Journal of Survey Statistics and Methodology (2026) ***00****, 1–43*

THREE-FOLD FAY-HERRIOT MODEL WITH UNEQUAL VARIANCE RANDOM EFFECTS

ESTEBAN CABELLO *
LAURA MARCIS
DOMINGO MORALES
MARIA CHIARA PAGLIARELLA
RENATO SALVATORE

This study extends the classical Fay–Herriot model to a threefold hierarchical structure incorporating heteroscedastic random effects. Three submodels are also introduced within this framework. Small area best linear unbiased predictors are derived for linear indicators, and their mean squared errors (MSEs) are estimated using both analytical and parametric bootstrap methods. Model performance and reliability are evaluated through diagnostic tools and influence measures, specifically designed for small area estimation. Simulation experiments are conducted to analyze the empirical properties of the predictors and MSE estimators. The proposed approach is applied to data from the 2019–2021 Spanish Living Conditions Survey to estimate the proportion of women and men below the poverty line, disaggregated by province, age group, and year.

Esteban Cabello is a Research Fellow at the Center of Operations Research, Miguel Hernández University of Elche, Edificio Torretamarit—Avda. de la Universidad s/n, Elche (Alicante) 03202, Spain. Laura Marcis is a Research Fellow at the Department of Economic and Political Sciences, University of the Aosta Valley, Via Monte Vodice, 11100 Aosta, Italy. Domingo Morales is Full Professor at the Center of Operations Research, Miguel Hernández University of Elche, Edificio Torretamarit—Avda. de la Universidad s/n, Elche (Alicante) 03202, Spain. Maria Chiara Pagliarella is Adjunct Professor at the Department of Economics and Law, University of Cassino and Southern Lazio, Campus Folcara, 03043 Cassino, Italy. Renato Salvatore is Researcher at the Department of Economics and Law, University of Cassino and Southern Lazio, Campus Folcara, 03043 Cassino, Italy.
This work was supported by the Conselleria d'Innovació, Universitats, Ciència i Societat Digital of the Generalitat Valenciana [CIPROM/2024/34] and by Agencia Estatal de Investigación [PID2022-136878NB-I00]. The authors declare that they have no conflicts of interest.
*Address correspondence to: Esteban Cabello, Centro de Investigación Operativa (CIO), Miguel Hernández University of Elche, Avenida de la Universidad, s/n, 03202 Elche (Alicante), Spain; E-mail: ecabello@umh.es.

https://doi.org/10.1093/jssam/smag013

KEYWORDS: Area-level models; Diagnostics; Fay–Herriot model; Heteroscedastic random effects; Living Conditions Survey; Poverty proportion; Small area estimation.

Statement of Significance

This research introduces a novel threefold extension of the Fay–Herriot model with heteroscedastic random effects, addressing a key limitation in traditional small area estimation methods. The proposed model captures complex heteroscedastic hierarchical structures—such as those defined by geographical area, age group, sex, and time—that arise in socio-economic surveys. It also incorporates model diagnostics and mean squared error (MSE) estimation procedures, improving both reliability and interpretability of the model inferential process. Through simulations and application to 2019–2021 Spanish Living Condition Survey data, we demonstrate that this approach yields accurate and reliable poverty estimates for disaggregated domains. The paper contributes to the statistical modelling of complex data surveys and provides new tools for applied researchers and official statistics programs.

1. INTRODUCTION

Small area estimation (SAE) is a field of inference in finite populations that has utility when sample data are used to predict indicators in subpopulations with a small sample size. In such subpopulations, called small areas or estimation domains, direct estimators will have large variances and therefore will not be very accurate. An alternative estimation procedure is to introduce auxiliary information through a statistical model. Area-level linear mixed models (LMMs) do this and are more adept at incorporating quality explanatory variables into the inferential process than other unit-based approaches. Fay and Herriot (1979) introduced the basic LMM for SAE. For more information, see Chapter 16 of Morales et al. (2021).

Finite populations are usually subdivided hierarchically into nested levels, with the time period being one of the possible levels. The corresponding hierarchical structure of the data then provides an important source of information that can be used to improve the efficiency of small area estimators. For this reason, Pfeffermann and Burck (1990), Rao and Yu (1994), Ghosh et al. (1996), Datta et al. (1999, 2002), You and Rao (2000), Singh et al. (2005), Esteban et al. (2012), Marhuenda et al. (2013), or

Benavent and Morales (2021) extended the Fay–Herriot (FH) model considering temporal correlation or hierarchical structures to two levels for sampling errors or random effects, time-varying random slopes, and state-space models.

The extension of the Fay–Herriot model to three levels of hierarchy to jointly model data from domains, subdomains, and time periods or sub-subdomains has had less attention from researchers because the usual periodicity of surveys in public statistics (quarterly or annual) does not allow obtaining cross-sectional data obtained with the same methodology over long periods of time. Some interesting contributions include Torabi and Rao (2014), Cai and Rao (2022), Cai et al. (2020), Krenzke et al. (2020), or Marcis et al. (2023). Another area in which there has not been a great research effort is the development of heteroscedastic LMMs. In this area, González-Manteiga et al. (2010) proposed a non-parametric estimator of the variance function of sampling errors, and Morales et al. (2015) introduced a partitioned FH temporal model with unequal variance random effects. More recently, Cabello et al. (2025) have extended the standard FH model by applying Box–Cox transformations to the random effects variances. However, the study of three-level area LMM with non-homoscedastic random effects has not yet been addressed in the scientific literature. This fact motivates the study of threefold Fay–Herriot models with unequal variance random effects, the application to the prediction of poverty indicators in small areas, the development of analytic and parametric bootstrap estimators of the MSEs of predictors and the derivation of influence and diagnostics analysis tools to validate the model and to get valuable information to interpret underlying data relationships.

Poverty is a very serious problem for society. This has led to numerous studies investigating statistical methodologies that allow poverty to be mapped at different levels of territorial aggregation. Regarding estimation techniques in small areas, we can cite the books edited by Pratesi (2016) and Betti and Lemmi (2013) that provide further contributions to the estimation of poverty indicators in small areas. This article introduces a new statistical methodology for poverty mapping and presents an application to data from the Spanish Living Conditions Survey (SLCS). This survey is designed to obtain reliable direct estimators in autonomous communities, but sample sizes are small in provinces. The study aims to estimate poverty ratios by sex and age group in Spanish provinces during the years 2019–2021. For this reason, this article introduces an area-level LMM with random effects across province crossed by age group, across provinces, age group, and sex, and across provinces, age group, sex, and year. The three-fold Fay–Herriot model, with unequal variance random effects, is fitted to the SLCS data, but the results suggest a simpler model without such a complex heteroscedasticity structure. For this reason, the article considers submodels where unequal variances follow simpler patterns.

The article is organized as follows. Section 2 introduces the SLCS data and motivates the use of SAE techniques for the problem of interest. Section 3 introduces the new threefold Fay–Herriot model, its submodels, the empirical best linear predictors (EBLUPs) of linear indicators, the analytic estimators of the MSEs and the specific influence and diagnostic tools. Section 4 presents some simulation experiments to investigate the performance of the fitting algorithm, the predictors, and the MSE estimators. Section 5 deals with the application to real data and the mapping of poverty proportions. Section 6 provides some conclusions. The article contains five appendices. Appendix A (please see the supplementary materials online) describes the Fisher-scoring fitting algorithm to calculate the residual maximum likelihood (REML) estimators of the model parameters. Appendix B (please see the supplementary materials online) derives the components of the analytic estimator of the MSEs and proposes a parametric bootstrap procedure. Appendix C (please see the supplementary materials online) describes the centered log-ratio transformation. Appendix D (please see the supplementary materials online) gives the simulation steps and the performance measures. Appendix E (please see the supplementary materials online) contains additional tables and maps to supplement the analysis of the SLCS data.

