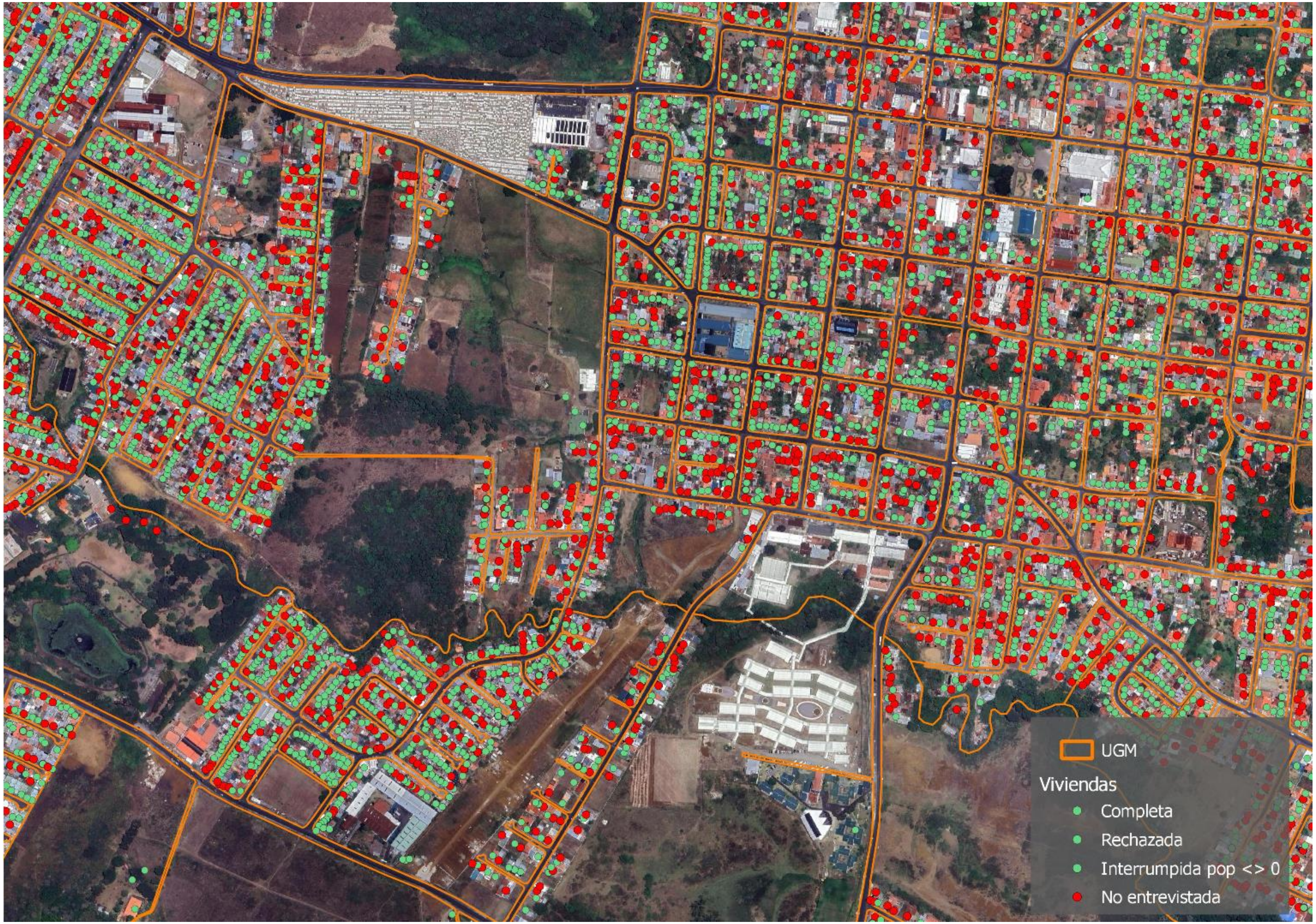

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-29094-x ARTICLE

ECLAC and UNFPA population models

- In our context, statistical models relate observed population data from the census to other data sets (available from administrative records or satellite imagery) in order to predict the population in areas where census information is incomplete.
- They are designed specifically for each country based on available inputs and expected objectives.
- Models can be designed to make estimates at grid level (1 km, 100 m, etc.), statistical sectors or other geographical or administrative levels, depending on the needs and the quality and quantity of information available.

Main characteristics

- Our population models have three characteristics:
  - They are Bayesian to be able to add previous information to the observed areas.
  - They are mixed to incorporate heterogeneity in unobserved areas.
  - Covariates always include satellite imagery (lights, building footprints), geospatial information (roads, infrastructure), or cartographic variables.


UGM
Viviendas
Completa
Rechazada
Interrumpida pop <> 0
No entrevistada

The Poisson GLMM for counts

We define the dwelling-level Poisson GLMM as in Berg (2022). Assume:

$$y_{ij} \mid \mu_{ij} \sim Poisson(\mu_{ij})$$
$$\mu_{ij} = N_j \, D_{ij}$$

Where y_{ij} represents the number of people in dwelling i and enumeration district j. N_j is the number of dwellings in enumeration district j. Also, D_{ij} is the average density in the dwelling and it related to the outcome through the following link function:

$$\log(D_{ij}) = \boldsymbol{x}_{ij}\boldsymbol{\beta} + u_j$$

Prior information and posterior distribution

The prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are as follows:

$$\beta_p \sim Normal\,(0{,}10000)$$
$$u_j \sim Normal\,(0{,}\sigma_u^2)$$
$$\sigma_u^2 \sim \ Inverse - Gamma(0.0001{,}0.0001)$$

Therefore, the Bayesian estimator for the number of people in dwelling i from ED j is given as

$$\tilde{\theta}_{ij} = E(y_{ij} \mid \mu_{ij})$$

The parameter of interest

The aim of the research will always be estimating the number of people in the country

$$t_y = \sum_{All\ EDs} \sum_{All\ Dwellings} y_{ij}$$

However, this parameter can be decomposed as follows:

$$t_y = \sum_{Complete\ EDs} \sum_{Complete\ Dwellings} y_{ij} + \sum_{Incomplete\ EDs} \sum_{Incomplete\ Dwellings} y_{ij}$$

Predictive approach

This way, the proposed Bayesian predictor is given by the following expression:

$$\hat{t}_y = \sum_{Complete\ EDs} \sum_{Complete\ Dwellings} y_{ij} \quad + \\ \sum_{Incomplete\ EDs} \sum_{Incomplete\ Dwellings} \tilde{\theta}_{ij}$$

This expression is similar to Molina and Rao (2010) Empirical Best Predictor in the context of poverty maps and small area estimation models.

The Multinomial GLMM for age-sex counts

We also define a municipal-level Multinomial GLMM to predict the probability of people being in each of the 40 age-sex groups (20 x 2). This way:

$$\boldsymbol{N}_d \sim Multinomial(\boldsymbol{p}_d)$$
$$\boldsymbol{p}_d = (p_{d,1,1}, \dots, p_{d,2,20})$$

Where $\boldsymbol{N}_d = (N_{d\,1\,1}, \dots, N_{d\,2\,20})'$, and $N_{d,k,l}$ represents the number of people in municipality d belonging to the sex k and age group l. Also,

$$\log\left(\frac{p_{d\,i\,j}}{p_{d\,1\,1}}\right) = \boldsymbol{z}_{dij}\boldsymbol{\gamma} + e_{dij}$$

Technical assistance in the region

- ECLAC and UNFPA join efforts have benefited the following countries in the last two years:
  - Costa Rica
  - Ecuador
  - Dominican Republic
- We are currently working with the following countries:
  - Barbados
  - Guyana
  - Jamaica

The role of covariates

Satellite Imagery (ED-level)

- We access this information trough Google Earth Engine, which provides facilities to analyze and obtain this data through the Javascript and Python programming languages, and recently since 2021 in R with the rgee package.
- Among the main advantages of information based on remote sensing is the ease of access to data with deep geographic coverage that is impossible to obtain by traditional means such as surveys or administrative records.

- *Building footprints*
- *WorldPop projections*
- *Urban cover fraction*
- *Rural cover fraction*
- *Crops_cover fraction*
- *Altitude in meters above sea level*
- *Travel time to the nearest medical center*
- *Travel time to the nearest school*

Administrative data (municipal-level)

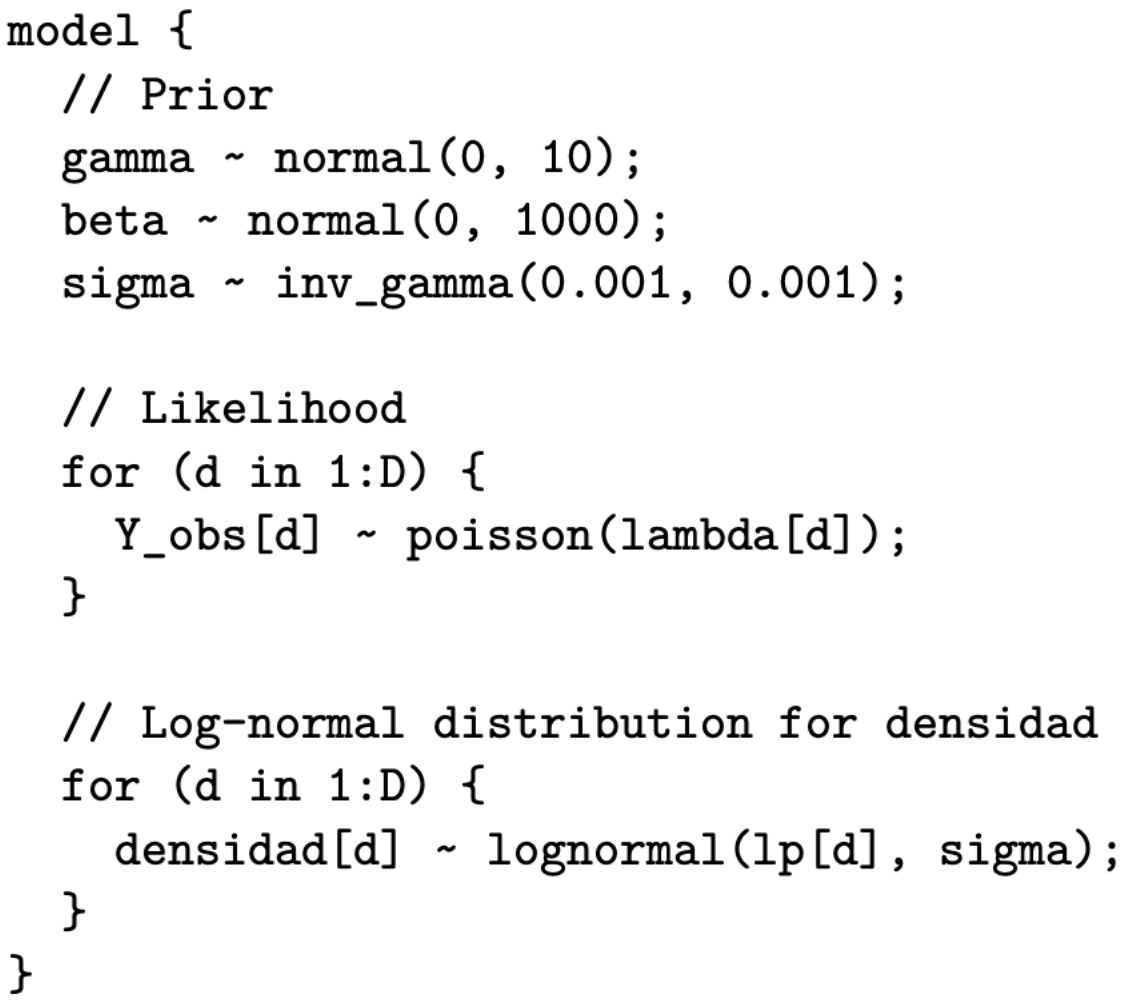
- In each country, valuable information can be found in administrative records.
- Also, we can find important covariates in the most recent census along with cartographic data available in the NSO.

