Modeling temporal changes in small area incomes under a random regression coefficients two-fold Fay-Herriot model

Dr. Naomi Diz Rosales
Dr. María José Lombardía and Dr. Domingo Morales

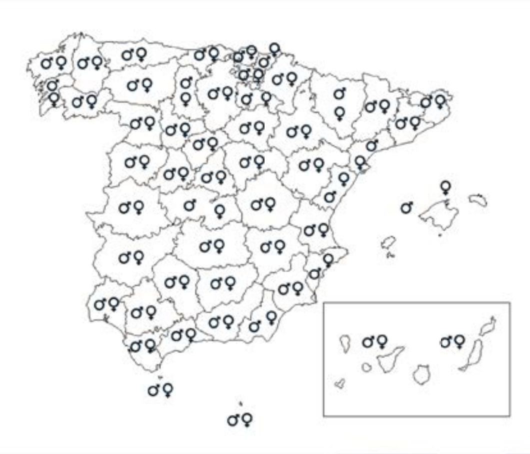
IPS 1261 – Advances in Methods for Scarce and Missing Data

04-06-2026

Introduction

Why this research?

Why this research?



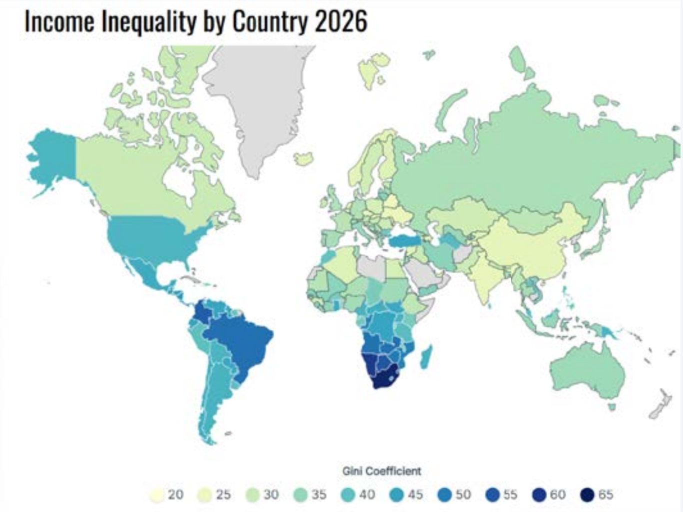


Figure 1: Income inequality by country 2026. Gini Coefficient. (Data from World Population Review (2026)).

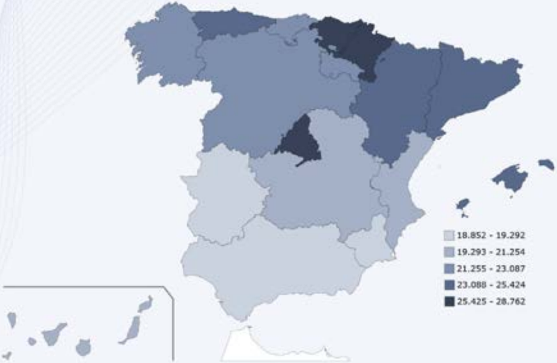


Figure 2: Income per person and consumption per household by autonomous community (data from INE (2026)).

Parameter of interest

$$\overline{Y}_{dt} = \frac{1}{N_{dt}} \sum_{h=1}^{N_{dt}} y_{dt,h}, \quad d = 1, \ldots, D, t = 1, \ldots, T.$$

- $\overline{Y}_{dt}$: Equivalised income in the d^{th} Autonomous Community (AC) in the t^{th} year.
- D and T: Total number of AC and total number of years.
- N_{dt}: Population size in the d^{th} AC in the t^{th} year.
- $y_{dt,h}$: Equivalised income of household to which individual h belongs in domain d and year t.

Parameter of interest

$$\overline{Y}_{dt} = \frac{1}{N_{dt}} \sum_{h=1}^{N_{dt}} y_{dt,h}, \quad d = 1, \ldots, D, t = 1, \ldots, T.$$

- $\overline{Y}_{dt}$: Equivalised income in the d^{th} Autonomous Community (AC) in the t^{th} year.
- D and T: Total number of AC and total number of years.
- N_{dt}: Population size in the d^{th} AC in the t^{th} year.
- $y_{dt,h}$: Equivalised income of household to which individual h belongs in domain d and year t.

Parameter of interest

$$\overline{Y}_{dt} = \frac{1}{N_{dt}} \sum_{h=1}^{N_{dt}} y_{dt,h}, \quad d = 1, \dots, D, t = 1, \dots, T.$$

- $\overline{Y}_{dt}$: Equivalised income in the d^{th} Autonomous Community (AC) in the t^{th} year.
- D and T: Total number of AC and total number of years.
- N_{dt}: Population size in the d^{th} AC in the t^{th} year.
- $y_{dt,h}$: Equivalised income of household to which individual h belongs in domain d and year t.

Parameter of interest

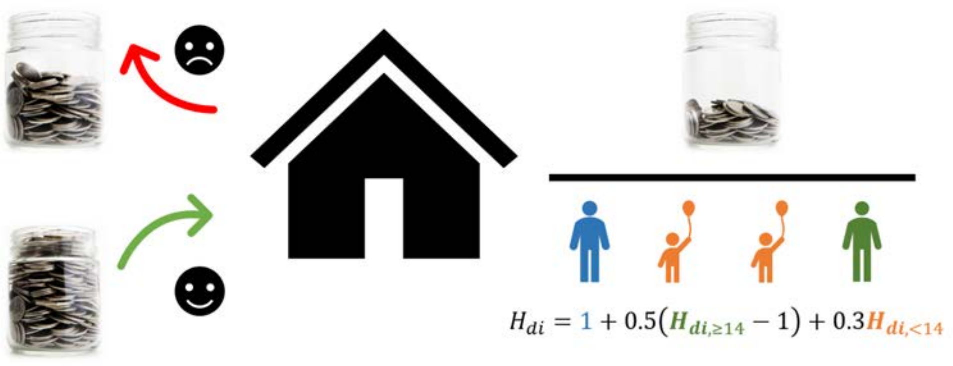
$$\overline{Y}_{dt} = \frac{1}{N_{dt}} \sum_{h=1}^{N_{dt}} y_{dt,h}, \quad d = 1, \ldots, D, t = 1, \ldots, T.$$

- $\overline{Y}_{dt}$: Equivalised income in the d^{th} Autonomous Community (AC) in the t^{th} year.
- D and T: Total number of AC and total number of years.
- N_{dt}: Population size in the d^{th} AC in the t^{th} year.
- $y_{dt,h}$: Equivalised income of household to which individual h belongs in domain d and year t.