2. DATA FRAMEWORK

This section introduces the socio-economic indicator of interest and describes the data. The target indicator is the poverty proportion, that is, the proportion of individuals under the poverty line. The estimation of this indicator by Spanish provinces, age groups, sexes and years, is illustrated with SLCS data from 2019 to 2021. The SLCS serves as an advanced statistical instrument for the European Commission, allowing studies on income, poverty, living conditions and inequality. Understanding these needs of the population allows governments and administrations to structure policies effectively. The SLCS is an annual survey, planned to provide statistical analyses at the NUTS 2 level (autonomous communities), given the inherent challenges in the NUTS 3 level (provinces), where sample sizes are insufficient to yield reliable direct estimates.

In mathematical terms, we consider a population U hierarchically partitioned in domains, subdomains and sub-subdomains; that is, $U = \cup_{d=1}^{D} U_d$, $U_d = \cup_{r=1}^{R} U_{dr}$, and $U_{dr} = \cup_{t=1}^{T} U_{drt}$, with population sizes N, N_d, N_{dr}, and N_{drt}, respectively. In our application to the SLCS, d represents the province crossed by age group, r indicates the sex ($r = 1$ for men and $r = 2$ for women), and t denotes the year ($t = 1$ for 2019, $t = 2$ for 2020, and $t = 3$ for 2021). Without loss of generality, we order the domains according to the 4 categories $[16, 44], [45, 54], [55, 64]$, and $[65, \infty)$ of the age group

variable. This is, $d \in \mathcal{D}_1 = \{1, \ldots, D/4\}$ is for people between 16 and 44 years old, $d \in \mathcal{D}_2 = \{D/4+1, \ldots, D/2\}$ is for people between 45 and 54 years old, $d \in \mathcal{D}_3 = \{D/2+1, \ldots, 3D/4\}$ is for people between 55 and 64 years old and $d \in \mathcal{D}_4 = \{3D/4+1, \ldots, D\}$ is for people over 64 years old. For ease of exposition, sub-subdomains are denoted as areas.

Let z_{drtj} be the normalized net annual income (NNAI) of the household where the individual $j \in U_{drt}$ lives. Let z_t be the poverty threshold at year t. We define the indicator function $y_{drtj} = I(z_{drtj} < z_t)$, which equals 1 if the individual j is under the poverty threshold and equals zero otherwise. The poverty proportion in U_{drt} is

$$\overline{Y}_{drt} = \frac{1}{N_{drt}} \sum_{j=1}^{N_{drt}} y_{drtj}. \tag{1}$$

The Spanish Statistical Office (INE—Instituto Nacional de Estadística) calculates z_{drtj} by summing up the annual net incomes of the household members and dividing by its normalized household size (NHS). The same value of NNAI is assigned to all the individuals in the same household. The definition of the NHS gives a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 14 in the household h. The NHS of the household h at the year t is the sum of the weights assigned to its members, that is

$$n_{ht} = 1 + 0.5(n_{ht}^1 - 1) + 0.3n_{ht}^0,$$

where n_h^1 is the number of members whose age is greater than or equal to 14 years and n_h^0 is the number of members under 14 years of age. INE sets up the poverty threshold at year t, z_t, as the 60 percent of the median of the NNAIs in Spain of the previous year. The Spanish poverty threshold (in euros) of 2019, 2020, and 2021 are $z_1 = 9009$, $z_2 = 9626$, and $z_3 = 9535$, respectively. Information on poverty thresholds is publicly available at the INE website: https://www.ine.es/en/prensa/ecv_2022_en.pdf.

INE calculates the SLCS inclusion probabilities, makes corrections because of non-responses and applies a calibration procedure to adjust the sample distribution to the population distribution of persons by autonomous community, age group and sex, provided by its demographic projections unit. For any sampled individual j of U_{drt}, w_{drtj} denotes its calibrated sampling weight. INE calculates these weights (elevation factors) and includes them in the SLCS data file. The direct estimator of the domain $\overline{Y}_{drt}$, is

$$\widehat{\overline{Y}}_{drt}^{dir} = \frac{1}{\widehat{N}_{drt}} \sum_{j \in s_{drt}} w_{drtj} y_{drt}, \quad \widehat{N}_{drt} = \sum_{j \in s_{drt}} w_{drtj}, \tag{2}$$

where s_{drt} is the sampled subset of domain U_{drt} and n_{drt} is the size of s_{drt}.

The direct estimators $\widehat{\overline{Y}}_{drt}^{dir}$ are used as the target variable of the new three-fold Fay–Herriot model with unequal variance random effects (4). The direct estimator of the variances of these estimators is

$$\widehat{\text{var}}_{\pi}\left(\widehat{\overline{Y}}_{drt}^{dir}\right) = \frac{1}{\left(\widehat{N}_{drt}^{dir}\right)^2} \sum_{j \in s_{drt}} w_{drtj}\left(w_{drtj} - 1\right)\left(y_{drtj} - \widehat{\overline{Y}}_{drt}^{dir}\right)^2. \tag{3}$$

By crossing the 52 Spanish provinces (including the autonomous cities of Ceuta and Melilla) with the four age groups, the two sexes and the three years, we have 1,248 areas in the SLCS data. Table 1 details the quartiles, $q_{0.0}, q_{0.25}, q_{0.5}, q_{0.75}$ and $q_{1.0}$, of the distribution of the sample sizes in the sets $\{n_{drt} : d = 1, \ldots, D\}$, $r = 1, 2$, $t = 1, 2, 3$. We observe that 90% of the samples s_{drt} have a size below 200, for both men and women. These sample sizes justify the use of model-based predictors rather than design-based direct estimators to estimate the poverty indicators of interest in all the domains.

The large number of areas and the small sample sizes make it likely that direct estimates of variances equal to zero exist with sample sizes greater than zero, that is, $\widehat{\text{var}}_{\pi}(\widehat{\overline{Y}}_{drt}^{dir}) = 0$ and $n_{drt} \neq 0$, for some $d = 1, \ldots, D, r = 1, 2, t = 1, 2, 3$. Specifically, there are 31 areas where this occurs. A notable case is the province of Soria ($d = 42$), which accounts for 13 of these 31 zeros. The areas that fulfill this condition are treated as missing data for model fitting and diagnosis. Operationally, each zero variance implies removing the six associated cross-classifications by sex and time (two sexes $\times$ three years), which would eliminate $31 \times 6 = 186$ areas in total (approximately 15% of the sample). Although smoothing the sampling variances using a generalized variance function (GVF) method is a standard approach for handling unreliable variances (You 2021), we decided not to apply this method in the present analysis. Since excluding these zero-variance areas does not affect the convergence of the model estimation algorithms in the real-data application, we prefer to retain the original design-based variances for the remaining areas.

Performance metrics such as the efficiency measure ε (18) and comparative RMSE plots (figure 12) are computed only for sampled domains with non-zero direct variance estimates, as the direct estimator serves as the benchmark.

3. THE MODEL

The basic SAE area-level models, such as the FH model, are of single-level structure with direct estimators explained by a onefold random effect and

Table 1. Area Sample Size Quartiles for Men and Women in SLCS From 2019 to 2021

t	n_{drt}	q_0	$q_{0.1}$	$q_{0.2}$	$q_{0.3}$	$q_{0.4}$	$q_{0.5}$	$q_{0.6}$	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	q_1
2019	n_{d11}	5.0	17.3	22.6	29.0	38.0	51.0	69.0	87.1	112.6	140.7	873.0
	n_{d21}	5.0	16.3	24.0	32.0	40.0	58.0	72.8	92.1	116.4	168.7	846.0
2020	n_{d12}	4.0	18.0	23.0	32.0	41.2	51.0	59.0	81.2	100.4	133.7	787.0
	n_{d22}	6.0	18.0	25.0	33.0	43.2	54.5	66.0	86.2	103.8	156.1	773.0
2021	n_{d13}	6.0	23.0	33.6	43.9	56.0	70.0	85.8	104.2	131.0	175.1	1200.0
	n_{d23}	7.0	23.3	36.0	45.0	60.2	78.0	94.0	114.2	137.0	196.4	1213.0

an independent error term. In most real-world applications, the data structure is naturally hierarchical with dependencies at multiple levels. Such a case is feasible in the estimation of the poverty proportion. We introduce a three-level Fay–Herriot model to account for hierarchical dependencies and unequal variances in the random effects at each level of aggregation. The FH3var model is