- *Telecommunication access*
- *Access problems*
- *High crime rates*
- *Primary education enrollment*
- *informal settlements*
- *Indigenous area*
- *Protected area*

MCMC convergence and predictions

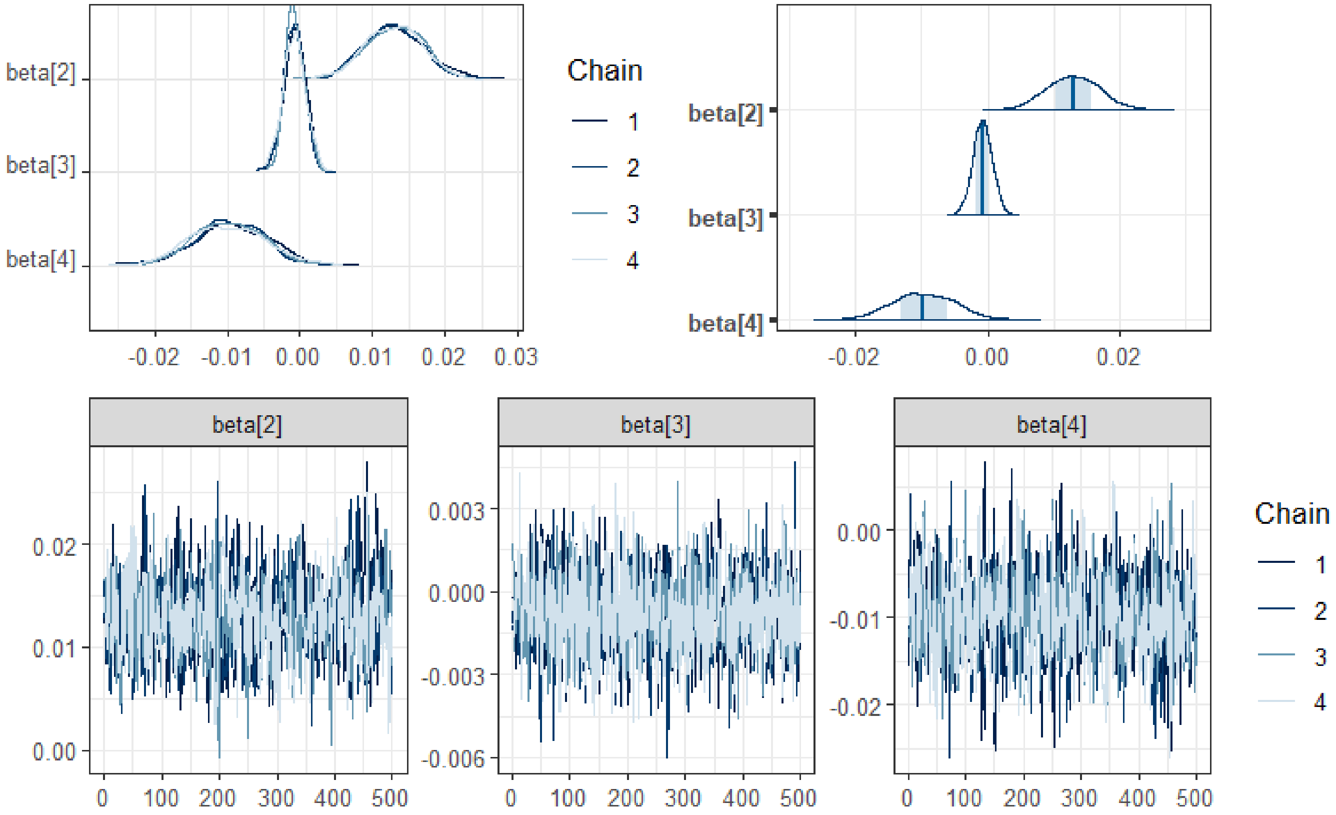
Software

- As these Bayesian computations are complex, we use our own coding in STAN.
  - STAN is an advanced Markov Chain Monte Carlo sampler that uses Hamiltonian algorithms.
  - It is easy to use and available in different platforms (Python, R, etc.)
  - It allows for computing parallelization making the process more efficient in the presence of this massive data sets.

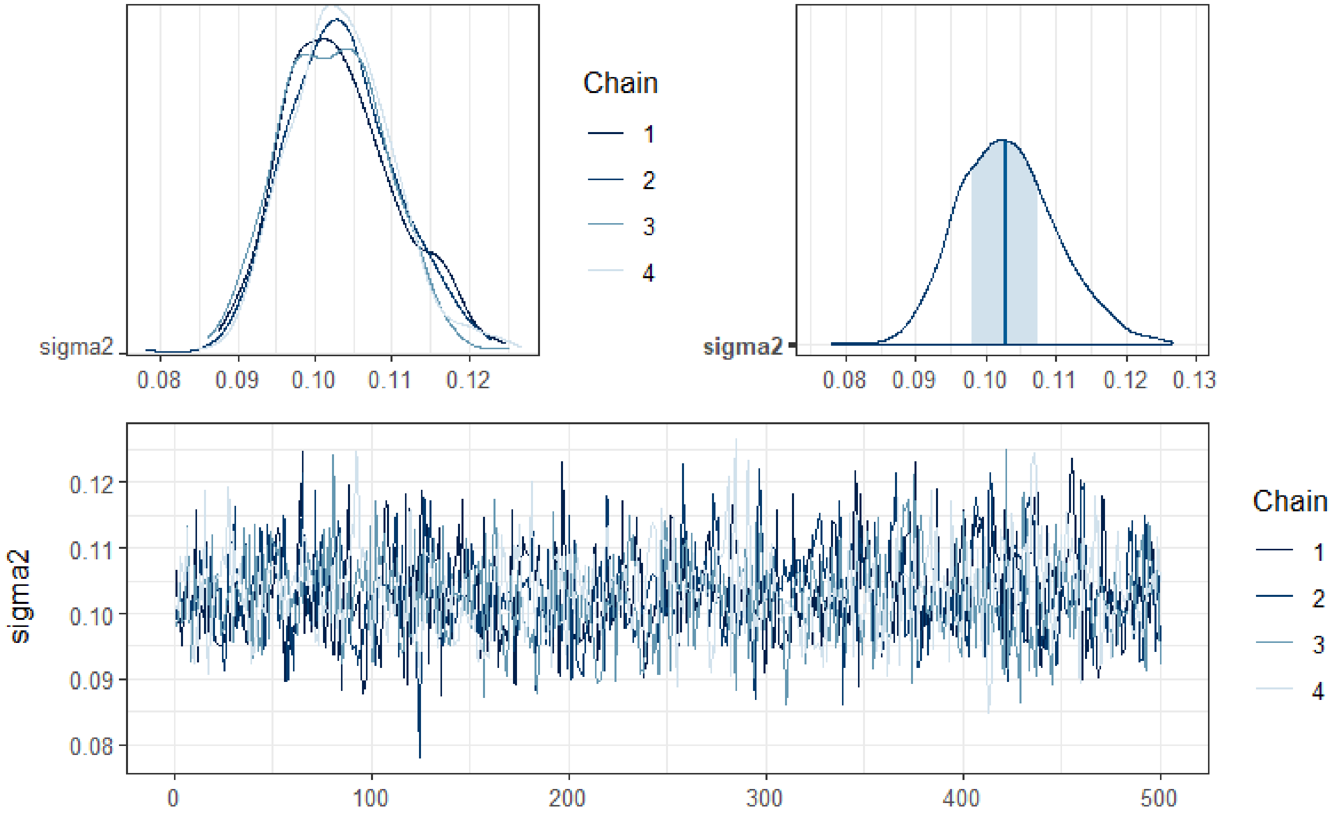


```
model {
  // Prior
  gamma ~ normal(0, 10);
  beta ~ normal(0, 1000);
  sigma ~ inv_gamma(0.001, 0.001);

  // Likelihood
  for (d in 1:D) {
    Y_obs[d] ~ poisson(lambda[d]);
  }

  // Log-normal distribution for densidad
  for (d in 1:D) {
    densidad[d] ~ lognormal(lp[d], sigma);
  }
}
```