Parameter of interest

$$\overline{Y}_{dt} = \frac{1}{N_{dt}} \sum_{h=1}^{N_{dt}} y_{dt,h}, \quad d = 1, \dots, D, t = 1, \dots, T.$$

- $\overline{Y}_{dt}$: Equivalised income in the d^{th} Autonomous Community (AC) in the t^{th} year.
- D and T: Total number of AC and total number of years.
- N_{dt}: Population size in the d^{th} AC in the t^{th} year.
- $y_{dt,h}$: Equivalised income of household to which individual h belongs in domain d and year t.



$$H_{di} = 1 + 0.5\left(H_{di,\geq 14} - 1\right) + 0.3H_{di,<14}$$

Parameter of interest

$$\overline{Y}_{dt} = \frac{1}{N_{dt}} \sum_{h=1}^{N_{dt}} y_{dt,h}, \quad d = 1, \dots, D, t = 1, \dots, T.$$

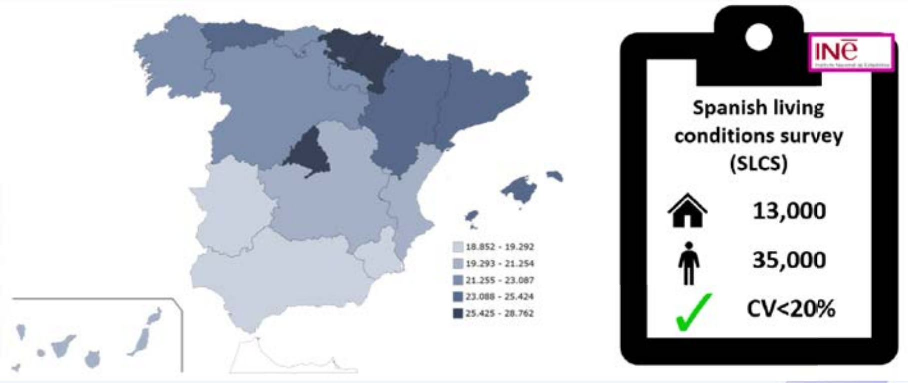
- $\overline{Y}_{dt}$: Equivalised income in the d^{th} Autonomous Community (AC) in the t^{th} year.
- D and T: Total number of AC and total number of years.
- N_{dt}: Population size in the d^{th} AC in the t^{th} year.
- $y_{dt,h}$: Equivalised income of household to which individual h belongs in domain d and year t.

Sampling

- Domain sample size: n_{dt}

Sampling

- Domain sample size: n_{dt}



Sampling

- Domain sample size: n_{dt}
- A Hájek direct estimator of $\overline{Y}_{dt}$ is:

$$y_{dt} = \hat{\overline{Y}}_{dt}^{dir} = \frac{1}{\hat{N}_{dt}^{dir}} \sum_{h=1}^{n_{dt}} w_{dt,h} y_{dt,h}, \quad \hat{N}_{dt}^{dir} = \sum_{h=1}^{n_{dt}} w_{dt,h}, \quad d = 1, \ldots, D, t = 1, \ldots, T.$$

Sampling

- Domain sample size: n_{dt}
- A Hájek direct estimator of $\overline{Y}_{dt}$ is:

$$y_{dt} = \hat{\overline{Y}}_{dt}^{dir} = \frac{1}{\hat{N}_{dt}^{dir}} \sum_{h=1}^{n_{dt}} w_{dt,h} y_{dt,h}, \quad \hat{N}_{dt}^{dir} = \sum_{h=1}^{n_{dt}} w_{dt,h}, \quad d = 1, \ldots, D, t = 1, \ldots, T.$$

$$\widehat{\text{Var}}_{\pi}\left(\hat{\overline{Y}}_{dt}^{\text{dir}}\right) = \frac{1}{\hat{N}_{dt}^{\text{dir}^2}} \sum_{h \in \mathbb{S}_{dt}} w_{dt,h}(w_{dt,h} - 1)\left(y_{dt,h} - \hat{\overline{Y}}_{dt}^{\text{dir}}\right)^2.$$

Sampling

- Domain sample size: n_{dt}
- A Hájek direct estimator of $\overline{Y}_{dt}$ is:

$$y_{dt} = \hat{\overline{Y}}_{dt}^{dir} = \frac{1}{\hat{N}_{dt}^{dir}} \sum_{h=1}^{n_{dt}} w_{dt,h} y_{dt,h}, \quad \hat{N}_{dt}^{dir} = \sum_{h=1}^{n_{dt}} w_{dt,h}, \quad d = 1, \ldots, D, t = 1, \ldots, T.$$

$$\widehat{\text{Var}}_{\pi}(\hat{\overline{Y}}_{dt}^{\text{dir}}) = \frac{1}{\hat{N}_{dt}^{\text{dir}^2}} \sum_{h \in \mathbb{S}_{dt}} w_{dt,h}(w_{dt,h} - 1)(y_{dt,h} - \hat{\overline{Y}}_{dt}^{\text{dir}})^2.$$

where the $w_{dt,h}$ are the elevation factors and n_{dt} is the survey sample size.

