

UNITED NATIONS



Producing Small Area Estimates for Labor Market Indicators in Latin America **A Bayesian Perspective**

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

Statistics Division UN-ECLAC



The need for disaggregations in labour force surveys

An SDG perspective





SUSTAINABLE GENALS

DECENT WORK AND ECONOMIC GROWTH

Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.

Target 8.3. Promote policies to support job creation and growing enterprises.

- Promote development-oriented policies that support productive activities, decent job creation, entrepreneurship, creativity and innovation, and enterprises, including through access to financial services.
 - by sector and sex.

encourage the formalization and growth of micro-, small- and medium-sized

Indicator 8.3.1: Proportion of informal employment in total employment,

Target 8.5. Full employment and decent work with equal pay.

- By 2030, achieve full and productive employment and decent work for all • women and men, including for young people and persons with disabilities, and equal pay for work of equal value.
 - Indicator 8.5.1: Average hourly earnings of employees, by sex, age, occupation and persons with disabilities
 - Indicator 8.5.2: Unemployment rate, by sex, age and persons with disabilities.



Target 8.6. Promote youth employment, education and training.

- By 2020, substantially reduce the proportion of youth not in employment, • education or training.
 - Indicator 8.6.1: Proportion of youth (aged 15–24 years) not in education, employment or training.

Target 8.7. End modern slavery, trafficking and child labor.

- Take immediate and effective measures to eradicate forced labour, end modern slavery and human trafficking and secure the prohibition and child soldiers, and by 2025 end child labour in all its forms.
 - engaged in child labour, by sex and age.

elimination of the worst forms of child labour, including recruitment and use of

Indicator 8.7.1: Proportion and number of children aged 5–17 years



Fundamental principle of data disaggregation

Sustainable Development Goal indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migration status, disability and geographic location, or other characteristics, in accordance with the Fundamental Principles of Official Statistics.

General Assembly resolution - 68/261

Fundamental principles of official statistics

The public's essential confidence in the integrity of official statistical systems and the credibility it gives to statistics depend to a large extent on respect for the fundamental values and principles that underpin any society that seeks to understand itself and respect the rights of its members and which, In this context, the professional independence and accountability of statistical agencies are crucial.

General Assembly resolution - 68/261

development.

PARNERSHPS EORIEGOALS

Strengthen the means of implementation and revitalize the global partnership for sustainable

SDG 17: Partnership for the goals

Target 17.1. Enhance availability of reliable data

• significantly the availability of high-quality, timely and reliable data contexts.

By 2020, enhance capacity-building support to developing countries, including for least developed countries and small island developing States, to increase disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national





What is it about?

Surveys depend on large sample size and a proper sampling strategy (sampling design and estimator); they also rely on a robust inferential system that provides precise and exact estimation in planned domains.

When the sample size of the survey is not enough for some subgroups of interest, it is necessary to resort to external auxiliary information (censuses, administrative records, satellite imagery) so that together (surveys and external data) a precise inferential system can be built.

UN-ECLAC uses SAE models to integrate data from different sources.



Coefficients of variation in different disaggregations. Source: NSO - Chile

What is a small area?

Estimates of the parameters of interest might be necessary for geographical disaggregation, allowing visualization on a map, or for sociodemographic subgroups.

- municipalities.
- Specific subgroups might involve combinations such as age x sex x ethnicity × immigration status.

In general, if the subgroups are not part of the survey design domains, their sample size is not planned in advance and will therefore be random, which increases the uncertainty of the direct estimate.

Geographical units could include states, provinces, departments, or

Parsimonious solution

The general idea is to support estimation within existing relationships in other areas by fitting statistical models that establish a connection between subgroups of interest through complementary information present in censuses, administrative records, or remote sensing data sources.

Confronted with the challenge of producing accurate estimates in small domains, models emerge as an alternative for estimating indicators of interest, even in domains where the sample size is small or nonexistent.



SAE models for labor market indicators: A non-comprehesnive review

Molina, Saei and Lombardía (2007) Small Area Estimates of Labour Force Participation under a Multinomial Logit Mixed Model

They proposed a methodology for estimating unemployment or employment characteristics in small areas, based on the assumption that the sample totals of unemployed and employed individuals follow a multinomial logit model with random area effects.