$$\begin{aligned} y_{drt} &= \boldsymbol{x}_{drt}\boldsymbol{\beta} + u_{1,d} + u_{2,dr} + u_{3,drt} + e_{drt}, \\ & d = 1,\ldots,D, r = 1,\ldots,R, t = 1,\ldots,T, \end{aligned} \tag{4}$$

where y_{drt} is a direct estimator of the characteristic of interest, $\boldsymbol{x}_{drt}$ is a row vector containing the aggregated population values of p auxiliary variables and $\boldsymbol{\beta}$ is a $p\times 1$ vector of regression parameters. We assume that $u_{1,d} \sim N(0,\sigma_1^2)$, $u_{2,dr} \sim N(0,\sigma_{2r}^2)$, $u_{3,drt} \sim N(0,\sigma_{3rt}^2)$, $e_{drt} \sim N(0,\sigma_{drt}^2)$, with $\sigma_{drt}^2 > 0$ known, $d = 1,\ldots,D$, $r = 1,\ldots,R$, and $t = 1,\ldots,T$, are mutually independent. In practice, the sampling error variances σ_{drt}^2 are unknown and must therefore be estimated. They are calculated as the design-based estimated variances from (3). This is possible using survey unit-level data and represents a crucial advantage of area-level models, as they benefit from information from the sampling design. The vector of variance parameters, $\boldsymbol{\theta} = (\sigma_1^2, \sigma_{21}^2,\ldots,\sigma_{2R}^2,\sigma_{311}^2,\ldots,\sigma_{3RT}^2)$, has $A = 1+R+RT$ components. The variance of y_{drt} is $\text{var}(y_{drt}) = \sigma_1^2+\sigma_{2r}^2+\sigma_{3rt}^2+\sigma_{drt}^2$, $d = 1,\ldots,D$, $r = 1,\ldots,R$, and $t = 1,\ldots,T$. The model (4) fits the nested population structure, with domains U_d, subdomains U_{dr} and subsubdomains U_{drt}, described in section 2. For example, d, r, and t may denote province crossed by age group, sex, and time period, respectively. The presence of heteroscedasticity in the random effects $u_{2,dr}$ and $u_{3,drt}$ makes the model (4) a generalization of the threefold Fay–Herriot model studied by Marcis et al. (2023).

At the subdomain level U_{dr}, we define the matrices and vectors

$$\boldsymbol{X}_{dr} = \underset{1\le t\le T}{\text{col}}(\boldsymbol{x}_{drt}), \boldsymbol{Z}_{1,dr} = \boldsymbol{Z}_{2,dr} = \boldsymbol{1}_T, \boldsymbol{Z}_{3,dr} = \boldsymbol{I}_T, \boldsymbol{V}_{3,dr} = \underset{1\le t\le T}{\text{diag}}\left(\sigma^2_{3rt}\right),$$
$$\boldsymbol{V}_{e,dr} = \underset{1\le t\le T}{\text{diag}}\left(\sigma^2_{drt}\right),\ \boldsymbol{y}_{dr} = \underset{1\le t\le T}{\text{col}}(y_{drt}), \boldsymbol{u}_{3,dr} = \underset{1\le t\le T}{\text{col}}\left(u_{3,drt}\right) \sim N_T\left(\boldsymbol{0}, \boldsymbol{V}_{3,dr}\right),$$
$$\boldsymbol{e}_{dr} = \underset{1\le t\le T}{\text{col}}(e_{drt}) \sim N_T\left(\boldsymbol{0}, \boldsymbol{V}_{e,dr}\right),$$

where $\boldsymbol{1}_a$ denotes the $a\times 1$ vector of ones and $\boldsymbol{I}_a$ is the $a\times a$ identity matrix. We can write model (4) in the subdomain form

$$\boldsymbol{y}_{dr} = \boldsymbol{X}_{dr}\boldsymbol{\beta} + \boldsymbol{Z}_{1,dr}u_{1,d} + \boldsymbol{Z}_{2,dr}u_{2,dr} + \boldsymbol{Z}_{3,dr}\boldsymbol{u}_{3,dr} + \boldsymbol{e}_{dr},$$
$$d = 1,\ldots,D, r = 1,\ldots,R.$$

The variance matrix of $\boldsymbol{y}_{dr}$ is

$$\boldsymbol{V}_{dr} = \text{var}(\boldsymbol{y}_{dr}) = \sigma^2_1\boldsymbol{1}_T\boldsymbol{1}'_T + \sigma^2_{2r}\boldsymbol{1}_T\boldsymbol{1}'_T + \boldsymbol{V}_{3,dr} + \boldsymbol{V}_{e,dr},$$
$$d = 1,\ldots,D, r = 1,\ldots,R.$$

At the domain level U_d, we define the vectors and matrices

$$\boldsymbol{y}_d = \underset{1\le r\le R}{\text{col}}(\boldsymbol{y}_{dr}), \boldsymbol{X}_d = \underset{1\le r\le R}{\text{col}}(\boldsymbol{X}_{dr}), \boldsymbol{Z}_{1,d} = \boldsymbol{1}_{RT}, \boldsymbol{Z}_{2,d} = \underset{1\le r\le R}{\text{diag}}(\boldsymbol{1}_T),$$
$$\boldsymbol{Z}_{3,d} = \boldsymbol{I}_{RT}, \boldsymbol{V}_{2,d} = \underset{1\le r\le R}{\text{diag}}(\sigma^2_{2r}), \boldsymbol{V}_{3,d} = \underset{1\le r\le R}{\text{diag}}\left(\underset{1\le t\le T}{\text{diag}}(\sigma^2_{3rt})\right),$$
$$\boldsymbol{V}_{e,d} = \underset{1\le r\le R}{\text{diag}}\left(\underset{1\le t\le T}{\text{diag}}(\sigma^2_{drt})\right), \boldsymbol{u}_{2,d} = \underset{1\le r\le R}{\text{col}}(u_{2,dr}) \sim N_R(\boldsymbol{0}, \boldsymbol{V}_{2,d}),$$
$$\boldsymbol{u}_{3,d} = \underset{1\le r\le R}{\text{col}}(\boldsymbol{u}_{3,dr}) \sim N_{RT}(\boldsymbol{0}, \boldsymbol{V}_{3,d}), \boldsymbol{e}_d = \underset{1\le r\le R}{\text{col}}(\boldsymbol{e}_{dr}) \sim N_{RT}(\boldsymbol{0}, \boldsymbol{V}_{e,d}).$$

We can write model (4) in the domain form

$$\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{Z}_{1,d}u_{1,d} + \boldsymbol{Z}_{2,d}\boldsymbol{u}_{2,d} + \boldsymbol{Z}_{3,d}\boldsymbol{u}_{3,d} + \boldsymbol{e}_d, \quad d = 1,\ldots,D.$$

The variance matrix of $\boldsymbol{y}_d$ is

$$\boldsymbol{V}_d = \text{var}(\boldsymbol{y}_d) = \boldsymbol{V}_{u,d} + \boldsymbol{V}_{e,d} \quad d = 1,\ldots,D,$$

where $\boldsymbol{V}_{u,d} = \sigma^2_1\boldsymbol{1}_{\text{RT}}\boldsymbol{1}'_{\text{RT}} + \text{diag}_{1\le r\le R}(\sigma^2_{2r}\boldsymbol{1}_T\boldsymbol{1}'_T) + \text{diag}_{1\le r\le R}(\text{diag}_{1\le t\le T}(\sigma^2_{3\text{rt}}))$.