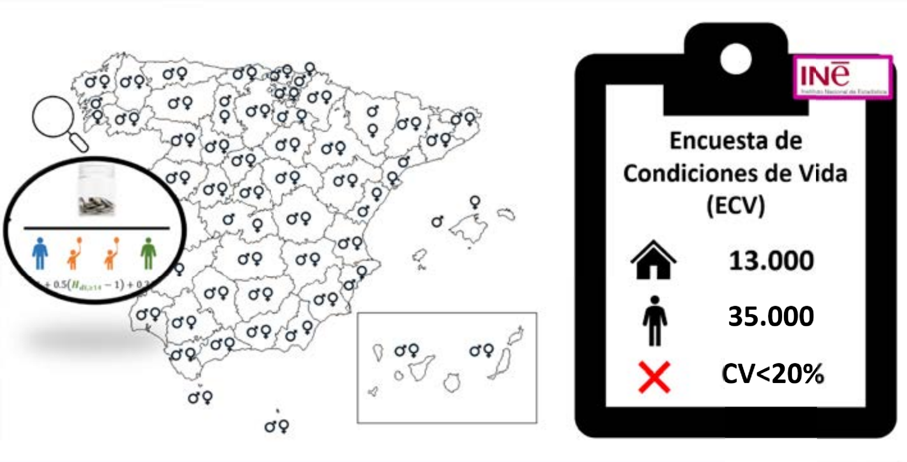
Why this research?

The World Bank Group works to end poverty in several ways:

- Funding projects that can have transformational impacts on communities.
- Collecting and analyzing the critical data and evidence needed to target these programs to reach the poorest and most vulnerable.
- Helping governments create more inclusive, effective policies that can benefit entire populations and lay the groundwork for prosperity for future generations.

Figure 3: Strategy to end poverty (The World Bank, 2023).

Why this research?



Small Area Estimation (SAE)

A multidisciplinary branch of statistics that aims to produce precise estimates of variables of interest in domains or areas with small or no sample size to apply traditional inferential methods that are efficient under asymptotic conditions.

Formal Framework of the SAE Problem

Let:

- $\mathbb{U} = \{1, \ldots, N\}$ be a finite population.
- Partitioned into D disjoint domains $\mathbb{U}_1, \ldots, \mathbb{U}_D$.
- Domain sizes $N_1, \ldots, N_D$.

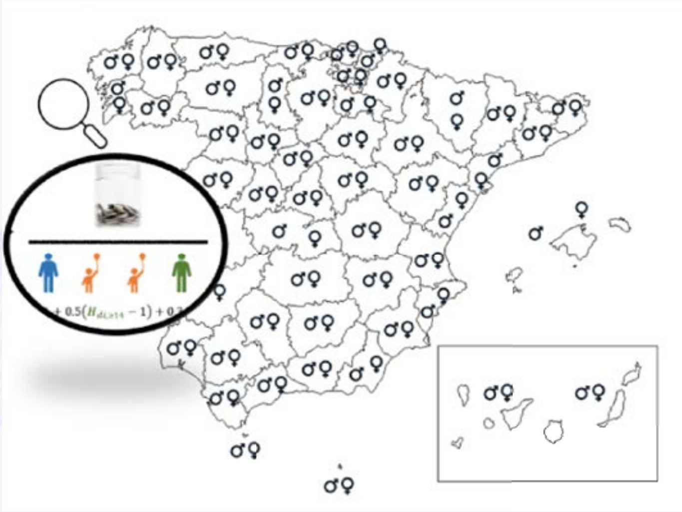
Small Area Estimation (SAE)

A multidisciplinary branch of statistics that aims to produce precise estimates of variables of interest in domains or areas with small or no sample size to apply traditional inferential methods that are efficient under asymptotic conditions.

Formal Framework of the SAE Problem

Let:

- $\mathbb{U} = \{1, \ldots, N\}$ be a finite population.
- Partitioned into D disjoint domains $\mathbb{U}_1, \ldots, \mathbb{U}_D$.
- Domain sizes $N_1, \ldots, N_D$.



Small Area Estimation (SAE)

A multidisciplinary branch of statistics that aims to produce precise estimates of variables of interest in domains or areas with small or no sample size to apply traditional inferential methods that are efficient under asymptotic conditions.

- Domain: a geographical, socio-demographic or other grouping with few available observations of an event of interest.
- It is strongly recommended to consult
  - Fundamental works by Fay and Herriot (1979) and Battese et al. (1988).
  - Reviews by Pfeffermann (2013) and Jiang and Rao (2020).
  - Monographs by Rao and Molina (2015) and Morales et al. (2021).

Area level Models

Random-intercept model

(Fay and Herriot, 1979)

$$y_d = (\beta_0 + u_{0d}) + \beta_1 x_d + e_d$$

Random Slope Models: Visualization

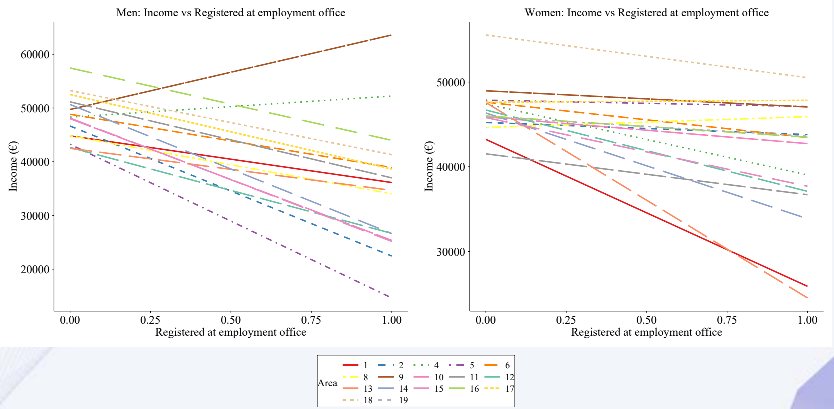


Figure 4: Income trends by area, sex and unemployment registration (data from Morales et al. (2021)).