Chaind for fixed effects coefficients

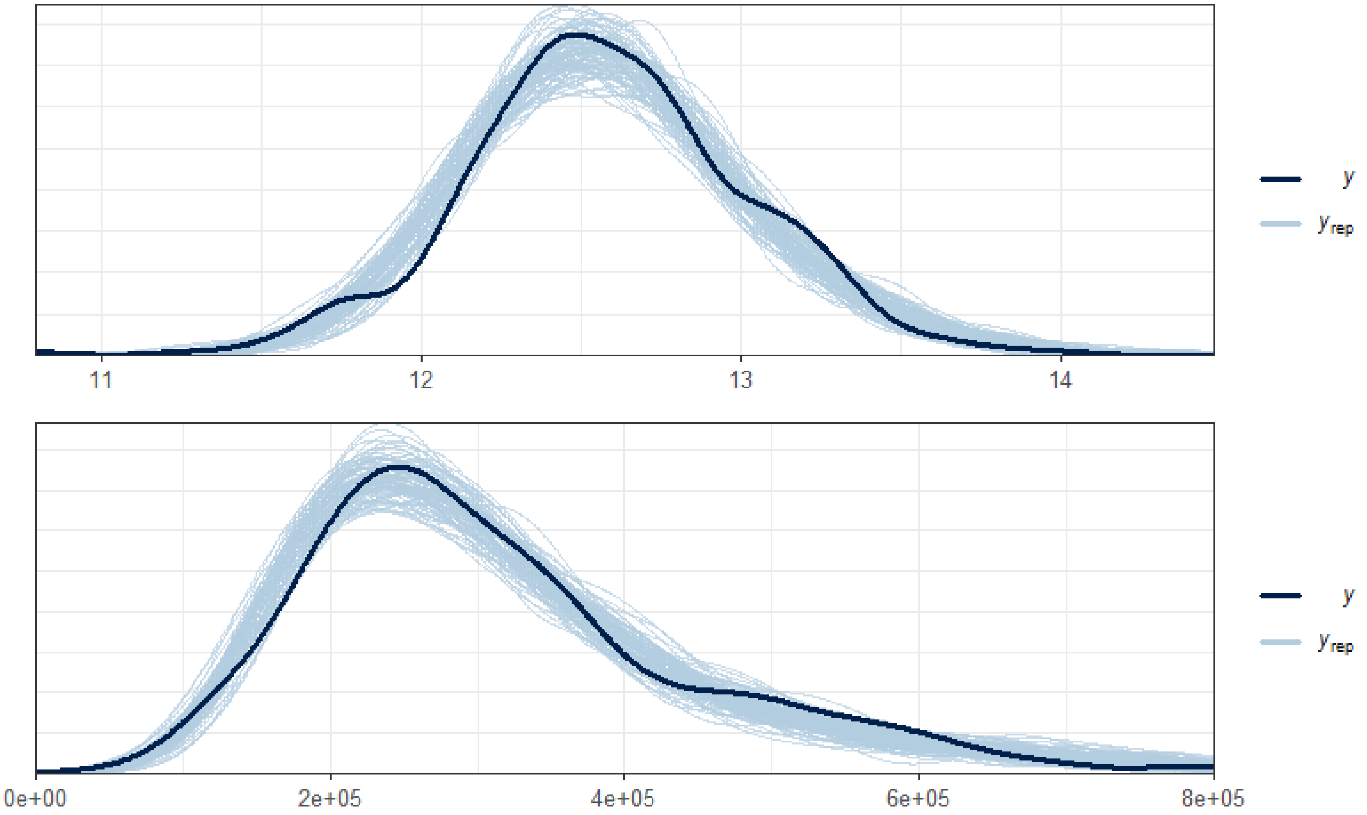


Chains for the variance of random effects

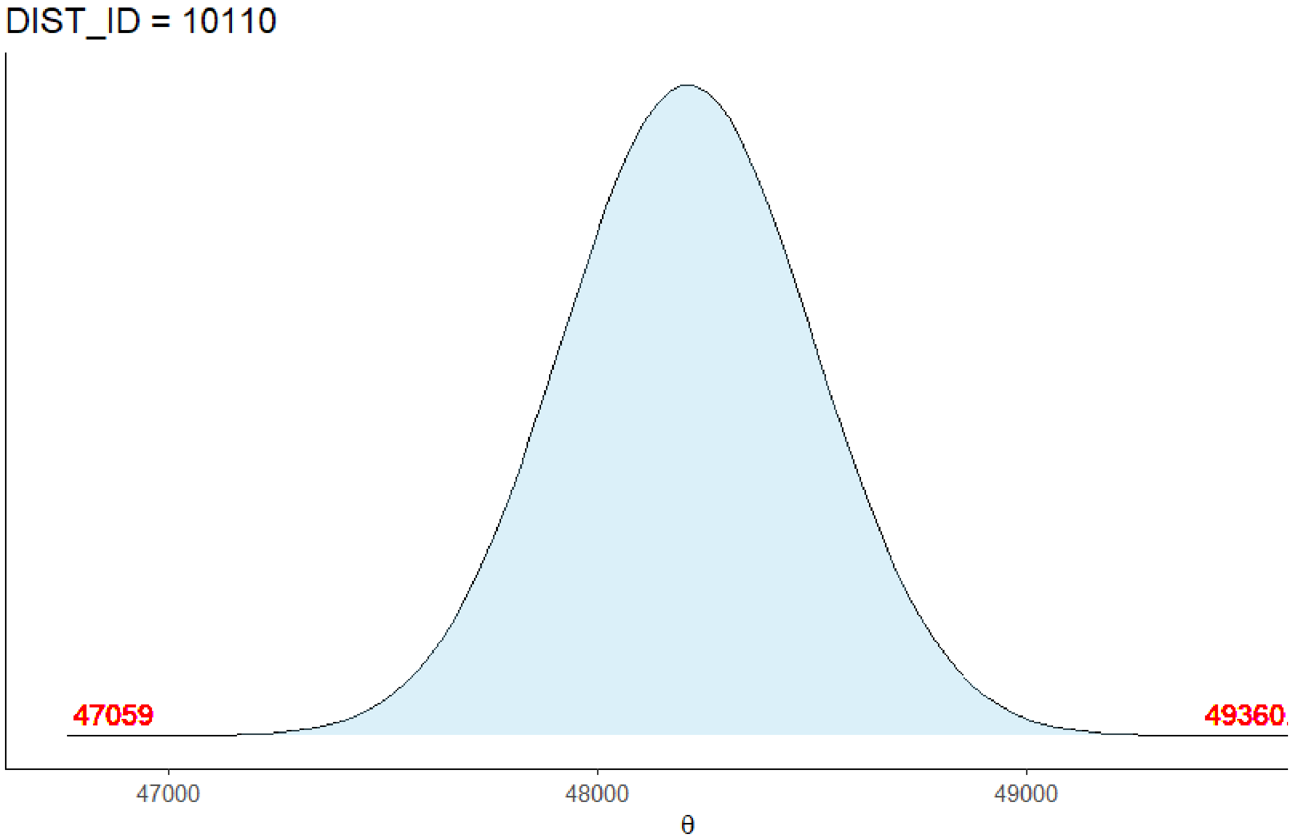


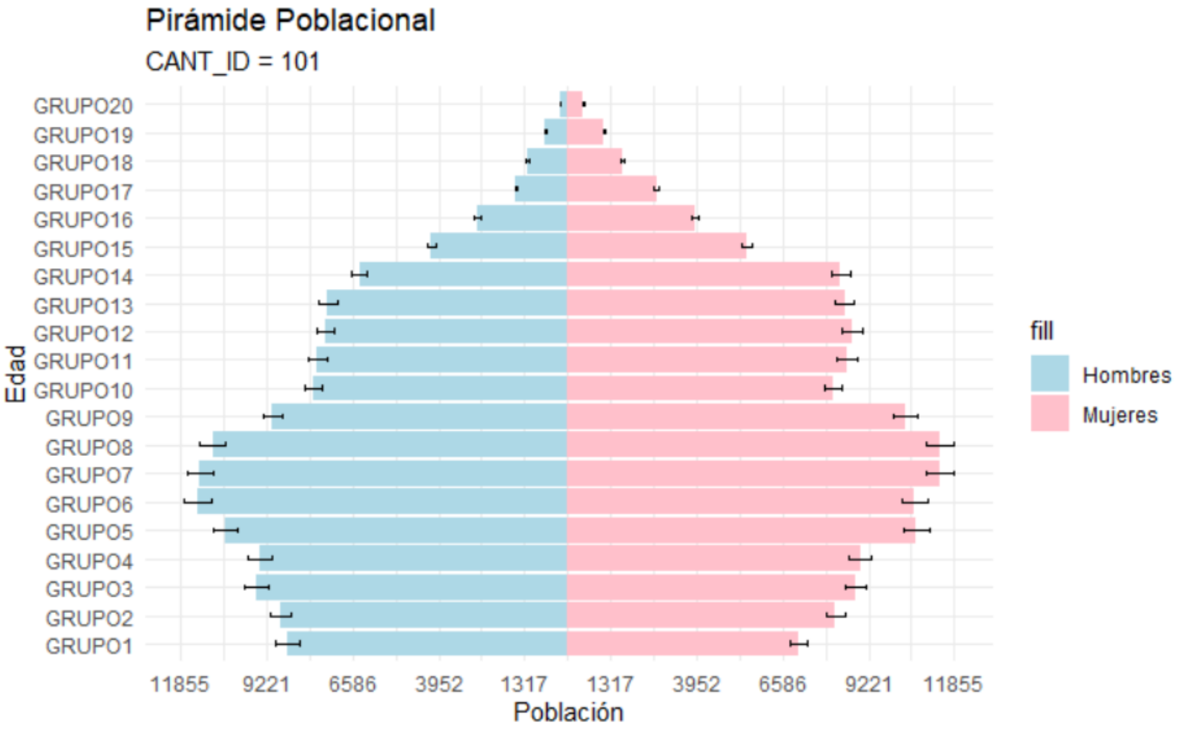
Posterior predictive checks

Log-scale and untransformed

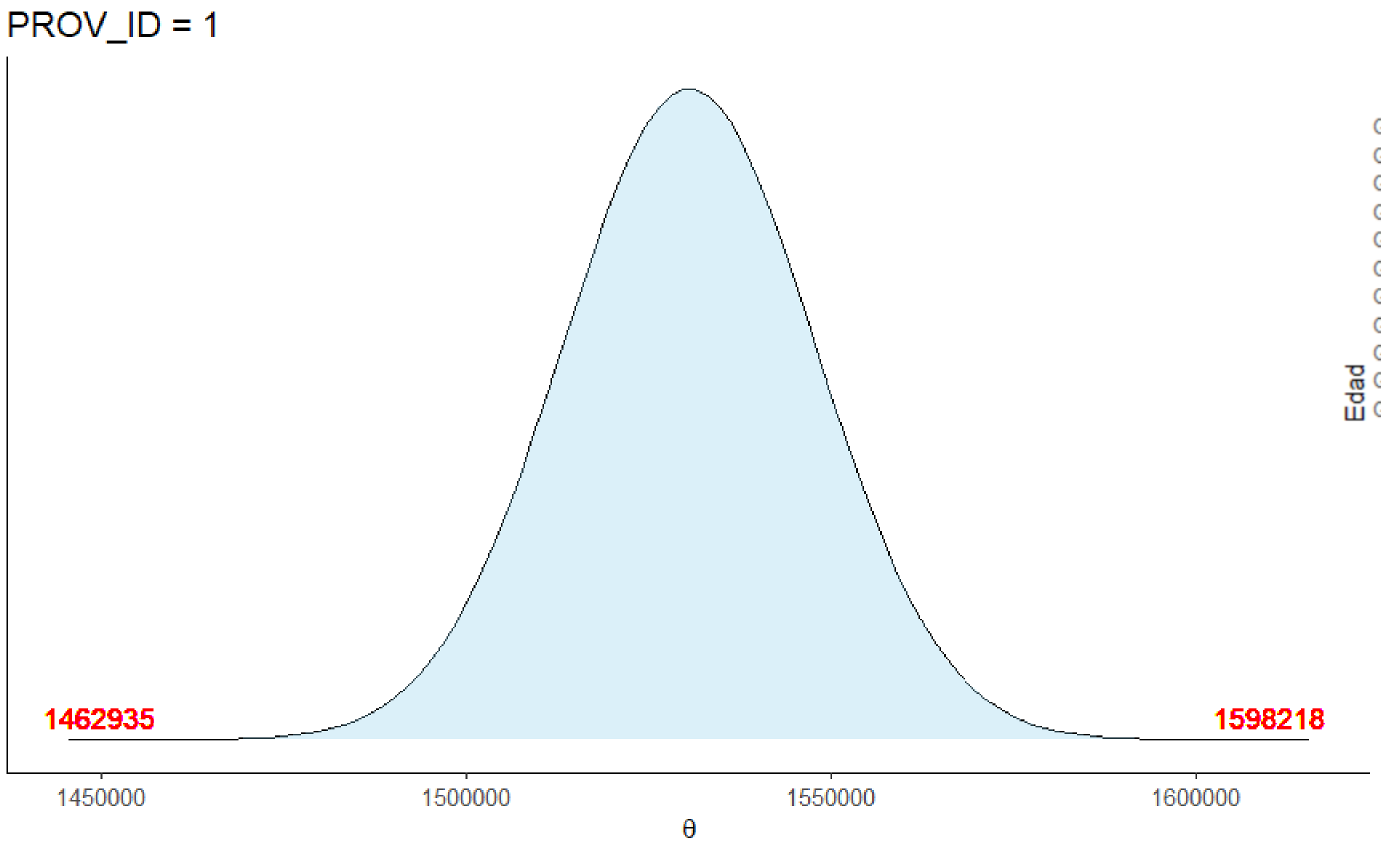