At the population level, we define the vectors and matrices

$$\boldsymbol{y} = \underset{1\le d\le D}{\text{col}}(\boldsymbol{y}_d), \boldsymbol{X} = \underset{1\le d\le D}{\text{col}}(\boldsymbol{X}_d), \boldsymbol{Z}_1 = \underset{1\le d\le D}{\text{diag}}(\boldsymbol{1}_{RT}), \boldsymbol{Z}_2 = \underset{1\le d\le D}{\text{diag}}\left(\underset{1\le r\le R}{\text{diag}}(\boldsymbol{1}_T)\right),$$
$$\boldsymbol{Z}_3 = \boldsymbol{I}_{DRT}, \boldsymbol{V}_{u_1} = \sigma_1^2\boldsymbol{I}_D, \boldsymbol{V}_{u_2} = \underset{1\le d\le D}{\text{diag}}\left(\underset{1\le r\le R}{\text{diag}}(\sigma_{2r}^2)\right),$$
$$\boldsymbol{V}_{u_3} = \underset{1\le d\le D}{\text{diag}}\left(\underset{1\le r\le R}{\text{diag}}\left(\underset{1\le t\le T}{\text{diag}}(\sigma_{3rt}^2)\right)\right), \boldsymbol{V}_e = \underset{1\le d\le D}{\text{diag}}\left(\underset{1\le r\le R}{\text{diag}}\left(\underset{1\le t\le T}{\text{diag}}(\sigma_{drt}^2)\right)\right),$$
$$\boldsymbol{u}_1 = \underset{1\le d\le D}{\text{col}}(u_{1,d}) \sim N_D(\boldsymbol{0}, \boldsymbol{V}_{u_1}), \boldsymbol{u}_2 = \underset{1\le d\le D}{\text{col}}(\boldsymbol{u}_{2,d}) \sim N_{DR}(\boldsymbol{0}, \boldsymbol{V}_{u_2}),$$
$$\boldsymbol{u}_3 = \underset{1\le d\le D}{\text{col}}(\boldsymbol{u}_{3,d}) \sim N_{DRT}(\boldsymbol{0}, \boldsymbol{V}_{u_3}), \boldsymbol{e} = \underset{1\le d\le D}{\text{col}}(\boldsymbol{e}_d) \sim N_{DRT}(\boldsymbol{0}, \boldsymbol{V}_e).$$

We can write model (4) in the population form

$$\boldsymbol{y} = \boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{u}_1 + \boldsymbol{Z}_2\boldsymbol{u}_2 + \boldsymbol{Z}_3\boldsymbol{u}_3 + \boldsymbol{e}, \tag{5}$$

where the vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$ and $\boldsymbol{e}$ are mutually independent. The variance matrix of $\boldsymbol{y}$ is

$$\begin{aligned}\boldsymbol{V} = \text{var}(\boldsymbol{y}) &= \underset{1\le d\le D}{\text{diag}}(\boldsymbol{V}_d)\\ &= \sigma_1^2 \underset{1\le d\le D}{\text{diag}}(\boldsymbol{1}_{RT}\boldsymbol{1}_{RT}') + \underset{1\le d\le D}{\text{diag}}\left(\underset{1\le r\le R}{\text{diag}}(\sigma_{2r}^2\boldsymbol{1}_T\boldsymbol{1}_T')\right) + \boldsymbol{V}_{u_3} + \boldsymbol{V}_e.\end{aligned}$$

Let us define the $DRT\times(D+DR+DRT)$ matrix $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3)$ and the $(D+DR+DRT)\times 1$ vector $\boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2', \boldsymbol{u}_3')'$. We can write model (5) in the LMM form

$$\boldsymbol{y} = \boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e},$$

where $\boldsymbol{V}_u = \text{var}(\boldsymbol{u}) = \text{diag}(\boldsymbol{V}_{u_1}, \boldsymbol{V}_{u_2}, \boldsymbol{V}_{u_3})$ is the covariance matrix of the random effects. If we assume that $\sigma_1^2 > 0$, $\sigma_{2r}^2 > 0$, $\sigma_{3rt}^2 > 0$, $r = 1, \ldots, R$, and $t = 1, \ldots,$T, are known, then the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and the best linear unbiased predictor (BLUP) of $\boldsymbol{u}$ are

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= \left(\sum_{d=1}^{D}\boldsymbol{X}_d'\boldsymbol{V}_d^{-1}\boldsymbol{X}_d\right)^{-1}\left(\sum_{d=1}^{D}\boldsymbol{X}_d'\boldsymbol{V}_d^{-1}\boldsymbol{y}_d\right),\\ \tilde{\boldsymbol{u}} &= \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})\\ &= \begin{pmatrix} \sigma_1^2 \underset{1\le d\le D}{\text{col}}\left(\boldsymbol{1}_{RT}'\boldsymbol{V}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\boldsymbol{\beta}})\right) \\ \underset{1\le d\le D}{\text{col}}\left(\underset{1\le r\le R}{\text{diag}}(\sigma_{2r}^2\boldsymbol{1}_T')\boldsymbol{V}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\boldsymbol{\beta}})\right) \\ \underset{1\le d\le D}{\text{col}}\left(\underset{1\le r\le R}{\text{diag}}\left(\underset{1\le t\le T}{\text{diag}}(\sigma_{3rt}^2)\right)\boldsymbol{V}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\boldsymbol{\beta}})\right)\end{pmatrix}.\end{aligned} \tag{6}$$

The empirical versions, EBLUE of $\boldsymbol{\beta}$ and EBLUP of $\boldsymbol{u}$, are obtained by plugging estimators $\widehat{\sigma}_1^2$, $\widehat{\sigma}_{2r}^2$ and $\widehat{\sigma}_{3rt}^2$ in the place of σ_1^2, σ_{2r}^2, and σ_{3rt}^2, respectively, that is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{y}, \quad \widehat{\boldsymbol{u}} = \widehat{\boldsymbol{V}}_u\boldsymbol{Z}'\widehat{\boldsymbol{V}}^{-1}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{7}$$

where

$$\widehat{\boldsymbol{V}}_u = \operatorname{diag}\left(\widehat{\sigma}_1^2\boldsymbol{I}_D, \operatorname*{diag}_{1\le d\le D}\left(\operatorname*{diag}_{1\le r\le R}\left(\widehat{\sigma}_{2r}^2\right)\right), \operatorname*{diag}_{1\le d\le D}\left(\operatorname*{diag}_{1\le r\le R}\left(\operatorname*{diag}_{1\le t\le T}\left(\widehat{\sigma}_{3rt}^2\right)\right)\right)\right),$$

$$\widehat{\boldsymbol{V}} = \widehat{\sigma}_1^2 \operatorname*{diag}_{1\le d\le D}\left(\mathbf{1}_{RT}\mathbf{1}_{RT}'\right) + \operatorname*{diag}_{1\le d\le D}\left(\operatorname*{diag}_{1\le r\le R}\left(\widehat{\sigma}_{2r}^2\mathbf{1}_T\mathbf{1}_T'\right)\right)$$
$$+ \operatorname*{diag}_{1\le d\le D}\left(\operatorname*{diag}_{1\le r\le R}\left(\operatorname*{diag}_{1\le t\le T}\left(\widehat{\sigma}_{3rt}^2\right)\right)\right) + \boldsymbol{V}_e.$$

Appendix A (please see the supplementary materials online) describes the Fisher-scoring algorithm to calculate the REML estimators of the model parameters, and their asymptotic distributions are provided to construct the asymptotic confidence intervals (CIs) and to compute the p-values. As starting values for $\sigma_1^2, \sigma_{2r}^2, \sigma_{3rt}^2$, we propose the REML estimates under the threefold homoscedastic model of Marcis et al. (2023), that is, $\widehat{\sigma}_1^{2(0)} = \tilde{\sigma}_1^2, \widehat{\sigma}_{2r}^{2(0)} = \tilde{\sigma}_2^2, \widehat{\sigma}_{3rt}^{2(0)} = \tilde{\sigma}_3^2, r = 1,\ldots,R, t = 1,\ldots,T$.