Random Slope Models

Random-intercept model

$$y_d = (\beta_0 + u_{0d}) + \beta_1 x_d + e_d$$

Random slope model:

$$y_d = (\beta_0 + u_{0d}) + (\beta_1 + u_{1d})\, x_d + e_d$$

with area-specific random effects:

$$\begin{pmatrix} u_{0d} \\ u_{1d} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma^2_{u_0} & \sigma_{u_0 u_1} \\ \sigma_{u_0 u_1} & \sigma^2_{u_1} \end{pmatrix}\right)$$

Random Slope Models in SAE

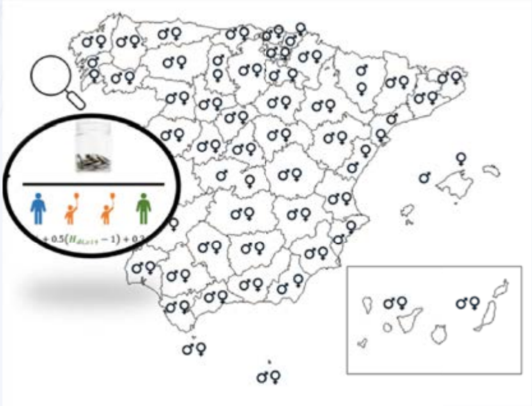
Year	Reference	Contribution
1981	Dempster	Random slope MMs defined in SAE
1990	Prasad and Rao	EBLUP and analytical MSE for random slope models
2013	Hobza and Morales	Recent LMM extensions in SAE
2021	Morales et al.	Further LMM developments
2023	Anjoy and Chandra; May et al.	Bayesian spatio-temporal random slopes

Research gap

Despite clear empirical evidence, random slopes have **not been extended** to complex indicators (poverty, income rates) under normal and non-normal distributions.

Objective

To estimate and **map** the **equivalised disposable income** in Spain by **province, sex and year (2013-2022)** using random slopes Fay-Herriot models for **small area estimation**.



Methodology

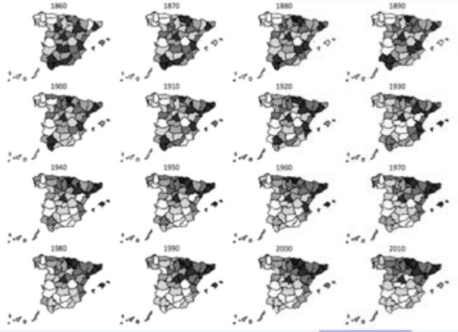
Two-fold Fay-Herriot: Motivation

The Starting Point: Two-fold FH

Esteban et al. (2012)

- **Structure:** Nested domains.
- **Limitation:** Parallel regression lines.

$$y_{dt} = \mu_{dt} + e_{dt},$$
$$\mu_{dt} = \mathbf{x}_{dt}\beta + u_{1,d} + u_{2,dt}.$$



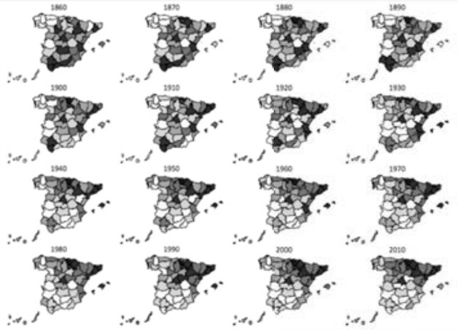
Two-fold Fay-Herriot: Motivation

The Starting Point: Two-fold FH

Esteban et al. (2012)

- **Structure:** Nested domains.
- **Limitation:** Parallel regression lines.

$$
\begin{aligned}
y_{dt} &= \mu_{dt} + e_{dt}, \\
\mu_{dt} &= \mathbf{x}_{dt}\beta + u_{1,d} + u_{2,dt}.
\end{aligned}
$$



Random regression coefficients and time effects two-fold Fay-Herriot (RRCFH2T) model

The RRCFH2T Model

Sampling Model

$$y_{dt} = \mu_{dt} + e_{dt}, \quad e_{dt} \sim N(0, \sigma^2_{edt}), \quad d = 1, \ldots, D, \ \ t = 1, \ldots, T$$

Linking Model with Random Slopes + AR(1)

$$\mu_{dt} = \underbrace{\mathbf{x}_{1,dt}\boldsymbol{\beta}_1 + \mathbf{x}_{1,dt}\mathbf{u}_{1,d}}_{\text{Random coefficients (domain)}} + \underbrace{\mathbf{x}_{2,dt}\boldsymbol{\beta}_2}_{\text{Fixed only}} + \underbrace{u_{2,dt}}_{\text{AR(1) time effect}}$$

- $\mathbf{u}_{1,d} \sim N_{p_1}(\mathbf{0}, \mathbf{V}_{1d})$: domain-level random slopes (province/sex).
- $\mathbf{u}_{2,d} \sim N_T(\mathbf{0}, \sigma^2_2\,\Omega_d(\rho_2))$: AR(1)-correlated time effects within domain d.
- σ^2_{edt}: known sampling variances.

</> **New R Package:** **`slopeSAE`**

Covariance Structure

Random slopes covariance

$$\boldsymbol{V}_{1d} = (\sigma_{1,ij})_{i,j=1,\ldots,p_1}$$

$$\boldsymbol{V}_{1d} = \begin{pmatrix} \sigma^2_{1,1} & \sigma_{1,1}\sigma_{1,2}\rho_{1,12} & \cdots & \sigma_{1,1}\sigma_{1,p}\rho_{1,1p} \\ \sigma_{1,2}\sigma_{1,1}\rho_{1,12} & \sigma^2_{1,2} & \cdots & \sigma_{1,2}\sigma_{1,p}\rho_{1,2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,p}\sigma_{1,1}\rho_{1,1p} & \sigma_{1,p}\sigma_{1,2}\rho_{1,2p} & \cdots & \sigma^2_{1,p} \end{pmatrix}$$

Time covariance within domain d

$$\boldsymbol{V}_{2d} = \sigma_2^2\,\Omega_d(\rho_2)$$

$$\Omega_d = \frac{1}{1-\rho_2^2}\begin{pmatrix} 1 & \rho_2 & \cdots & \rho_2^{T-1} \\ \rho_2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_2 \\ \rho_2^{T-1} & \cdots & \rho_2 & 1 \end{pmatrix}_{T\times T}$$