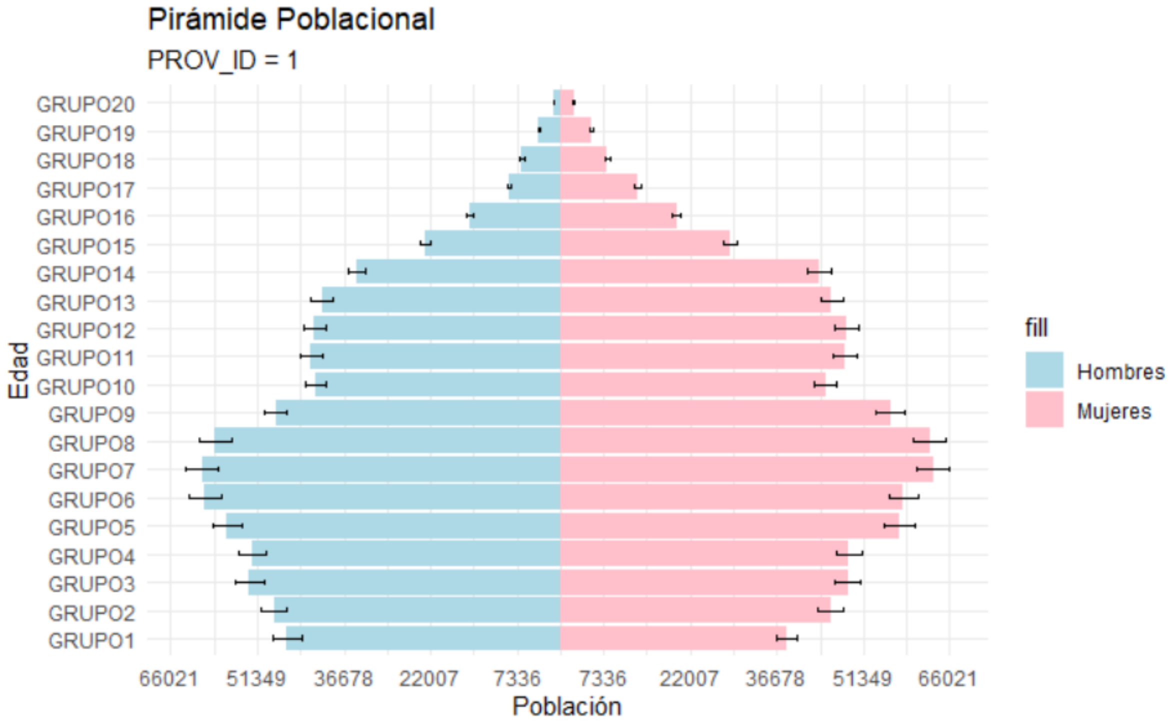
Municipal estimates



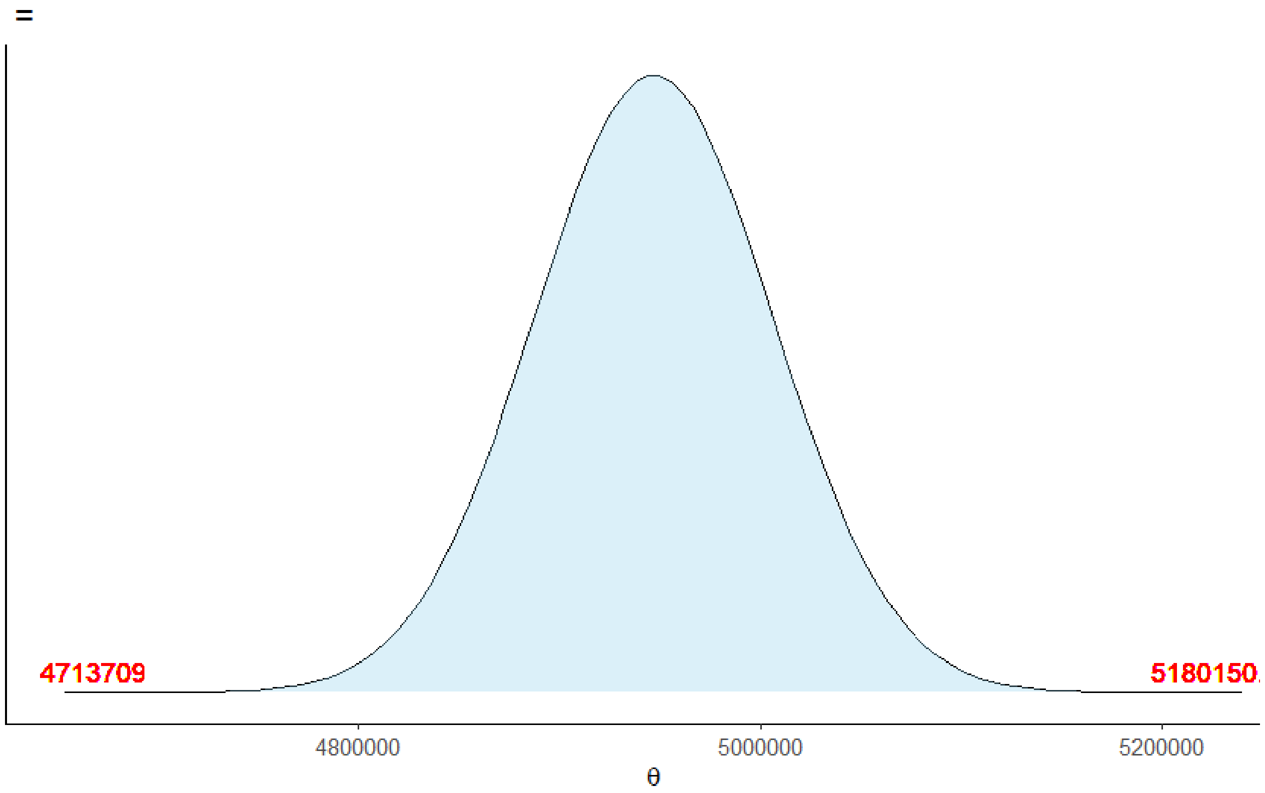


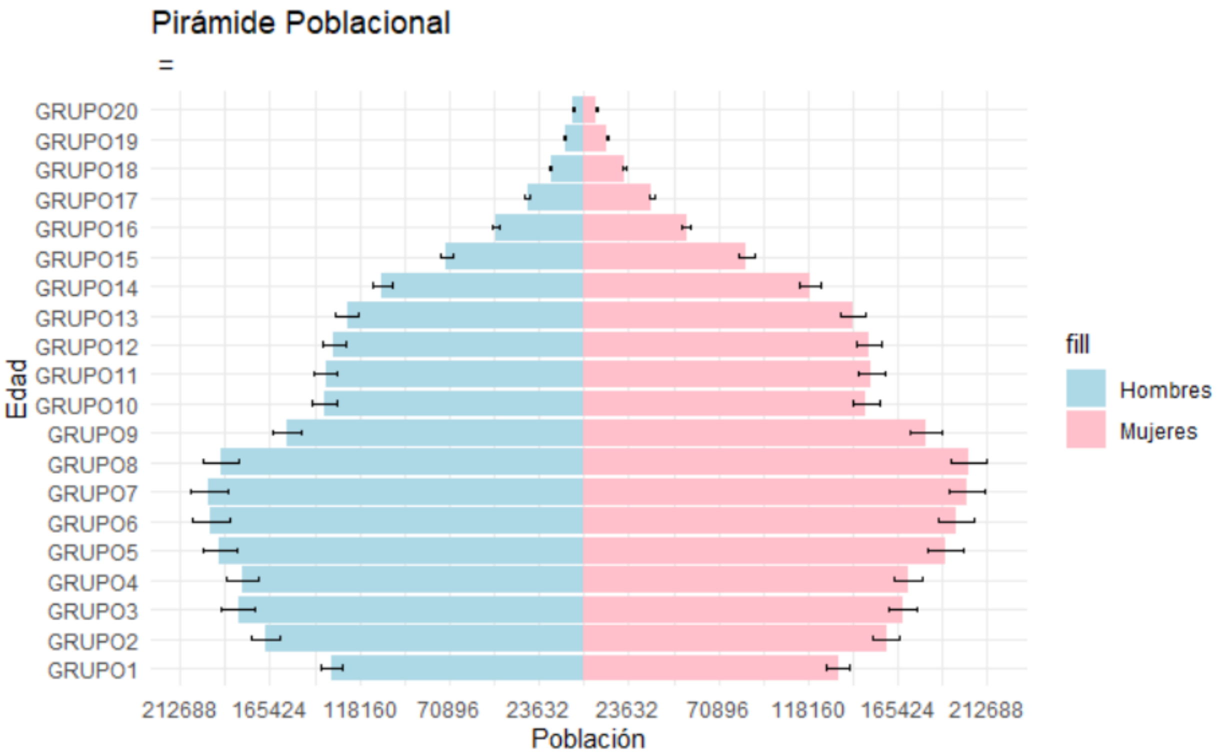
State estimates

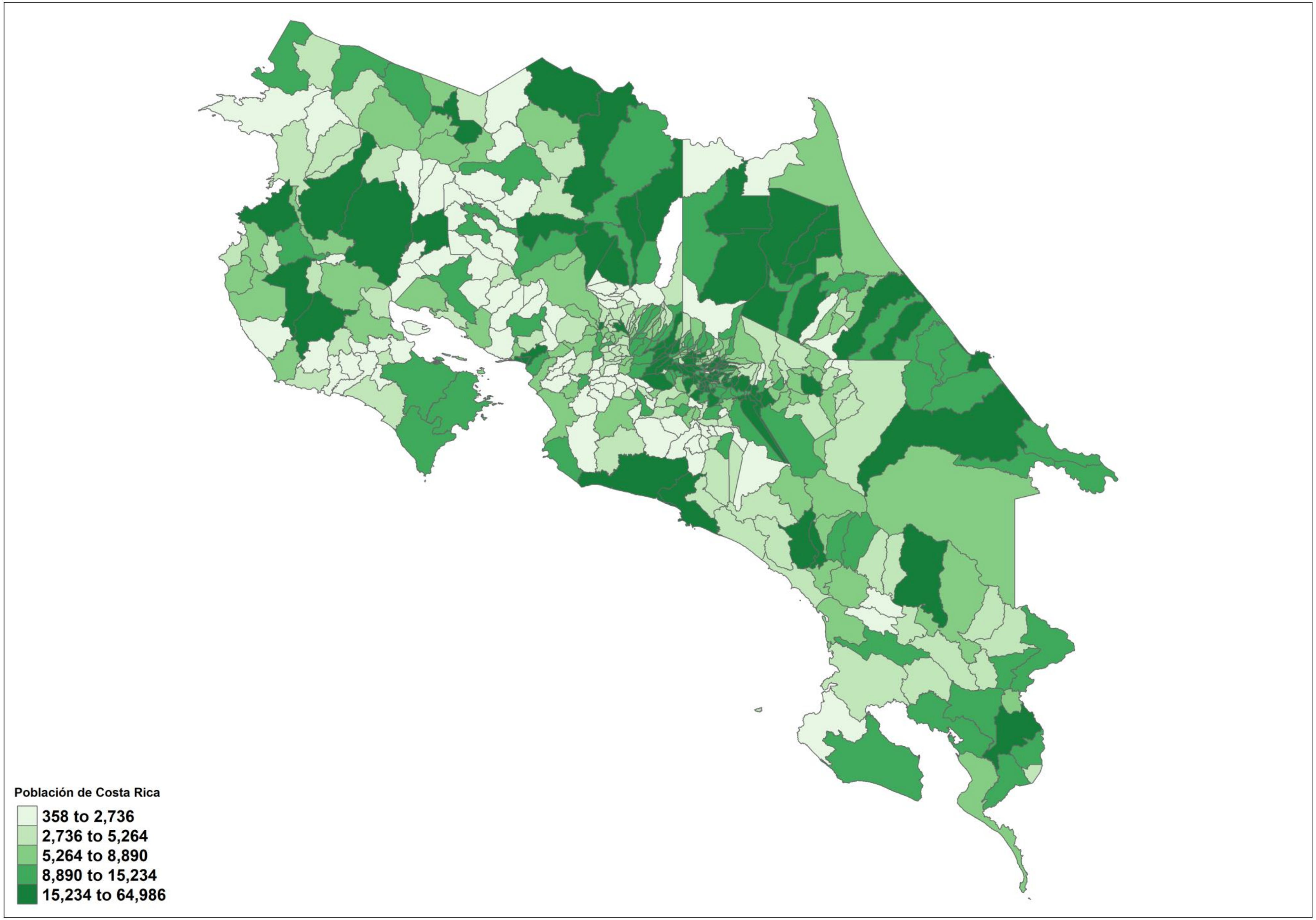




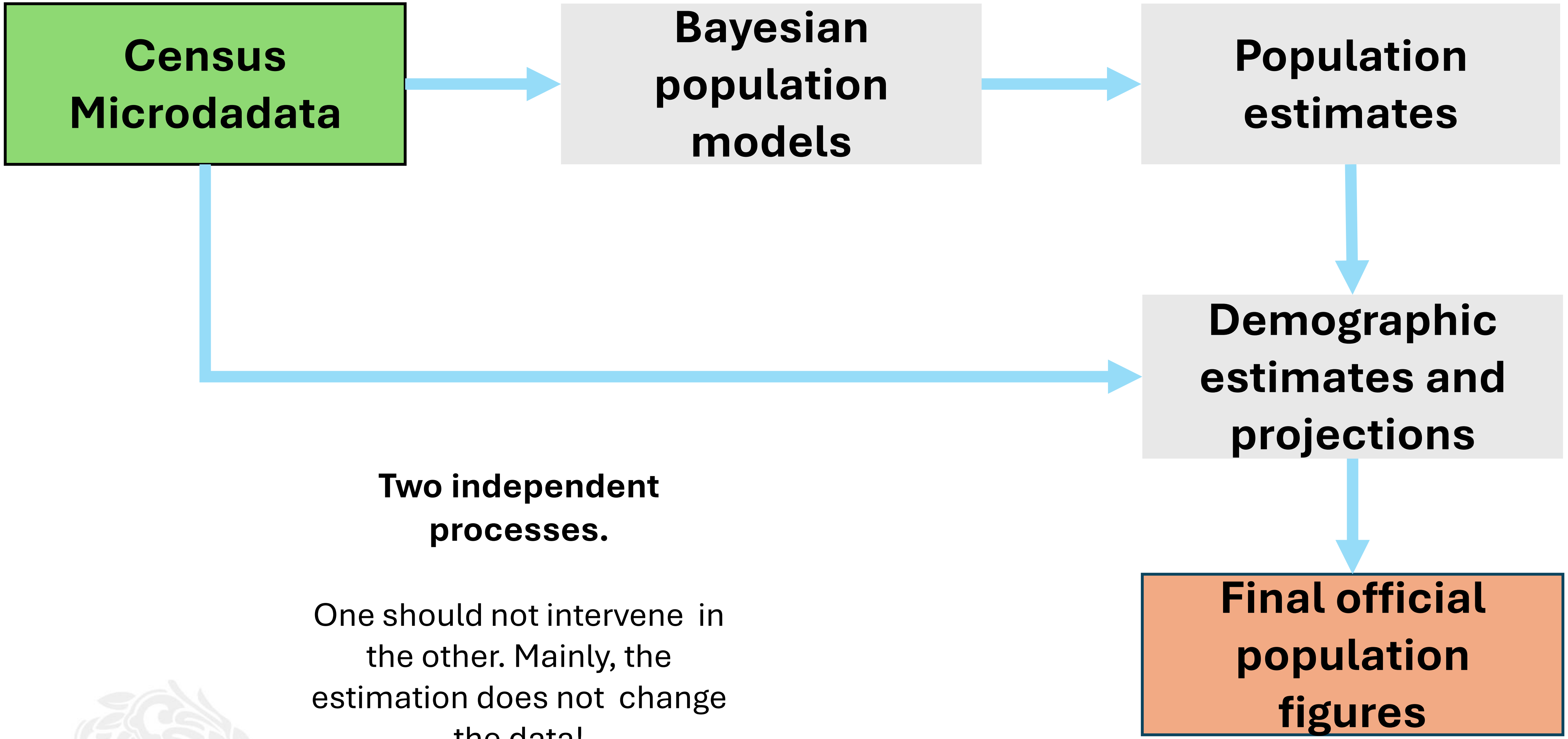
National estimates






Población de Costa Rica
358 to 2,736
2,736 to 5,264
5,264 to 8,890
8,890 to 15,234
15,234 to 64,986

One final word!



Census Microdadata

Bayesian population models

Population estimates

Demographic estimates and projections

Final official population figures

Two independent processes.

One should not intervene in the other. Mainly, the estimation does not change the data!