The FH3var model provides a flexible variance structure of the random effects, although in some cases it may produce over-parametrizations. In some real data problems, a less complex model can be equally effective with fewer parameters. Thus, we consider three submodels:

(1) *Submodel 0*: $\sigma_{2r}^2 = \sigma_2^2$, $\sigma_{3rt}^2 = \sigma_3^2$, $r = 1,\ldots,R$, $t = 1,\ldots,T$. This is the FH3 model studied by Marcis et al. (2023), which assumes that the random effects u_{2dr}'s and u_{3drt}'s are homoscedastic with common variances σ_2^2 and σ_3^2, respectively. There are $A = 3$ variance parameters: $\sigma_1^2, \sigma_2^2, \sigma_3^2$.
(2) *Submodel 1*: $\sigma_{3rt}^2 = \sigma_{3t}^2$, $r = 1,\ldots,R$, $t = 1,\ldots,T$. This is the FH3var2r3t submodel, which assumes that the u_{2dr}'s are heteroscedastic with variances σ_{2r}^2, $r = 1,\ldots,R$, and the u_{3drt}'s are heteroscedastic with variances with variances σ_{3t}^2, $t = 1,\ldots,T$. There are $A = 1+R+T$ variance parameters $\sigma_1^2, \sigma_{2r}^2, \sigma_{3t}^2$, $r = 1,\ldots,R$, and $t = 1,\ldots,T$.
(3) *Submodel 2*: $\sigma_{2r}^2 = \sigma_2^2$, $r = 1,\ldots,R$. This is the FH3var23rt model, which assumes that the u_{2dr}'s are homoscedastic with common variances σ_2^2 and the u_{3drt}'s are heteroscedastic with variances depending on r and t. There are $A = 2+RT$ variance parameters: $\sigma_1^2, \sigma_2^2, \sigma_{3rt}^2$, $r = 1,\ldots,R, t = 1,\ldots,T$.

The selection of the submodel that best fits the data can be made from the FH3var model using statistical inference and diagnostic tools, ultimately

adopting one of them as the model generating the data on which to base the predictions. In any case, to use the more complex FH3var model, it is recommended that the number of subdomains DRT be much greater than the number of variances $1+R+RT$ to be estimated (for example, at least 10 times greater). This ensures that the REML log-likelihood of the model is sufficiently peaked in a neighborhood of the maximum, so that the Fisher-scoring maximizing algorithm will not encounter convergence problems when starting from that neighborhood.

3.1 EBLUP and MSE

This section considers the problem of predicting the linear combination of fixed and random effects $\mu_{drt} = \boldsymbol{x}_{drt}\boldsymbol{\beta} + u_{1,d} + u_{2,dr} + u_{3,drt}$. If the variance components are known, the BLUP of μ_{drt} is $\tilde{\mu}_{drt} = \boldsymbol{x}_{drt}\tilde{\boldsymbol{\beta}} + \tilde{u}_{1,d} + \tilde{u}_{2,dr} + \tilde{u}_{3,drt}$. The corresponding EBLUP of μ_{drt} is obtained by substituting the estimators of the variance components $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_A)$ and it has the form

$$\widehat{\mu}_{drt} = \boldsymbol{x}_{drt}\widehat{\boldsymbol{\beta}} + \widehat{u}_{1,d} + \widehat{u}_{2,dr} + \widehat{u}_{3,drt},$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{u}}$ are given in (7). The EBLUP of μ_{drt} can be also employed for estimating the sub-subdomain mean $\overline{Y}_{drt}$, in this case, the notation $\widehat{\overline{Y}}_{drt}^{eblup} = \widehat{\mu}_{drt}$ is used. In linear form, we have $\mu_{drt} = \boldsymbol{a}'(\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u})$, where $\boldsymbol{a} = \boldsymbol{a}_{drt} = \mathrm{col}_{1\le\ell\le D}(\mathrm{col}_{1\le i\le R}(\mathrm{col}_{1\le j\le T}(\delta_{d\ell}\delta_{ri}\delta_{tj})))$ is a vector having a one in the cell $t+(r-1)T+(d-1)RT$ and having zeros in the remaining cells. The symbol $\delta_{d\ell}$ denotes the Kronecker's delta, that is, $\delta_{d\ell} = 1$ if $d = \ell$ and $\delta_{d\ell} = 0$ otherwise.

Note that the EBLUPs of domain proportions based on FH models do not theoretically constrain predictions to the unit interval $[0,1]$, which could cause boundary issues given that the target variable is a proportion. However, in our real-data application, we found that the poverty proportion predictions under the FH3var model and submodels remained strictly within the valid boundaries without requiring transformation. Nevertheless, to address the theoretical concern and provide a comprehensive solution for possible cases where predictions under the model might violate these bounds, Appendix C (please see the supplementary materials online) details the formulation of the Fay–Herriot model obtained by applying the centered log-relationship (clr) transformation to the direct estimators of the proportions.

For approximating the MSE of $\widehat{\overline{Y}}_{drt}^{eblup}$, we follow the same steps as in section 17.2.3 of Morales et al. (2021). We obtain the formulas

$$
\begin{aligned}
MSE(\widehat{\overline{Y}}_{drt}^{eblup}) &= g_{1drt}(\boldsymbol{\theta}) + g_{2drt}(\boldsymbol{\theta}) + g_{3drt}(\boldsymbol{\theta}),\ g_{1drt}(\boldsymbol{\theta}) \\
&= \boldsymbol{a}'\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{a}, g_{2drt}(\boldsymbol{\theta}) \\
&= [\boldsymbol{a}'\mathbf{X} - \boldsymbol{a}'\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{V}_e^{-1}\boldsymbol{X}]\boldsymbol{Q}[\boldsymbol{X}'\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{V}_e^{-1}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{a}], g_{3drt}(\boldsymbol{\theta}) \\
&\approx \mathrm{tr}\{(\nabla \boldsymbol{b}')\boldsymbol{V}(\nabla \boldsymbol{b}')'E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})']\},
\end{aligned} \tag{8}
$$

where $\boldsymbol{T} = \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{V}_u$, $\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$, and $\boldsymbol{b}' = \boldsymbol{a}'\boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}$. The MSE of $\widehat{\overline{Y}}_{drt}^{eblup}$ can be estimated by parametric bootstrap or by using the analytic estimator

$$
mse(\widehat{\overline{Y}}_{drt}^{eblup}) = g_{1drt}(\widehat{\boldsymbol{\theta}}) + g_{2drt}(\widehat{\boldsymbol{\theta}}) + 2g_{3drt}(\widehat{\boldsymbol{\theta}}), \tag{9}
$$

where $\widehat{\boldsymbol{\theta}}$ is the REML estimator. Appendix B (please see the supplementary materials online) derives the calculations of $g_{1drt}(\widehat{\boldsymbol{\theta}})$, $g_{2drt}(\widehat{\boldsymbol{\theta}})$, and $g_{3drt}(\widehat{\boldsymbol{\theta}})$ and, as an alternative, gives a parametric bootstrap procedure.

3.2 Model Diagnostics

Statistical models are only as good as their ability to capture underlying data structures while avoiding overfitting, excessive influence from outliers or violations of key assumptions. This section draws on the diagnostic measures introduced and studied by Marcis et al. (2023). It focuses on four main aspects of model validation: (1) leverage analysis, identifying areas with excessive influence; (2) domain deletion assessment, evaluating robustness to domain-specific effects; (3) residual diagnostics, checking model assumptions; and (4) model efficiency and error reduction, analyzing the goodness-of-fit.

From an applied perspective, these diagnostic tools are essential for evaluating model robustness and detecting anomalies that may introduce bias into the estimates. Residual analysis, particularly using Cholesky residuals, is the most informative tool for verifying core model assumptions such as normality. Cook's distance and leverage points are essential for detecting influential domains or subdomains that may disproportionately affect parameter estimates; while leverages distinguish between the impact on fixed and random effects, Cook's distance provides a holistic measure of influence due to extreme residual points. Finally, efficiency measures, such as the correlation between conditional residuals and errors, are crucial for measuring the extent of "borrowing of strength" and confirming whether the EBLUP yields a substantial reduction in variability relative to direct estimators. In applied works, we recommend a sequential approach: first confirming model validity through residuals,

then identifying influential cases with Cook's distance, and finally assessing practical gains through efficiency metrics.