Matrix Representation: $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$

Design matrices:

- $\boldsymbol{X} = \operatorname{col}_d(\boldsymbol{X}_d), \quad \boldsymbol{X}_d = (\boldsymbol{X}_{1,d}, \boldsymbol{X}_{2,d})$
- $\boldsymbol{Z}_1 = \operatorname{diag}_d(\boldsymbol{Z}_{1d}), \quad \boldsymbol{Z}_{1d} = \operatorname{col}_t(\boldsymbol{x}_{1,dt})$
- $\boldsymbol{Z}_2 = \boldsymbol{I}_M, \quad M = DT$

Covariance structure:

- $\boldsymbol{V}_1 = \operatorname{diag}_d(\boldsymbol{V}_{1d})$
- $\boldsymbol{V}_2 = \operatorname{diag}_d(\sigma_2^2 \Omega_d(\rho_2))$
- $\boldsymbol{V}_e = \operatorname{diag}_{d,t}(\sigma_{edt}^2)$ (known)

Total variance-covariance (domain level)

$$\boldsymbol{V}_d = \operatorname{var}(\boldsymbol{y}_d) = \boldsymbol{Z}_{1d}\boldsymbol{V}_{1d}\boldsymbol{Z}_{1d}' + \sigma_2^2\Omega_d + \boldsymbol{V}_{ed}$$

Population level

$$\boldsymbol{V} = \operatorname{var}(\boldsymbol{y}) = \boldsymbol{Z}_1\boldsymbol{V}_1\boldsymbol{Z}_1' + \boldsymbol{V}_2 + \boldsymbol{V}_e = \operatorname{diag}_d(\boldsymbol{V}_d)$$

Estimation: BLUP and EBLUP

If $\boldsymbol{V}_{1d}$, σ_2^2, ρ_u are known, the **BLUP** of μ_{dt} is $\tilde{\mu}_{dt} = \boldsymbol{x}_{1,dt}\tilde{\beta}_1 + \boldsymbol{x}_{2,dt}\tilde{\beta}_2 + \boldsymbol{x}_{1,dt}\tilde{\boldsymbol{u}}_{1d} + \tilde{u}_{2,dt}$, where

$$\tilde{\beta} = \left(\sum_d \boldsymbol{X}_d' \boldsymbol{V}_d^{-1} \boldsymbol{X}_d\right)^{-1} \left(\sum_d \boldsymbol{X}_d' \boldsymbol{V}_d^{-1} \boldsymbol{y}_d\right)$$

$$\tilde{\boldsymbol{u}} = \begin{pmatrix} \underset{d}{\mathrm{col}}\left(\boldsymbol{V}_{1d}\boldsymbol{Z}_{1d}'\boldsymbol{V}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\beta})\right) \\ \sigma_2^2 \underset{d}{\mathrm{col}}\left(\Omega_d \boldsymbol{V}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\beta})\right) \end{pmatrix}$$

EBLUP (plug-in REML estimates $\hat{\boldsymbol{V}}_{1d}$, $\hat{\sigma}_2^2$, $\hat{\rho}_2$)

$$\hat{\beta} = (\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}, \qquad \hat{\boldsymbol{u}} = \hat{\boldsymbol{V}}_u\boldsymbol{Z}'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})$$

$$\hat{\mu}_{dt} = \boldsymbol{x}_{1,dt}\hat{\beta}_1 + \boldsymbol{x}_{2,dt}\hat{\beta}_2 + \boldsymbol{x}_{1,dt}\hat{\boldsymbol{u}}_{1d} + \hat{u}_{2,dt}$$

Mean Squared Error Estimation

We define three MSE estimators of the EBLUP $\hat{\mu}_{dt}$:

- **Analytic MSE** (mse^{A}_{dt}): Prasad and Rao (1990) decomposition adapted for the RRCFH2T structure under REML:

$$MSE(\hat{\mu}_{dt}) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3(\boldsymbol{\theta})$$
$$\widehat{mse}^{A}_{dt} = g_1(\hat{\boldsymbol{\theta}}) + g_2(\hat{\boldsymbol{\theta}}) + 2g_3(\hat{\boldsymbol{\theta}})$$

- **Parametric Bootstrap** (mse^{PB}_{dt}): González-Manteiga et al. (2008), adapted to account for the AR(1) structure in the bootstrap data generation.

- **Bias-corrected Bootstrap** (mse^{EF}_{dt}): Erciulescu and Fuller (2013) correction applied to reduce bootstrap bias in the MSE estimate.

Simulations

Simulation Studies — Overview

Sim.	Focus	Setup	Key Finding
1	**Parameter estimators**	(R)Bias and (R)RMSE of $\hat{\theta}$ via REML	Estimators unbiased and precise.
2	**Predictor behaviour**	(R)Bias and (R)RMSE of EBLUP	Practically unbiased and RRMSE $<$ 5%.
3	**Sensitivity analysis**	RRCFH2T with different $\mathbf{V}_{1d}$ structures	Robust. Best with random slopes and AR(1).
4	**MSE estimators**	**A** vs. **PB** vs. **EF**	Analytic MSE estimator best choice.
5	**Robustness**	Random-effects with skew-normal distributions	EBLUP and analytic MSE show robustness.

Simulation Studies — Overview

Sim.	Focus	Setup	Key Finding
1	**Parameter estimators**	(R)Bias and (R)RMSE of $\hat{\theta}$ via REML	Estimators unbiased and precise.

Simulation Studies — Overview

Sim.	Focus	Setup	Key Finding
1	**Parameter estimators**	(R)Bias and (R)RMSE of $\hat{\theta}$ via REML	Estimators unbiased and precise.
2	**Predictor behaviour**	(R)Bias and (R)RMSE of EBLUP	Practically unbiased and RRMSE $< 5\%$.