They considered the multinomial vector that counts the number of sampled sexage group.

$$\log\left(\frac{p_{dij}}{p_{di3}}\right) = \boldsymbol{x}_{di}\boldsymbol{\beta}_j + u_d$$

unemployed (y_1) , employed (y_2) , and inactive (y_3) individuals within each area-

j = 1, 2; i = 1, ..., 6; d = 1, ..., 406

Scealy (2010) - ABS Small Area Estimation Using a Multinomial Logit Mixed Model with Category Specific **Random Effects**

She defined a model-based approach to producing small area estimates of counts for different categories of the Australian labour force based on a multinomial logit mixed model with category specific random effects.

with the employed category and the other associated with the unemployed category.

$$\log\left(\frac{p_{dij}}{p_{di3}}\right) = \boldsymbol{x}_{di}\boldsymbol{\beta}_j + u_{dj}$$

- Within each small area there are two correlated random effects, one associated

 $u_d \sim Normal(\mathbf{0}, W)$

López-Vizcaíno, Lombardía, Morales (2013) Multinomial-based small area estimation of labour force indicators

on the categories of the response vector.

$$\log\left(\frac{p_{dij}}{p_{di3}}\right) = \boldsymbol{x}_{di}\boldsymbol{\beta}_j + u_{dj}$$

They proposed small area estimators of labour force characteristics (totals of employed, unemployed and inactive people and unemployment rates) that were derived from a multinomial logit mixed model with independent random effects

 $u_d \sim Normal(0, V = diag(\phi_1, \phi_2))$

López-Vizcaíno, Lombardía, Morales (2014) **Multinomial-based small area estimation of labour force indicators**

logit mixed model that included correlated time and area random effects

$$\log\left(\frac{p_{dij}}{p_{di3}}\right)$$

It is also assumed that u_1 and u_2 are independent, with:

$$u_1 \sim Normal(0)$$

- They proposed small area estimators of labour force derived from a multinomial
 - $= \mathbf{x}_{di} \mathbf{\beta}_i + u_{di}$

 - $(0, V_1 = diag(\phi_1, \phi_2))$
 - $\boldsymbol{u}_2 \sim Normal\left(\boldsymbol{0}, \boldsymbol{V}_2 \sim AR(1)\right)$

A non-informative Bayesian approach Under revision

ECLAC in collaboration with Franco

We considered a sampling and a linking model in an area-level set up, where the random effects are correlated

 $\log\left(\frac{p_{dij}}{p_{di3}}\right) = \mathbf{x}_{di}\mathbf{\beta}_j + u_{dj}$

n_{dij} ~

 $\tilde{y}_{dij} = \tilde{n}_{dij} p_{dij} \sim Multinomial(\sum \tilde{y}_{dij}, \boldsymbol{p} = (p_{di1}, p_{di2}, p_{di3})),$

- $u_d \sim Normal(\mathbf{0}, W)$
- In the sampling model, we considered the design-effects in the following way:

$$\frac{p_{dij}(1-p_{dij})}{v_{dij}}$$

Cholesky decomposition and prior distributions

Under the Cholesky decomposition, we assume that **L** is the Cholesky factor of the correlation matrix ρ . This way:

$$W = diag(\sigma_1, \sigma_2)$$

We considered a traditional noninformative approach

$$\boldsymbol{\beta}_j \sim N$$

 $\sigma_j \sim Inverse - gan$

- σ_2 × ρ × diag(σ_1, σ_2) $= L \times L'$

 $Vormal(\mathbf{0}, \mathbf{A}_{i})$ mma (0.00001, 0.000001)

STAN - Parameters

parameters {
 matrix[P-1, K] beta;
 vector<lower=0>[P-1] sigma_u;
 cholesky_factor_corr[P-1] L_u;
 matrix[P-1, D] z_u;
}

transformed parameters {
 simplex[P] theta[D];
 matrix[P-1, D] u; // random effect matrix
 u = diag_pre_multiply(sigma_u, L_u) * z_u;
 }

STAN - Model

model { L u ~ lkj corr cholesky(1); to vector(z u) ~ normal(0, 10000); sigma u ~ inv gamma(0.0001, 0.0001);

beta $[p-1, k] \sim normal(0, 10000);$ target += multinomial lpmf(hat y[d,] | theta[d,]);

Benchmarking

Once the chains have reached convergence, we use the direct estimates at the national and regional levels to benchmark the small area estimates. This process is carried out for each iteration of the MCMC process.

Further research should be conducted regarding the benchmarking process and how it can take into account the model-based variance.