IAOS-ISI 2024
MEXICO CONFERENCE

Thank you

INEGI

International
Statistical
Institute

IAOS
IMPROVING OFFICIAL STATISTICS

TIME SERIES MODELS FOR REPEATED SURVEYS

DENISE SILVA

SOCIETY FOR THE DEVELOPMENT OF SCIENTIFIC RESEARCH (SCIENCE)

ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS (ENCE/IBGE)

May 2024

REPEATED SURVEYS

- Collect data at several points in time

- **Repeated sample surveys: rotating panels**
    - Some units are retained on the sample from one occasion to the next → Sample overlap between occasions
    - Inclusion/exclusion of units in the sample on distinct survey rounds

- **Panel**: set of sampling units that join and leave the sample at the same time

BRAZILIAN LABOUR FORCE SURVEY (BLFS)

- Publishes official unemployment figures since 2016
- Two-stage cluster design – census enumeration areas are PSUs and households are SSUs
- Rotating panel survey with a partially overlapping sample of households – rotation pattern 1-2(5)
- Planned sample overlap between quarters: 80% of households
- Each household is interviewed once every quarter
- National estimates are released monthly (based on rolling quarterly data), and subnational estimates are published quarterly

CONTEXT:

- Users have been calling for:
  - estimates based solely on a single-month sample
  - greater frequency of subnational releases
- Single-month estimates are needed to monitor the labour market after the COVID-19 pandemic for both national and state levels
- BLFS sample size big enough for monthly **national** estimates but not for state-level
- Alternative data sources as potential data for producing official statistics

BLFS ROTATION PATTERN

Year	Month	Panel																													
		1A	1B	1C	1D	1E	1F	1G	1H	1I	1J	1K	1L	1M	1N	1O	2A	2B	2C	2D	2E	2F	2G	2H	2I	2J	2K	2L	2M	2N	2O
t	1	3			2			1																							
t	2		3			2			1																						
t	3			3			2			1																					
t	4	4			3			2			1																				
t	5		4			3			2			1																			
t	6			4			3			2			1																		
t	7	5			4			3			2			1																	
t	8		5			4			3			2			1																
t	9			5			4			3			2			1															
t	10				5			4			3			2			1														
t	11					5			4			3			2			1													
t	12						5			4			3			2			1												
t+1	1							5			4			3			2			1											
t+1	2								5			4			3			2			1										
t+1	3									5			4			3			2			1									
t+1	4										5			4			3			2			1								
t+1	5											5			4			3			2			1							
t+1	6												5			4			3			2			1						
t+1	7													5			4			3			2			1					
t+1	8														5			4			3			2			1				
t+1	9															5			4			3			2			1			
t+1	10																5			4			3			2			1		
t+1	11																	5			4			3			2			1	
t+1	12																		5			4			3			2			1

THE BLFS TIME SERIES

- Time series of a repeated survey with sample overlap
- Rotation pattern affects the correlation between survey estimates over time
- Observed series are subjected to sampling errors
- The sampling errors are correlated over time due to sample overlap

In the beginning....

Blight and Scott (1973); Scott and Smith (1974);

Scott, Smith and Jones(1977)

Binder and Hidiroglou (1988); Binder and Dick (1989) ; Tiller (1989); Pferffermann, Burck and Ben-Tuvia (1989)

MODELLING TIME SERIES FROM REPEATED SURVEYS

Usual Approach $\hat{y}_t = T_t + S_t + I_t$

Standard time series procedures fail to account for the effect of the sampling error autocorrelation

Signal Extraction $\hat{y}_t = \theta_t + e_t$

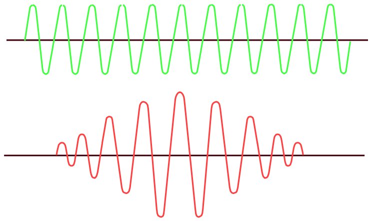
$$\theta_t = T_t + S_t + I_t$$

$\hat{y}_t$ is the design unbiased survey estimate

θ_t is the unknown population quantity

e_t is the survey error

SIGNAL EXTRACTION APPROACH: COMBINES TWO MODELS



- One to describe the evolution of the unobservable population quantity $\{\theta_t\}$ over time (signal)
- One to represent the time series relationship between the sampling errors $\{e_t\}$ of the survey estimators (noise)

MODELLING PROCEDURE

- Formulate time series models for the signal θ_t and the noise e_t
- Combine the models using a state-space formulation
- The models contemplate the survey design
- Estimate unobserved model components using the Kalman Filter

In the beginning….

Binder and Dick (1990); Pferffermann (1991); Tiller (1992);
Pferffermann and Bleuer (1993);
Pfeffermann, Feder and Signorelli (1998)

TIME SERIES MODEL-BASED ESTIMATORS

Flexibility to meet historical demands and new challenges

- Estimation of trend and seasonality
- Production of labour force indicators based solely on the cases surveyed in the reference month (instead of rolling quarters)
- Production of small area estimates
- Estimation of the effect of the SARS-COV-2 pandemic (higher volatility)
- Incorporation of auxiliary and alternative data sources, such as big data
- Nowcasting

STATE-SPACE MODEL FOR UNEMPLOYMENT RATE ($\hat{y}_t$)

$\hat{y}_t$: design-based estimate at month t

Signal extraction: $\hat{y}_t = \theta_t + e_t$

Unobserved components of unknown population quantity : $\theta_t = T_t + S_t + I_t \qquad I_t \sim N(0, \sigma_I^2)$

Trend: $T_t = T_{t-1} + R_{t-1}$

Seasonal Component

$$R_t = R_{t-1} + \eta_{R,t} \qquad S_t = \sum_{l=1}^{\frac{s}{2}=6} S_{l,t} + \eta_{S,t}$$

$$\eta_{R,t} \sim N(0, k_t \sigma_R^2) \qquad \eta_{S,t} \sim N(0, \sigma_S^2)$$

k_t fator increases the variance for more flexibility in trend

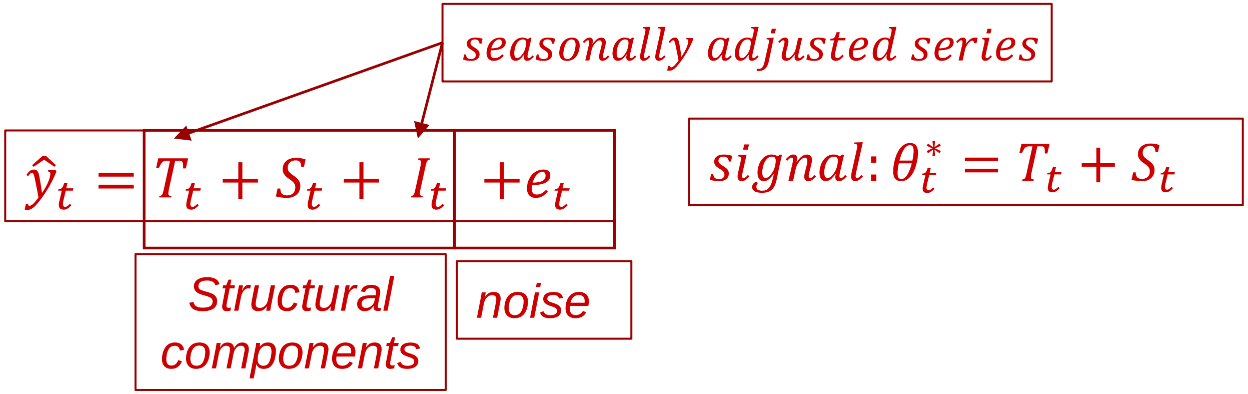
MODEL FOR BLFS SAMPLING ERROR (e_t)

Sampling error e_t: $$e_t = \hat{c}_t \tilde{e}_t$$

$\hat{c}_t$: standard error of design-based estimates

$$\tilde{e}_t = \phi \tilde{e}_{t-3} + \eta_{\tilde{e},t}, \qquad \eta_{\tilde{e}} \sim N\left(0, \sigma^2_{\tilde{e}}\right)$$

Each household is interviewed once every quarter



$\mathit{seasonally\ adjusted\ series}$

$$\hat{y}_t = T_t + S_t + I_t + e_t$$

signal: $\theta^*_t = T_t + S_t$

Structural components

noise

REGIONAL MULTIVARIATE MODEL

$\hat{y}_{j,t}$: design-based estimate for unemployment rate at month t in the state j

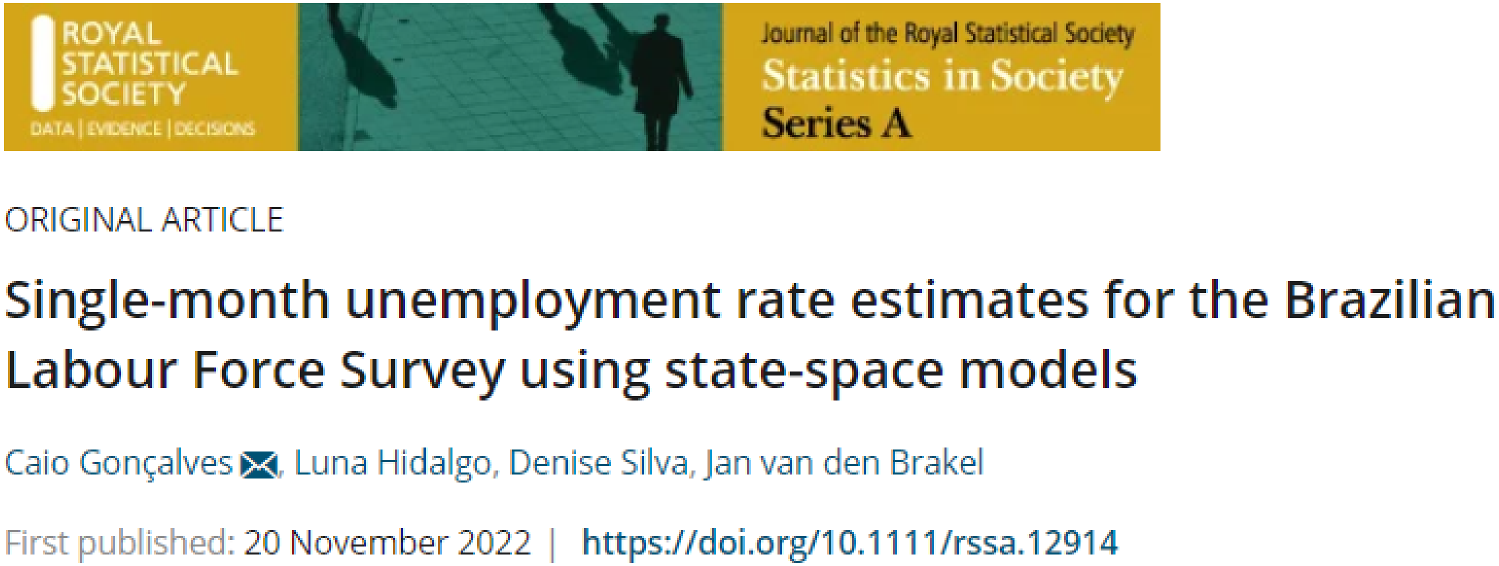
$$\begin{pmatrix} \hat{y}_{1,t} \\ \vdots \\ \hat{y}_{J,t} \end{pmatrix} = \begin{pmatrix} \theta_{1,t} \\ \vdots \\ \theta_{J,t} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ \vdots \\ e_{J,t} \end{pmatrix} \qquad j = 1, \dots, J$$