Simulation Studies — Overview

Sim.	Focus	Setup	Key Finding
1	**Parameter estimators**	(R)Bias and (R)RMSE of $\hat{\theta}$ via REML	Estimators unbiased and precise.
2	**Predictor behaviour**	(R)Bias and (R)RMSE of EBLUP	Practically unbiased and RRMSE $< 5\%$.
3	**Sensitivity analysis**	RRCFH2T with different $\mathbf{V}_{1d}$ structures	Robust. Best with random slopes and AR(1).

Simulation Studies — Overview

Sim.	Focus	Setup	Key Finding
1	**Parameter estimators**	(R)Bias and (R)RMSE of $\hat{\theta}$ via REML	Estimators unbiased and precise.
2	**Predictor behaviour**	(R)Bias and (R)RMSE of EBLUP	Practically unbiased and RRMSE $< 5\%$.
3	**Sensitivity analysis**	RRCFH2T with different $\mathbf{V}_{1d}$ structures	Robust. Best with random slopes and AR(1).
4	**MSE estimators**	**A** vs. **PB** vs. **EF**	Analytic MSE estimator best choice.

Simulation Studies — Overview

Sim.	Focus	Setup	Key Finding
1	**Parameter estimators**	(R)Bias and (R)RMSE of $\hat{\theta}$ via REML	Estimators unbiased and precise.
2	**Predictor behaviour**	(R)Bias and (R)RMSE of EBLUP	Practically unbiased and RRMSE $< 5\%$.
3	**Sensitivity analysis**	RRCFH2T with different $\mathbf{V}_{1d}$ structures	Robust. Best with random slopes and AR(1).
4	**MSE estimators**	**A** vs. **PB** vs. **EF**	Analytic MSE estimator best choice.
5	**Robustness**	Random-effects with skew-normal distributions	EBLUP and analytic MSE show robustness.

Application to real data

Selected model

RRCFH2T Model. REML Estimation

$$\mu_{dt} = \beta_{1,1}x_{1,dt1} + \beta_{2,1}x_{2,dt1} + \beta_{2,2}x_{2,dt2} + \beta_{2,3}x_{2,dt3} + x_{1,dt1}u_{1,d1} + u_{2,dt}.$$

Table 2: REML estimates and asymptotic 95 % confidence intervals.

	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,1}$	$\hat{\beta}_{2,2}$	$\hat{\beta}_{2,3}$	$\hat{\sigma}_{1,11}$	$\hat{\sigma}_2$	$\hat{\rho}_2$
2.5%	18.563	6.533	−18.489	2.046	0.500	0.341	0.797
Est.	**22.042**	**8.204**	**−13.819**	**6.489**	**8.000**	**0.432**	**0.880**
97.5%	25.522	9.875	−9.149	10.932	15.501	0.522	0.963

Selected model

RRCFH2T Model. REML Estimation

$$\mu_{dt} = \beta_{1,1} x_{1,dt1} + \beta_{2,1} x_{2,dt1} + \beta_{2,2} x_{2,dt2} + \beta_{2,3} x_{2,dt3} + x_{1,dt1} u_{1,d1} + u_{2,dt}.$$

Table 2: REML estimates and asymptotic 95 % confidence intervals.

	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,1}$	$\hat{\beta}_{2,2}$	$\hat{\beta}_{2,3}$	$\hat{\sigma}_{1,11}$	$\hat{\sigma}_{2}$	$\hat{\rho}_{2}$
2.5%	18.563	6.533	−18.489	2.046	0.500	0.341	0.797
Est.	**22.042**	**8.204**	**−13.819**	**6.489**	**8.000**	**0.432**	**0.880**
97.5%	25.522	9.875	−9.149	10.932	15.501	0.522	0.963
	edu3						

Selected model

RRCFH2T Model. REML Estimation

$$\mu_{dt} = \beta_{1,1}x_{1,dt1} + \beta_{2,1}x_{2,dt1} + \beta_{2,2}x_{2,dt2} + \beta_{2,3}x_{2,dt3} + x_{1,dt1}u_{1,d1} + u_{2,dt}.$$

Table 2: REML estimates and asymptotic 95 % confidence intervals.

	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,1}$	$\hat{\beta}_{2,2}$	$\hat{\beta}_{2,3}$	$\hat{\sigma}_{1,11}$	$\hat{\sigma}_2$	$\hat{\rho}_2$
2.5%	18.563	6.533	−18.489	2.046	0.500	0.341	0.797
Est.	**22.042**	**8.204**	**−13.819**	**6.489**	**8.000**	**0.432**	**0.880**
97.5%	25.522	9.875	−9.149	10.932	15.501	0.522	0.963
	edu3		**lab2**				

Selected model

RRCFH2T Model. REML Estimation

$$\mu_{dt} = \beta_{1,1}x_{1,dt1} + \beta_{2,1}x_{2,dt1} + \beta_{2,2}x_{2,dt2} + \beta_{2,3}x_{2,dt3} + x_{1,dt1}u_{1,d1} + u_{2,dt}.$$

Table 2: REML estimates and asymptotic 95 % confidence intervals.

	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,1}$	$\hat{\beta}_{2,2}$	$\hat{\beta}_{2,3}$	$\hat{\sigma}_{1,11}$	$\hat{\sigma}_2$	$\hat{\rho}_2$
2.5%	18.563	6.533	−18.489	2.046	0.500	0.341	0.797
Est.	**22.042**	**8.204**	**−13.819**	**6.489**	**8.000**	**0.432**	**0.880**
97.5%	25.522	9.875	−9.149	10.932	15.501	0.522	0.963
	edu3		**lab2**	**cit1**			