3.2.1 Leverages.
In mixed models, certain small areas may induce excessive influence in the estimation of the model parameters. Identifying such influential observations is necessary to prevent excessive bias in model-based predictions. Following Demidenko and Stukel (2005) and Zewotir and Galpin (2007), the leverage matrix of domain d of the model (4) is given by

$$\boldsymbol{L}_d = \boldsymbol{L}_d(\widehat{\boldsymbol{\theta}}) = \frac{\partial \widehat{\boldsymbol{\mu}}_d}{\partial \boldsymbol{y}_d} = \frac{\partial \boldsymbol{X}_d \widehat{\boldsymbol{\beta}}}{\partial \boldsymbol{y}_d} + \frac{\widehat{\boldsymbol{u}}_d}{\partial \boldsymbol{y}_d} \triangleq \boldsymbol{L}_{f,d} + \boldsymbol{L}_{r,d}, \tag{10}$$

where $\boldsymbol{L}_{f,d}$ and $\boldsymbol{L}_{r,d}$ denotes the corresponding fixed-effect and random-effect leverages. Computing the corresponding partial derivates, $\boldsymbol{L}_{f,d}$ and $\boldsymbol{L}_{r,d}$ are given by

$$\boldsymbol{L}_{f,d} = \boldsymbol{X}_d(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}_d'\boldsymbol{V}_d^{-1}, \boldsymbol{L}_{r,d} = \boldsymbol{Z}_d\boldsymbol{V}_{ud}\boldsymbol{Z}_d'\boldsymbol{V}_d^{-1}(\boldsymbol{I} - \boldsymbol{L}_{f,d}). \tag{11}$$

By averaging the diagonal elements of the leverage matrices, can be defined fixed-effects and random-effects average leverages for domain d, that is,

$$L_{f,d} = \frac{1}{RT}\mathrm{tr}(\boldsymbol{L}_{f,d}), L_{r,d} = \frac{1}{RT}\mathrm{tr}\ (\boldsymbol{L}_{r,d}), \tag{12}$$

where $\mathrm{tr}(\boldsymbol{M})$ denotes the trace of the square matrix $\boldsymbol{M}$. Following Demidenko and Stukel (2005) and Marcis et al. (2023), the usual cut-off values for the leverage of fixed-effects and random-effects are, respectively,

$$\mathrm{co}_f = \frac{2}{D}\sum_{d=1}^{D}\mathrm{tr}(\boldsymbol{L}_{f,d}) = \frac{2p}{D},\ \mathrm{co}_r = \frac{2}{D}\sum_{d=1}^{D}\mathrm{tr}(\boldsymbol{L}_{r,d}). \tag{13}$$

Exceeding these cut-off points indicates a substantial influence of the direct estimator on the predicted value of the EBLUP vector $\widehat{\boldsymbol{\mu}}_d$, which may disproportionately impact, therefore, the model's inference.

Figures 7 and 8 illustrate the use of fixed-effect and random-effect leverages as diagnostic tools in application to real data.

3.2.2 Domain deletion.
The impact of individual domains on model stability is assessed by computing the effect of domain deletion on fixed effect estimates. Specifically, when domain U_d is omitted, the shift in the estimated coefficient vector $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(d)} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}_d\boldsymbol{V}_d^{-1}(\boldsymbol{I} - \boldsymbol{L}_{f,d})^{-1}\boldsymbol{r}_d^m,$$

where the subscript (d) is used to denote that the estimator has been calculated with the aggregated data from $U - U_d$ and $\boldsymbol{r}_d^m = \boldsymbol{y}_d - \boldsymbol{X}_d\widehat{\boldsymbol{\beta}}$ is the vector of marginal residuals. To quantify the degree of influence, the Cook's distance is

$$\Delta_d = \frac{1}{\text{rank}(\boldsymbol{X})}(\boldsymbol{r}_d^m)'(\boldsymbol{I} - \boldsymbol{L}_{f,d})^{-1}\boldsymbol{V}_d^{-1}\boldsymbol{X}_d(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(d)}). \tag{14}$$

The most used cutoff values in linear regression models are $\Delta_d > \frac{4}{D}$ (fixed-effects) or $\Delta_d > \frac{4}{D-p}$ (mixed-effects). Additionally, comparing the MSE estimates of the EBLUPs before and after domain removal provides valuable information about the robustness of the model-based predictions, that is,

$$\overline{mse}(\widehat{\theta}) = \frac{1}{DRT}\sum_{\ell=1}^{D}\sum_{r=1}^{R}\sum_{t=1}^{T} mse(\widehat{\mu}_{\ell rt}(\widehat{\boldsymbol{\theta}})),$$

$$\overline{mse}(\widehat{\theta}_{(d)}) = \frac{1}{(D-1)RT}\sum_{\ell=1,\ell\neq d}^{D}\sum_{r=1}^{R}\sum_{t=1}^{T} mse(\widehat{\mu}_{\ell rt}(\widehat{\boldsymbol{\theta}}_{(d)})).$$

Figure 7 (left) illustrates the use of the Cook's distance in detecting the most influential areas in application to real data.

3.2.3 Residuals.

Residuals are used to examine model assumptions and to detect outliers and potentially influential data points. Although the true model errors have zero mean and constant variance, this may not be true for the residuals. They may show correlation, and their variances may be different. In general, and especially concerning linear mixed-effects models, various types of studentization are used (Calvin and Sedransk 1991; Nobre and Singer 2007). For the threefold Fay–Herriot with unequal variances models—that are block-diagonal LMMs with correlated observations—we employ the analysis of the model-scaled residuals, after using the Cholesky root of the covariance matrix of the vector of target variables.

This approach, also used by Jacqmin-Gadda et al. (2007) in the context of LMMs, involves premultiplying the model by the Cholesky root of the covariance matrix to obtain residuals derived from this decomposition. As highlighted by Rao and Molina (2015), this method is advantageous because the residuals are approximately uncorrelated, allowing the residual plots to be interpreted similarly to those in linear regression models (figure 6). With this method, the obtained model has a uniform dispersion matrix.

Given the block-diagonal covariance matrix, $\widehat{\boldsymbol{V}} = \boldsymbol{V}(\widehat{\boldsymbol{\theta}})$, and the corresponding Cholesky root, $\widehat{\boldsymbol{V}} = \widehat{\boldsymbol{C}}'\boldsymbol{C}$, the transformed model is

$$\boldsymbol{y}_{ch} = \boldsymbol{C}'^{-1}\boldsymbol{y} = \overline{\boldsymbol{X}}\boldsymbol{\beta}_{ch} + \boldsymbol{\epsilon}. \tag{15}$$

where $\overline{\boldsymbol{X}} = \widehat{\boldsymbol{C}}'^{-1}\boldsymbol{X}, \boldsymbol{\beta}_{ch} = \widehat{\boldsymbol{C}}'^{-1}\boldsymbol{\beta}, \boldsymbol{\epsilon} = \widehat{\boldsymbol{C}}'^{-1}(\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e})$. The residuals and the studentized residuals of the "reduced" model (15) are

$$r^m_{ch,drt} = \widehat{\boldsymbol{C}}'^{-1} r^m_{drt} = \widehat{\boldsymbol{C}}'^{-1}(y_{drt} - \boldsymbol{x}'_{drt}\widehat{\boldsymbol{\beta}}), \tilde{r}^m_{ch,drt} = \frac{r^m_{ch,drt}}{\sqrt{v_{drt}}}, \tag{16}$$

where v_{drt} is the diagonal element of the matrix $\widehat{\sigma}_\epsilon(drt)^2\left(\boldsymbol{I} - \overline{\boldsymbol{X}}(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}})^{-1}\overline{\boldsymbol{X}}'\right)$ and $\widehat{\sigma}_\epsilon(drt)^2$ is the estimated variance of the model (15), treated as a fixed-effect linear model.

Exploring Cholesky residuals can also be a valuable approach to discriminate the best model among several candidates. This can be achieved by comparing standard tests for heteroscedasticity (White's test and Breusch—Pagan test, Wooldridge 2016) and independence (Durbin–Watson test, Wooldridge 2016) as in the regression analysis—as well as by employing the Locally Estimated Scatterplot Smoothing (LOESS) method (Harrell 2015). The latter may serve as an effective tool to identify which model demonstrates greater regularity of the residuals, allowing for the control of which model is furthest from showing evidence of patterns in the residuals and from a non-linear condition.

Figures 4, 5, 6, and 8 illustrate the use of residuals to analyze the predictive level of the model, examine the normality hypothesis, or detect anomalous areas.