Borrowing strength from the time and space:

$$cov\left(\eta_{R,y_j,t}, \eta_{R,y_{j'},t}\right) = \rho^{R}_{y_j,y_{j'}}\ \sigma_{R,y_j,t}\ \sigma_{R,y_{j'},t} \qquad j \neq j'$$

$\rho^{R}_{y_j,y_{j'}}$ is the correlation between the slope disturbance terms of the unemployment rate of states j and j'

MODEL-BASED SINGLE MONTH ESTIMATES



ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

ORIGINAL ARTICLE

Single-month unemployment rate estimates for the Brazilian Labour Force Survey using state-space models

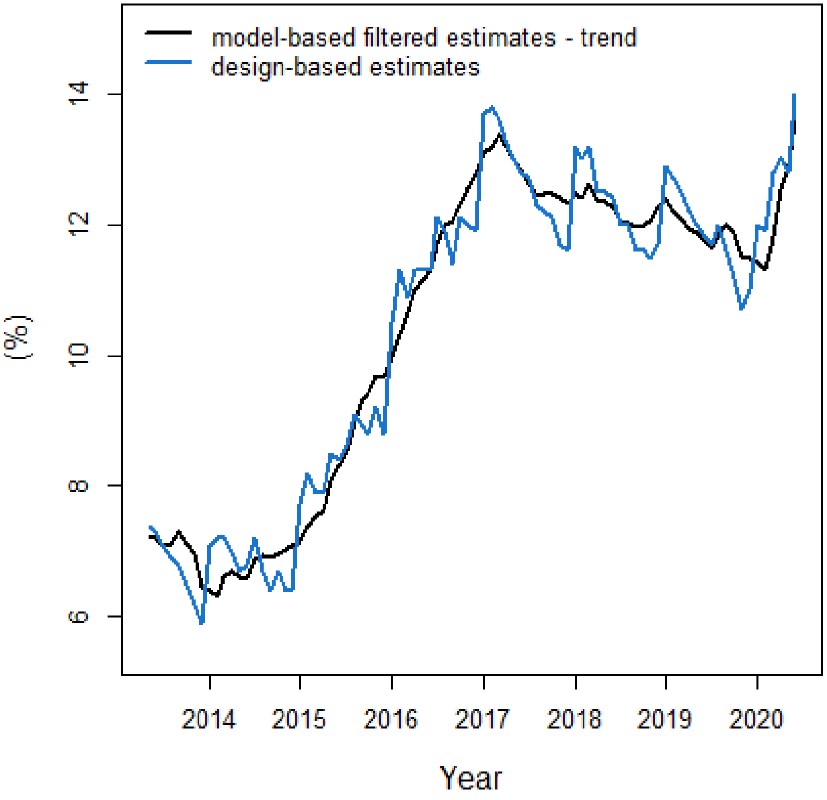
Caio Gonçalves, Luna Hidalgo, Denise Silva, Jan van den Brakel

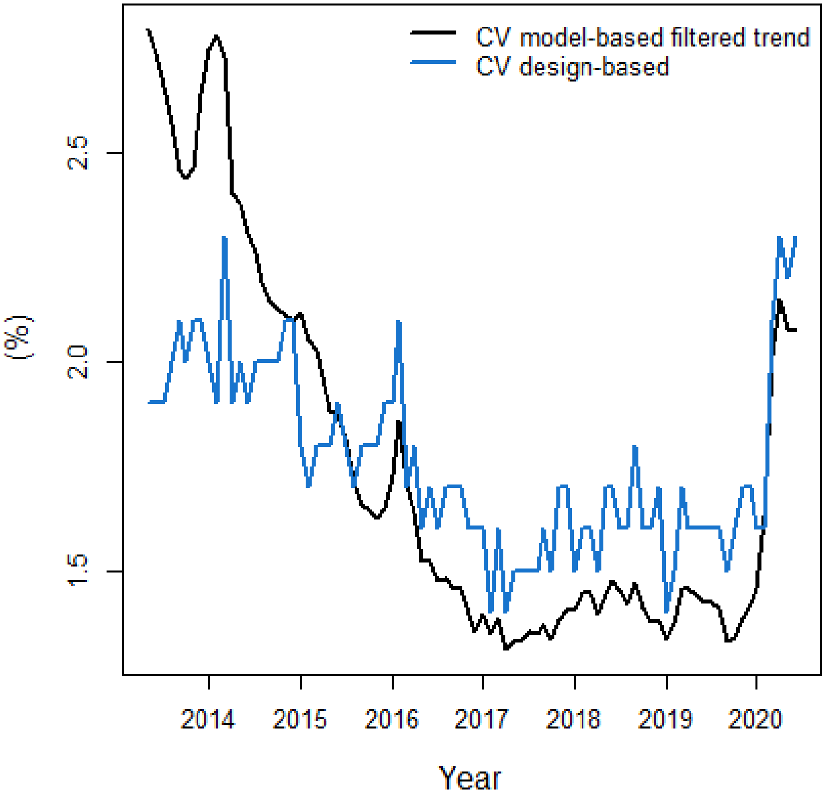
First published: 20 November 2022 | https://doi.org/10.1111/rssa.12914

Users have been calling for estimates based solely on a single-month sample, and for a greater frequency of subnational releases

RESULTS

Unemployment rate design-based and model-based (trend) estimates, and coefficients of variation – Brazil