Income predictions by province, sex and year

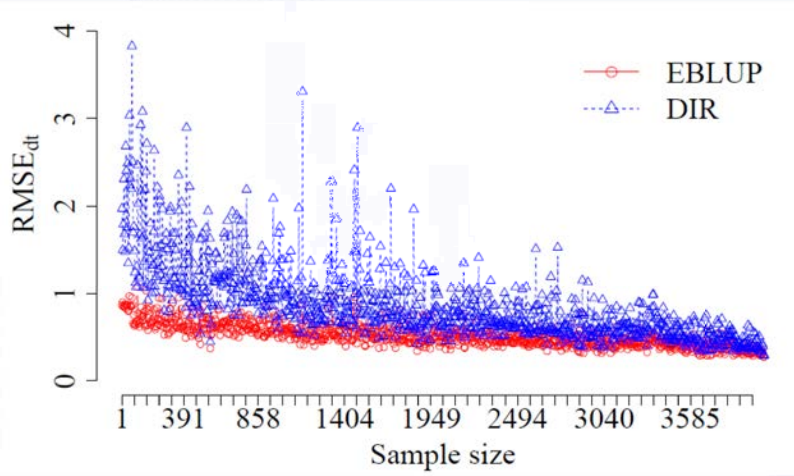


Figure 5: RMSE comparison between the EBLUP and the direct estimator (DIR).

Income mapping

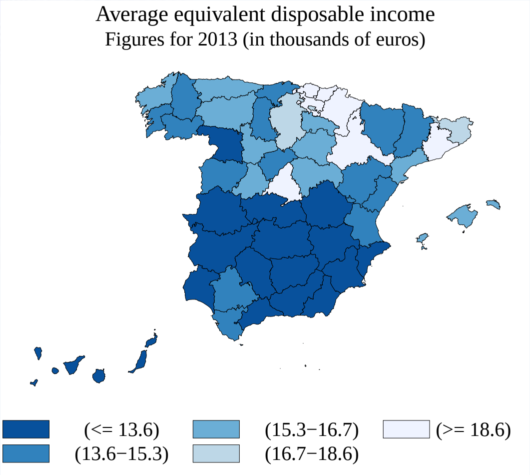


Figure 6: Estimated equivalised disposable income by province in Spain for 2013-2022.

Mapping heterogeneity

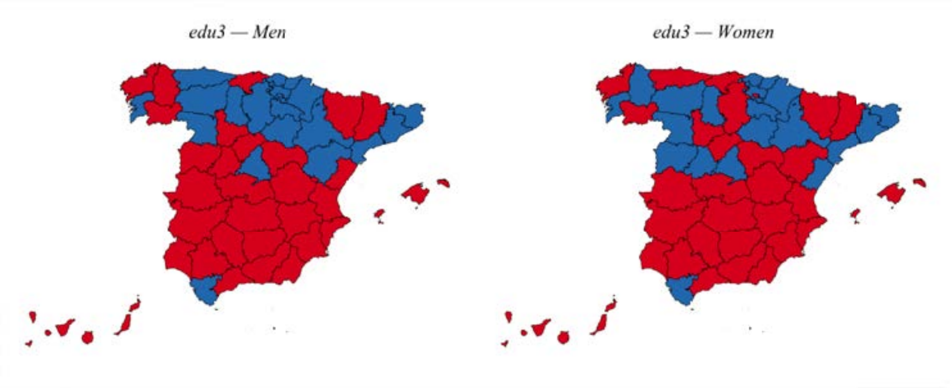


Figure 7: Blue indicates a higher-than-average income return to tertiary education. Red indicates a negative return.

Income sex divergence

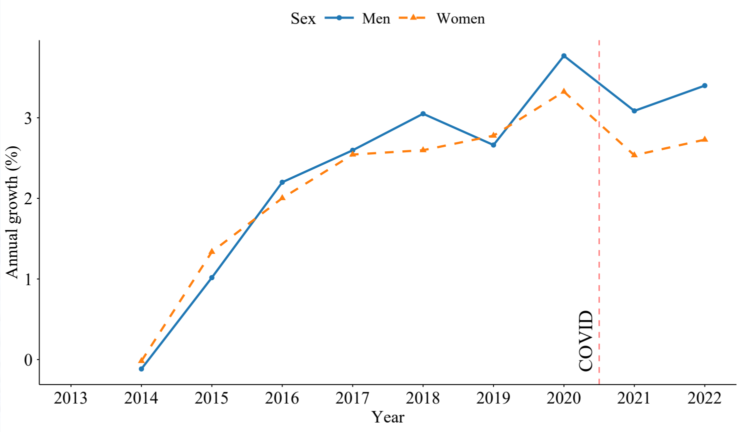


Figure 8: Annual growth in estimated equivalent disposable income by sex, highlighting the impact of the COVID-19 pandemic.

Research transfer

Publishing our research in Open Access mode

All the methodology and results illustrated in this presentation are reported in a paper under second review.

Diz-Rosales, N., Lombardía, M.J., and Morales, D. (2025).Random regression coefficients two-fold Fay–Herriot time models for small area income estimation.

NDiz-Rosales/SmallArea-RandomSlopes


Random regression coefficients time two-fold Fay-Herriot models for small area income estimation

Naomi Diz-Rosales[1], María José Lombardía[2], Domingo Morales[3]

2025-11-15

Structure

Conclusions and Future work

Conclusions

- We propose the RRCFH2T model, which integrates two-fold structures, random slopes, and AR(1) temporal dependence into the Fay-Herriot framework for the first time.
- The proposed EBLUP estimator outperforms classical models in terms of accuracy and stability, especially in domains with small sample sizes.
- The analytical MSE estimator turns out to be the most balanced option, offering high precision with a significantly lower computational cost than bootstrap methods.
- The application confirms a persistent territorial and gender gap in income, which has been exacerbated by the COVID-19 pandemic.

Conclusions

- We propose the RRCFH2T model, which integrates two-fold structures, random slopes, and AR(1) temporal dependence into the Fay-Herriot framework for the first time.
- The proposed EBLUP estimator outperforms classical models in terms of accuracy and stability, especially in domains with small sample sizes.
- The analytical MSE estimator turns out to be the most balanced option, offering high precision with a significantly lower computational cost than bootstrap methods.
- The application confirms a persistent territorial and gender gap in income, which has been exacerbated by the COVID-19 pandemic.