S



🔶 Directo 🔶 Mod 🔶 Bench













Implementation of the model in LAC region

ECLAC's Household Survey Data Bank (BADEHOG)

This is a repository of household surveys from 18 Latin American countries maintained by the Statistics Division.

- In the case of Chile, the 2017 National Social and Economic Survey (CASEN) survey) corresponds to a representative sample at the national, regional, urban, and rural levels.
- For Colombia, the Great Integrated Household Survey of 2018, which is representative of the national, urban, rural, and regional levels, along with departments and their capitals, was used.
- In the case of Peru, the 2017 National Household Survey (ENAHO), which is representative of the national, urban, rural, and departmental levels, was considered. Table 1 shows a comprehensive summary of the household surveys used in this system.

Country	
ARG	Permanent H
BOL	Nationa
BRA	National Survey by
CHL	National Socioeconomic
COL	Large Integr
CRI	National Hou
DOM	National Continuous
ECU	National Survey on Employment, Ur
GTM	National Sur
HND	Multipurpo
MEX	National Household Inco
NIC	National Household Surv
PAN	Mult
PER	National Household Sur
PRY	Continuous Permar
SLV	Multipurpo
URY	Continuo

Survey

- Household Survey (EPH)
- I Household Survey
- Continuous Household Sample
- c Characterization Survey (CASEN)
- rated Household Survey
- usehold Survey (ENAHO)
- s Labour Force Survey (ENCFT)
- nemployment and Underemployment (ENEMDU)
- vey on Living Conditions
- ose Household Survey
- me and Expenditure Survey (ENIGH)
- vey on Living Standard Measurement
- ipurpose Survey
- rvey Living Conditions and Poverty
- nent Household Survey (EPHC)
- ose Household Survey
- us Household Survey

Year



ECLAC's census data bank (CELADES')

This is a repository of LAC censuses maintained by the ECLAC Population Division (CELADE), which has pursued its ongoing activity of disseminating the census information derived from its data bank.

Most Latin American countries have shared their databases to CELADE, allowing the availability of microdata from the previous and current rounds of censuses.

We used the software Redatam for statistical processing specialized in microdata of population and housing censuses developed by CELADE.





Country	
ARG	National Census of Po
BOL	Population
BRA	Dem
CHL	Population
COL	National Census of I
CRI	X National Censu
DOM	IX National Cense
ECU	VII Population
GTM	XII National Census of
HND	XVII Populat
MEX	Population
NIC	Population and H
PAN	
PER	XII Census of Population, VII Housing
PRY	II National Indigenous
SLV	VI Population Ce
URY	Por

Census

opulation, Households, and Housing		
on and Housing Census		
nographic Census		
on and Housing Census		
f Population and Housing (CNPV)		
sus of Population and Housing		
sus of Population and Housing		
tion and Housing Census		
f Population and VII Housing Census		
ation and Housing Census		
on and Housing Census		
Housing Census of Nicaragua		
Census 2010		
g Census, and III Indigenous Communities Census		
s Census of Population and Housing		
ensus and V Housing Census		

pulation Census

Year



Satellite Imagery

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.

- variables that can be proxies for different economic aspects.

 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.

Data panels can be built at a low marginal cost of variables as diverse as night lights, rainfall, wind speed, floods, topography, forest cover, types of crops, urban development, kind of road services, among many other

Some preliminary maps Use with caution

Unemployment Rate

0.00 to 0.05

0.05 to 0.10

0.10 to 0.15

0.15 to 0.20

0.20 to 0.25





Occupancy Rate

0.2 to 0.3

0.3 to 0.4

0.4 to 0.5

0.5 to 0.6

0.6 to 0.7

0.7 to 0.8

0.8 to 0.9





Participation Rate

0.4 to 0.5

0.5 to 0.6

0.6 to 0.7

0.7 to 0.8

0.8 to 0.9





Unemployment Rate

0.00 to 0.05

0.05 to 0.10

0.10 to 0.15

0.15 to 0.20

0.20 to 0.25

0.25 to 0.30





Occupancy Rate

0.4 to 0.5

0.5 to 0.6

0.6 to 0.7

0.7 to 0.8

0.8 to 0.9

0.9 to 1.0











. .











Xavier Mancero Chief of the Social Statistics Unit

> **Stalyn Guerrero Consultant and R expert**

Carolina Franco Principal Statistician at NORC

Teamwork

Rolando Ocampo Director of the Statistics Division



UNITED NATIONS



Thank you for your attendance!