3.2.4 MSEs and residuals' correlations.

In the context of small area estimation, the reduction of the MSE is of great importance as it ensures more accurate estimates. The comparison between the real parameter to be estimated and its prediction is possible when simulations are performed. If a real case is analyzed, however, it makes sense to understand whether the predicted MSE of the EBLUP has reduced the variability of the direct estimator. A useful and practical approach proposed by Marcis et al. (2023) is to employ a measure based on the correlation between the conditional residuals $r^c_{drt} = y_{drt} - \tilde{\mu}_{drt}$ and the errors e_{drt}, called $\boldsymbol{\epsilon}$. In particular, we can focus on the diagonal elements, $c_{drt,drt}$, of the matrix $C = V_e P V_e = cov(r^c, e)$, that is $c_{drt,drt} = cov(r^c_{drt}, e_{drt}) = \sigma^2_{drt} - g_{1drt}(\boldsymbol{\theta}) - g_{2drt}(\boldsymbol{\theta}), d = 1, \ldots, D, r = 1, \ldots, R$, $t = 1, \ldots, T$. The correlation between conditional residuals and errors with respect to the diagonal elements of $\boldsymbol{C}$ is given by the ratio

$$corr(r^c_{drt}, e_{drt}) = \varepsilon_{drt} = \frac{cov(r^c_{drt}, e_{drt})}{\sqrt{var(r^c_{drt})var(e_{drt})}} = \frac{c_{drt,drt}}{\sqrt{c_{drt,drt}\sigma^2_{drt}}} = \sqrt{\frac{c_{drt,drt}}{\sigma^2_{drt}}}. \tag{17}$$

Then, the measure of the model efficiency, $\boldsymbol{\epsilon}$ is given by

$$\boldsymbol{\epsilon} = \frac{1}{DRT}\sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{t=1}^{T} \text{corr}(r^c_{drt}, e_{drt}) = \frac{1}{DRT}\sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{t=1}^{T}\varepsilon_{drt}. \tag{18}$$

Note that when the correlation between conditional residuals and error tends to one, not only $c_{drt,drt} \to \sigma^2_{drt}$, but the conditional residuals tend to the direct estimator. This means that the drtth small area contributes more to the reduction of the MSE, and therefore, it reaches the best performance of the model regarding the decrease in the expected MSE of the EBLUP. Taking the MSE estimator $mse(\widehat{\mu}_{drt}) = g_{1drt}(\widehat{\boldsymbol{\theta}}) + g_{2drt}(\widehat{\boldsymbol{\theta}}) + 2g_{3drt}(\widehat{\boldsymbol{\theta}})$ from section 3.1, can be defined the mse-ratio:

$$\pi_{drt} = \frac{2g_{3drt}(\widehat{\boldsymbol{\theta}})}{g_{1drt}(\widehat{\boldsymbol{\theta}}) + g_{2drt}(\widehat{\boldsymbol{\theta}})}, d = 1, \ldots, D, r = 1, \ldots, R, t = 1, \ldots, T. \tag{19}$$

A scatter plot of π_{drt} versus ε_{drt} jointly indicates the efficiency of the model and the contribution of the variability of the variance component estimators to the estimation of the MSEs of the EBLUPs $\widehat{\mu}_{drt}$.

Figures 9 and 10 show the use of efficiency measures in Bland–Altman plots. Table 9 shows the efficiency of the FH3, FH3var23rt, and FH3var models, providing evidence in favor of the FH3var23rt model. Figure 11 presents two plots of mse-ratios versus efficiencies, explaining the conclusions drawn from them.

4. SIMULATIONS

This section presents three simulation experiments to study the behavior of the Fisher scoring algorithm that calculates the REML estimators of the model parameters, the EBLUPs of linear indicators, and the analytic and bootstrap estimators of the MSEs. The simulations evaluate the performance of the new statistical methodology under different data-generating processes. We are primarily interested in investigating the impact of heteroscedasticity on the bias and MSE of the EBLUPs, including the impact of model misspecification. The question to be investigated is what happens if we use predictors based on a model with an incorrect variance structure. The study includes both homoscedastic and heteroscedastic cases to determine which specification produces the most accurate and reliable

predictors, while also demonstrating the consequences of incorrect model assumptions. For $d = 1, \ldots, D$, $r = 1, \ldots, R$, $t = 1, \ldots, T$, the explanatory and dependent variables are

$$x_{drt} = \frac{1}{5}(b_{drt} - 1)U_{drt} + 1, U_{drt} = \frac{t}{T+1}, b_{drt} = 1 + \frac{1}{DR}((d-1)RT + (r-1)T + t),$$

$$y_{drt} = \beta_1 + \beta_2 x_{drt} + u_{1,d} + u_{2,dr} + u_{3,drt} + e_{drt}, \beta_1 = 1, \beta_2 = 1,$$

$$u_{1,d} \sim N(0, \sigma_1^2), u_{2,dr} \sim N(0, \sigma_{2r}^2), u_{3,drt} \sim N(0, \sigma_{3rt}^2), e_{drt} \sim N(0, \sigma_{drt}^2), \sigma_1^2 = 0.5,$$

$$\sigma_{drt}^2 = 0.8 + \frac{2}{25}\frac{1}{DRT-1}((d-1)RT + (r-1)T + t - 1),$$

where $R = 2$ and $T = 3$ imitating the application to real data. For the variances of $u_{2,dr}$ and $u_{3,drt}$, $d = 1, \ldots, D$, $r = 1, 2$, $t = 1, 2, 3$, we consider three different scenarios:

S0 Homoscedasticy: $\sigma_{2r}^2 = 0.5$, $\sigma_{3rt}^2 = 0.5$.
S1 Low of heteroscedasticity: $\sigma_{2r}^2 = 0.5 + 0.1(r-1)$, $\sigma_{3rt}^2 = 0.5 + 0.1(r-1) + 0.05(t-1)$.
S2 High heteroscedasticity: $\sigma_{2r}^2 = 0.5 + 1(r-1)$, $\sigma_{3rt}^2 = 0.5 + 1(r-1) + 1(t-1)$.

4.1 Simulation 1

Simulation 1 investigates the behavior of the Fisher scoring algorithm under Scenarios S1 and S2. It is carried out with $I = 10^3$ Monte Carlo replicates. Tables 2 and 3 present the empirical relative biases (RBIAS) and relative root-MSEs (RRMSE), in %, of the REML estimators of the model parameters $\widehat{\tau} \in \{\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}_1, \widehat{\sigma}_{2r}, \widehat{\sigma}_{3rt}; r = 1, \ldots, R, t = 1, \ldots, T\}$ for $D = 100$, $200, 300, 400$, under Scenarios S1 and S2, respectively.

Tables 2 and 3 show that the relative biases and root-MSEs tend to decrease when D increases. This result agrees with asymptotic theory, since REML estimators are consistent in LMMs. Appendix D.1 (please see the supplementary materials online) shows the simulation steps and provides the absolute biases and the absolute root-MSEs.