Conclusions

- We propose the RRCFH2T model, which integrates two-fold structures, random slopes, and AR(1) temporal dependence into the Fay-Herriot framework for the first time.
- The proposed EBLUP estimator outperforms classical models in terms of accuracy and stability, especially in domains with small sample sizes.
- The analytical MSE estimator turns out to be the most balanced option, offering high precision with a significantly lower computational cost than bootstrap methods.
- The application confirms a persistent territorial and gender gap in income, which has been exacerbated by the COVID-19 pandemic.

Conclusions

- We propose the RRCFH2T model, which integrates two-fold structures, random slopes, and AR(1) temporal dependence into the Fay-Herriot framework for the first time.
- The proposed EBLUP estimator outperforms classical models in terms of accuracy and stability, especially in domains with small sample sizes.
- The analytical MSE estimator turns out to be the most balanced option, offering high precision with a significantly lower computational cost than bootstrap methods.
- The application confirms a persistent territorial and gender gap in income, which has been exacerbated by the COVID-19 pandemic.

Future work

- Higher-order AR structures
- Time-varying random slopes
- Variance smoothing (GVF)
- Measurement error in covariates

Future work

- Higher-order AR structures
- Time-varying random slopes
- Variance smoothing (GVF)
- Measurement error in covariates

Future work

- Higher-order AR structures
- Time-varying random slopes
- Variance smoothing (GVF)
- Measurement error in covariates

Future work

- Higher-order AR structures
- Time-varying random slopes
- Variance smoothing (GVF)
- Measurement error in covariates

Acknowledgements

The paper is part of grant PRE2021-100857, funded by MCIN/AEI/10.13039/501100011033 and ESF+.

MINISTERIO DE CIENCIA E INNOVACIÓN

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN

Acknowledgements

This research is part of the R&D project PID2020-113578RB-I00, funded by MCIN/AEI/10.13039/501100011033/. It has also been supported by the Spanish R&D project PID2023-147127OB-I00 (funded by MICIU/AEI/10.13039/501100011033/ and "FEDER/UE"), and PID2022-136878NB-I00, the Valencian grant Prometeo/2021/063, by the Xunta de Galicia (Competitive Reference Groups ED431C-2024/14) and by CITIC , as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, receives subsidies from the Department of Education, Science, Universities, and Vocational Training of the Xunta de Galicia. In addition, we thank the Centro de Supercomputación de Galicia (CESGA) for providing their services for part of the simulations in this work.

Thank you!

naomi.diz.rosales@udc.es

References

References

- Anjoy, P., and Chandra, H. (2023). Hierarchical Bayes estimation of small areas means under a spatial nonstationary Fay-Herriot model. *Communications in Statistics-Simulation and Computation*, *52*(7), 3043-3061. https://doi.org/10.1080/03610918.2021.1926501.
- Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, *83*(401), 28-36. https://doi.org/10.2307/2288915.

References

- Dempster, A.P., Rubin, D.B., and Tsutakawa, R.K. (1981). Estimation in covariance components models.*Journal of the American Statistical Association, 76*(374), 341-353. https://doi.org/10.2307/2287835.
- Erciulescu, A. L., and Fuller, W. A. (2013, August 3-8). Small-area prediction of the mean of a binomial random variable. In *Proceedings of the Survey Research Methods Section. Joint Statistical Meetings* (pp. 855-863). American Statistical Association, Montréal, Québec, Canada. http://www.asasrms.org/Proceedings/y2013/files/307932_79947.pdf.

References

- Esteban, M.D., Morales, D., Pérez, A., and Santamaría, L. (2012). Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis, 56*(10), 2840-2855. https://doi.org/10.1016/j.csda.2011.10.015.
- Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association, 74*(366a), 269-277. https://doi.org/10.1080/01621459.1979.10482505.

References

- González-Manteiga, W., Lombardía, M.J., Molina, I., Morales, D., and Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis, 52*(12), 5242-5252. https://doi.org/10.1016/j.csda.2008.04.031.
- Hall, P., and Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 68*(2), 221-238. https://doi.org/10.1111/j.1467-9868.2006.00541.x.

References

- Hobza, T., and Morales, D. (2013). Small area estimation under random regression coefficient models. *Journal of Statistical Computation and Simulation, 83*(11), 2160-2177. https://doi.org/10.1080/00949655.2012.684094.
- Instituto Nacional de Estadística (2026). *Living Conditions Survey (LCS). Income per person and consumption per household by autonomous community*. https://ine.es/jaxiT3/Datos.htm?t=9947#_tabs-mapa.

References

- Jiang, J., and Rao, J.S. (2020). Robust small area estimation: An overview. *Annual review of statistics and its application, 7*, 337-360. https://doi.org/10.1146/annurev-statistics-031219-041212.
- May, P., McConville, K.S., Moisen, G.G., Bruening, J., and Dubayah, R. (2023). A spatially varying model for small area estimates of biomass density across the contiguous United States. *Remote Sensing of Environment, 286*, 113420. https://doi.org/10.1016/j.rse.2022.113420.

References

- Morales, D., Esteban, M.D., Pérez, A., and Hobza, T. (2021). *A course on small area estimation and mixed models. Methods, theory and applications in R.* Springer.
- Pfeffermann, D. (2013). New important developments in small area estimation.*Statistical Science, 28*(1), 40-68. http://dx.doi.org/10.1214/12-STS395.

References

- Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of the mean squared error of small area estimators. *Journal of the American Statistical Association*, *85*(409), 163-171.
- Tirado, D.A., Díez-Minguela, A., and Martínez-Galarraga, J. (2016). Regional inequality and economic development in Spain, 1860-2010.*Journal of Historical Geography,54*, 87-98. https://doi.org/10.1016/j.jhg.2016.09.005.

References

- World Bank Group. (2023). *What are quintiles?* https://datahelpdesk.worldbank.org/knowledgebase/articles/1986160-what-are-quintiles
- World Population Review. (2026). *Income Inequality by Country 2026*.https://worldpopulationreview.com/country-rankings/income-inequality-by-country.