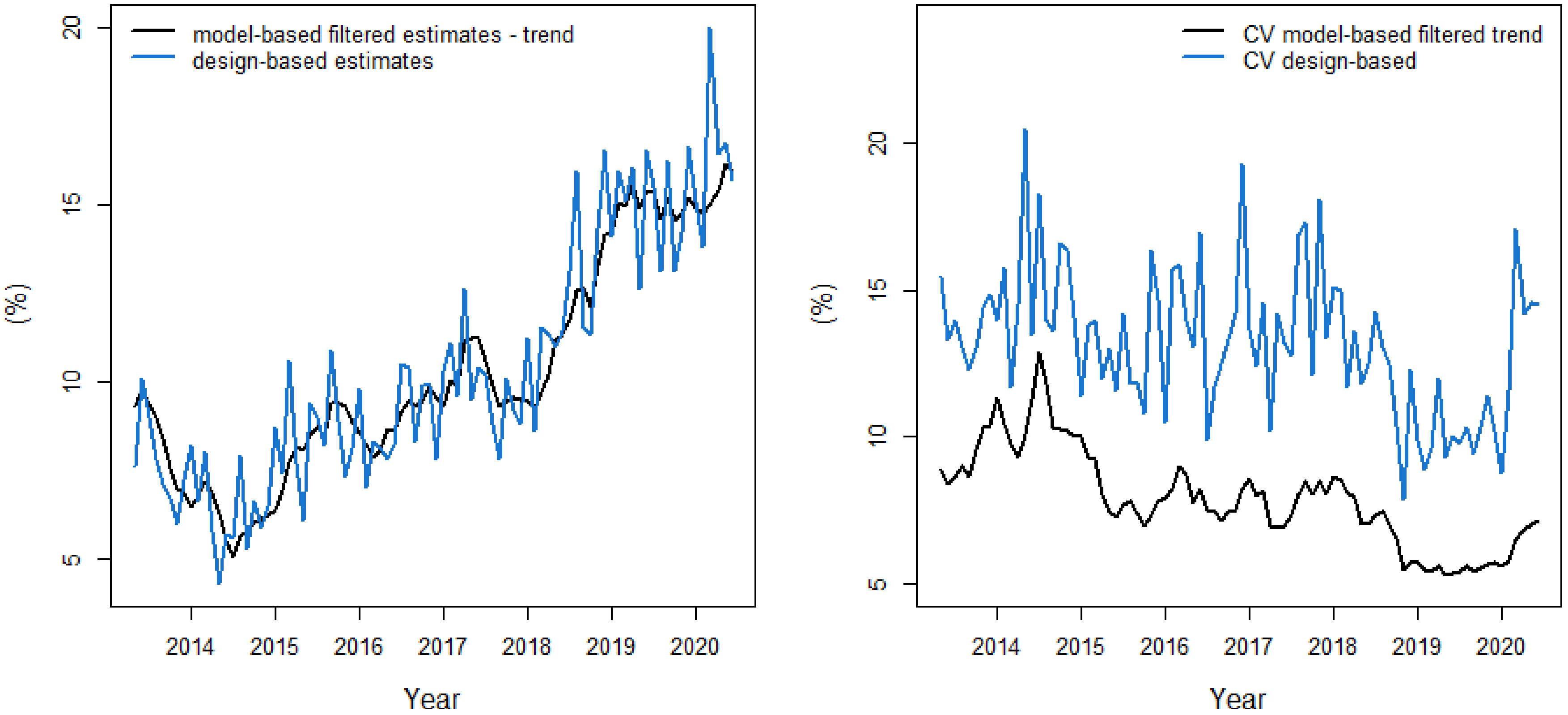




Single month sample size ~ 60,000 households

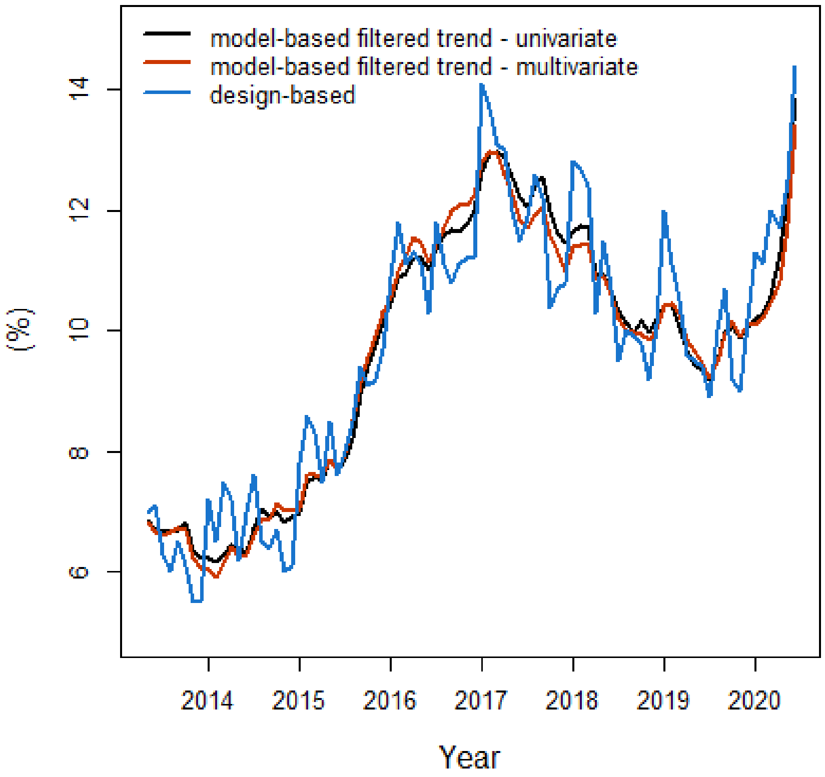
RESULTS

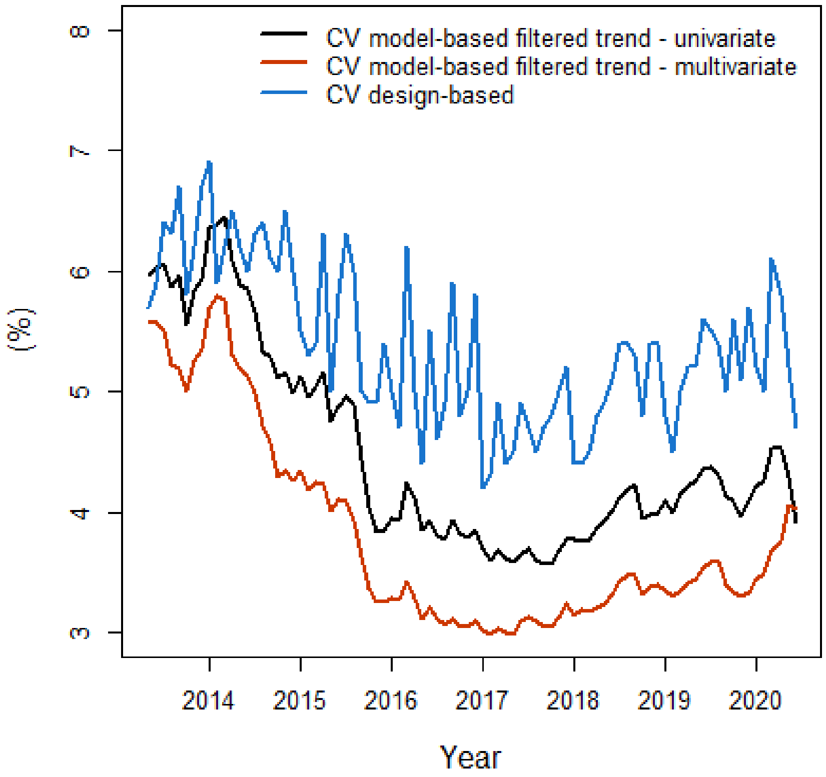
Unemployment rate design-based and model-based (trend) estimates, and coefficients of variation – Roraima



Single month sample size ~ 1,000 households

Design-based, trend model-based (univariate and multivariate) estimates, and coefficients of variation Minas Gerais





Correlation matrix of slope disturbance terms - Southeast region

States	Minas Gerais	Espírito Santo	Rio de Janeiro	São Paulo
Minas Gerais	1			
Espírito Santo	0.9520	1		
Rio de Janeiro	0.5436	0.7677	1	
São Paulo	0.8162	0.8581	0.7587	1

REMARKS ON MODEL-BASED ESTIMATES

National level

- Precision is already adequate for publication using design-based
- Model-based approach produces trend and seasonally adjusted series taking into account the sampling error
- The advantages of abandoning the use of three-month rolling estimates are wide

State level

- Opportunity to produce monthly estimates for states with acceptable precision
- Regional multivariate models presented advantages

TIME SERIES MODELS FOR SMALL AREA ESTIMATION

Small Area Quarterly Unemployment Estimates for The Brazilian Labour Force Survey Using State-space Models

OTTAWA 2023
64TH WORLD STATISTICS CONGRESS

$\hat{y}_{j,t}$: Direct estimate of the total number of unemployed for quarter t in area j.

Signal extraction: $\hat{y}_{j,t} = \theta_{j,t} + e_{j,t}$

Unobservable Components of the unknown population quantity $\theta_{t,j}$:

$$\theta_{t,j} = T_{t,j} + S_{t,j} + I_{t,j} \qquad I_{t,j} \sim N(0, \sigma^2_{I,j})$$

MULTIVARIATE TIME SERIES MODELS TO INTEGRATE SURVEY DATA WITH BIG DATA

Model-based Single-month Unemployment Estimates for the Brazilian Labour Force Survey Incorporating Google Trends Data

- Common trend models to combine survey data and Google Trends time series

$$\begin{pmatrix}\hat{y}_t\\ \boldsymbol{x}_t\end{pmatrix} = \begin{pmatrix}\theta_t\\ \hat{\boldsymbol{\Lambda}}\,\boldsymbol{f}_t\end{pmatrix} + \begin{pmatrix}e_t\\ \boldsymbol{u}_t\end{pmatrix}$$

- $\boldsymbol{x}_t$: Google Trends series
- $\boldsymbol{f}_t$: factors obtained via principal components analysis of $\boldsymbol{x}_t$

COMPOSITIONAL TIME SERIES OF LABOUR FORCE STATUS

Proportion of Unemployed People y_{1t}

Proportion of Employed People y_{2t}

Proportion of Economically Inactive y_{3t}

Unemployment Rate $\frac{y_{1t}}{y_{1t}+y_{2t}}$

$$\sum_{m=1}^{3} y_{m,t} = 1 \qquad 0 \leq y_{m,t} \leq 1$$

Vector of proportions subjected to unity-sum constraint

- Multivariate time series comprising observations of compositions at each time point
- Variables have to be modelled concurrently satisfying a unity-sum constraint
- Use multivariate state space models taking into account the sampling errors

COMPOSITIONAL TIME SERIES OF LABOUR FORCE STATUS

Considering the 3 series simultaneously :

$$\boxed{\underline{y}_t = \underline{\theta}_t + \underline{e}_t}$$

$$\underline{y}_t = (y_{1,t}, y_{2,t}, y_{3,t})'$$

$$\underline{\theta}_t = (\theta_{1,t}, \theta_{2,t}, \theta_{3,t})'$$

$$\underline{e}_t = (e_{1,t}, e_{2,t}, e_{3,t})'$$

$$\sum_{m=1}^{3} y_m = \sum_{m=1}^{3} \theta_m = 1$$

Compositional Time Series Analysis of Labour Force Status in the Brazilian National Household Survey

Denise Silva [1] *, Eduardo Rosseti* [2] *and Antônio Teixeira Júnior* [3]

1. Escola Nacional de Ciências Estatísticas – ENCE/IBGE
2. FGV Projetos
3. Senac Departamento Nacional

Conference in honour of Fred Smith and Chris Skinner
Online, 7-9 July 2021

ENCE Instituto Brasileiro de Geografia e Estatística IBGE IBGE

CONCLUDING REMARKS

- Flexible and powerful method to produce reliable and precise model-based estimates
- Models can borrow strength from different survey occasions (in time), areas, as well as incorporate administrative or alternative data sources
- Models already tested and/or implemented by BLS-US, Statistics Netherlands, Statistics Canada, ONS-UK, ABS

REFERENCES

Binder, R. and Dick, J. (1989) Modelling and estimation for repeated surveys. *Survey Methodology*, 15, 29–45.

Boonstra, H. J. and van den Brakel, J. A. (2019) Estimation of level and change for unemployment using structural time series models. *Survey Methodology*, 45, 395–425.

Durbin, J.; Koopman S. J. Time Series Analysis by State Space Methods. 2. ed. Oxford: Oxford University Press, 2012.

van den Brakel, J. A. and Krieg, S. (2009) Estimation of the monthly unemployment rate through structural time series modelling in a rotating panel design. *Survey Methodology*, 35, 177–190.

Pfeffermann, D. and Tiller, R. (2006) Small area estimation with state space models subject to benchmark constraints. *J. Am. Statist. Ass.*, 101, 1387–1397.

Scott, A. J. and Smith, T. M. F. (1974) Analysis of repeated surveys using time series methods. *J. Am. Statist. Ass.*, 69, 674–678.

Scott, A. J., Smith, T. M. F. and Jones, R. G. (1977) The application of time series methods to the analysis of repeated surveys. Int. Statist. Rev., 45, 13–28.

Rosseti, E. S. & Silva, D. B. N (2017). Modelos de séries temporais para pesquisas amostrais repetidas. In: Albieri, S. & Dias, A. J. R. (Org.). *40 Anos da Unidade de Métodos Estatísticos do IBGE Alguns Passos*. Documentos para Disseminação - Memória Institucional. Rio de Janeiro: IBGE, 22, p. 171-188.

Schiavoni, C., Palm, F., Smeekes, S., & van den Brakel, J. (2019). *A dynamic factor model approach to incorporate Big Data in state space models for official statistics*. arXiv.org at Cornell University Library.arXiv e-prints, No. 1901.11355

Silva, D.B.N. and Smith, T.M.F, (2001), " Modelling Compositional Time Series from Repeated Surveys", Survey methodology, Vol. 27, No. 2, pp 205-215.

Tiller, R.B (1992), "Time Series Modeling of Sample Survey Data From the U.S. Current Population Survey," Journal of Official Statistics, 8, pp149-166