4.2 Simulation 2

Simulation 2 investigates the behavior of the EBLUPs based on models FH3, FH3var, FH3var2r3t, and FH3var23rt when data are generated under scenarios S0, S1, and S2, respectively. Under scenario S0, the correct model is FH3. Under scenario S2, the correct model is FH3var. Under the low-heteroscedasticity scenario S1, all models would be potentially usable

Table 2. RBIAS and RRMSE, in %, of REML Estimators Under Scenario S1

	$RBIAS(\widehat{\tau})$				$RRMSE(\widehat{\tau})$			
τ	$D=100$	$D=200$	$D=300$	$D=400$	$D=100$	$D=200$	$D=300$	$D=400$
β_1	0.6791	0.1029	−0.6574	−0.4258	23.1380	16.2542	13.9191	11.5572
β_2	−0.4685	−0.0456	0.1393	0.2249	11.9174	8.3228	7.2169	5.9444
σ_1^2	−0.7232	0.4274	0.4946	0.1108	31.8704	23.4651	18.6762	16.9127
σ_{21}^2	2.1837	0.2638	−0.7568	−0.5168	41.4605	28.4546	23.9198	20.5227
σ_{22}^2	0.9528	0.3645	−0.6458	−1.1723	36.8422	25.8675	20.6950	18.5397
σ_{311}^2	−0.3121	−0.1530	−0.3792	1.4796	48.2499	34.5727	28.4646	24.6514
σ_{312}^2	1.2572	−0.9560	−0.5451	−0.1439	44.5124	32.1666	26.4375	22.5205
σ_{313}^2	−2.4125	1.5237	−0.6613	−0.3802	43.2892	30.1114	24.2555	21.7431
σ_{321}^2	1.3844	0.9043	−0.1629	−0.2605	44.8081	31.4625	25.2098	22.5346
σ_{322}^2	−1.0191	−0.3668	1.0008	0.7106	42.1144	29.6643	24.2521	21.2184
σ_{323}^2	−1.3594	0.2445	0.1200	0.0257	39.8513	29.5752	22.6983	19.6132

Table 3. RBIAS and RRMSE, in %, of REML Estimators Under Scenario S2

	$RBIAS(\widehat{\tau})$				$RRMSE(\widehat{\tau})$			
τ	$D=100$	$D=200$	$D=300$	$D=400$	$D=100$	$D=200$	$D=300$	$D=400$
β_1	−1.6848	−0.2888	−0.0170	−0.4664	29.7447	20.4454	17.4801	14.0854
β_2	0.7456	−0.0782	0.0339	0.1640	16.3117	11.3198	9.5701	7.9702
σ_1^2	−4.4803	−2.0095	0.9699	0.1150	43.4366	31.6241	26.0253	23.5242
σ_{21}^2	11.5887	2.8046	−0.6110	1.7728	49.9465	37.7729	30.7476	28.3611
σ_{22}^2	−0.4524	0.7952	0.1260	−0.5634	28.9712	20.3248	17.9581	14.3463
σ_{311}^2	7.8082	1.6741	2.1256	−1.5436	51.5259	37.7442	33.2604	27.3091
σ_{312}^2	0.3896	0.2859	0.2255	0.2008	25.9730	19.0602	15.0332	13.3837
σ_{313}^2	0.4053	−0.7698	0.0016	0.1996	21.1694	15.2186	12.2363	11.1874
σ_{321}^2	−2.1649	−2.2154	0.6337	−0.0425	31.6744	22.7272	19.1580	15.6531
σ_{322}^2	−0.9696	−1.1283	−0.1384	0.1573	24.3199	17.4059	14.6568	12.5428
σ_{323}^2	−1.1519	0.3063	0.0010	−0.2260	21.4199	15.2822	12.1926	10.3195

as working models. Simulation 2 analyzes the effect of heteroscedasticity on EBLUPs based on models that do or do not incorporate heteroscedasticity. Simulation 2 calculates averages across sub-subdomains of the empirical average absolute relative biases (AARBIAS) and average relative root-MSEs (ARRMSE), in %, with $I = 10^3$ Monte Carlo replicates.

Tables 4 and 5 present the AARBIAS (top) and the ARRMSE (bottom) for the homoscedastic Scenario S0 and the heteroscedastic Scenarios S1 and S2, respectively. Under Scenario S0, the best performance is achieved by the EBLUP based on model FH3. However, heteroscedastic

Table 4. AARBIAS (Top) and ARRMSE (Bottom), in %, of EBLUPs Under Scenario 0

EBLUP	$D = 100$	$D = 200$	$D = 300$	$D = 400$
FH3	0.6404	0.6119	0.6291	0.6217
FH3var2r3t	0.6452	0.6170	0.6304	0.6221
FH3var23rt	0.6518	0.6190	0.6305	0.6223
FH3var	0.6550	0.6188	0.6309	0.6224
FH3	24.8539	24.7563	24.7260	24.7179
FH3var2r3t	24.9523	24.8104	24.7619	24.7436
FH3var23rt	25.0564	24.8708	24.8034	24.7739
FH3var	25.0748	24.8798	24.8093	24.7780

Table 5. AARBIAS (Top) and ARRMSE (Bottom), in %, of EBLUPs Under Scenarios 1 and 2

	Scenario 1				Scenario 2			
EBLUP	$D = 100$	$D = 200$	$D = 300$	$D = 400$	$D = 100$	$D = 200$	$D = 300$	$D = 400$
FH3	0.6428	0.6422	0.6522	0.6277	0.8083	0.7767	0.7465	0.7554
FH3var2r3t	0.6432	0.6419	0.6524	0.6293	0.8069	0.7655	0.7338	0.7460
FH3var23rt	0.6504	0.6426	0.6537	0.6301	0.8157	0.7665	0.7336	0.7436
FH3var	0.6494	0.6419	0.6530	0.6302	0.8144	0.7652	0.7299	0.7431
FH3	25.5543	25.5537	25.5365	25.5120	30.2801	30.2756	30.2201	30.2057
FH3var2r3t	25.6404	25.5962	25.5612	25.5263	29.9378	29.9132	29.8434	29.8274
FH3var23rt	25.7432	25.6418	25.5876	25.5441	29.8497	29.8083	29.7237	29.6957
FH3var	25.7589	25.6494	25.5914	25.5464	29.8140	29.7643	29.6809	29.6543

specifications (FH3var, FH3var2r3t, and FH3var23rt) do not perform markedly worse under homoscedasticity, since the more complex variance structures ought to accommodate the constant-variance case as a special configuration. Under Scenario S1, we confirm that all models can be potentially used as working models, as the differences are negligible. Finally, under Scenario S2, the EBLUPS based on the models that take into account the heteroscedasticity obtain the best results. Although the efficiency gains are modest, the proposed FH3var models enable a crucial balance between predictive performance and model interpretability. These findings demonstrate that it is possible to rigorously capture complex variance structures to improve accuracy without sacrificing the explanatory power of the LMM family.

These results are in line with expectations, as the EBLUPs based on models closest to the data-generating model are those that obtain the best

results. Appendix D.1 (please see the supplementary materials online) shows the simulation steps and provides the absolute biases and root-MSEs.

In summary, Simulation 2 provides two key practical insights for heterogeneity under FH3var models. First, the proposed heteroscedastic models offer advantages when heterogeneity is present in the variance components, leading to more accurate and reliable predictors. Second, these models remain competitive even when the true data-generating process is homoscedastic, showing negligible loss of efficiency compared to the standard homoscedastic FH3 model. These findings demonstrate that the FH3var framework allows for capturing complex variance structures without compromising accuracy under simpler conditions, effectively balancing model flexibility with predictive performance.

4.3 Simulation 3

Simulation 3 studies the behavior of the MSE estimators under Scenario S2 for $D = 100$. The goal is to compare the accuracy of the analytical and parametric bootstrap MSE estimators when applied to linear predictors. More concretely, the analyzed estimators are $mse(\widehat{\mu}_{drt})$ and $mse^*(\widehat{\mu}_{drt})$, described in appendix B (please see the supplementary materials online). We investigate the bias and the MSE of the MSE estimators. For the bootstrap-based estimator, we explore the influence of the number B of bootstrap iterations on the stability of the estimator. To achieve this goal, the two estimators are compared with the empirical MSE of $\widehat{\mu}_{drt}$ obtained from the output of Simulation 2. The performance measures are the averages across sub-subdomains of the empirical average absolute relative biases (AARB) and average relative root-MSEs (ARRE), in %, with $I = 10^3$ Monte Carlo replicates.

Table 6 shows the obtained AARBs and ARRE for the analytic ($AARB^0$, $ARRE^0$) and the parametric bootstrap ($AARB^1$, $ARRE^1$) estimators of the MSEs of the EBLUPs. The number of bootstrap replicates is $B = 50, 100, 200, 400$. Whereas the analytical approximation remains stable as B increases, as it does not depend on the bootstrap replicates, bootstrap-based MSE estimates exhibit high variability for small B values. Increasing B to at least 400 iterations significantly reduces the relative error, aligning bootstrap estimates with the analytical approximation. Therefore, we recommend at least $B = 400$ replicates for estimating the MSE by parametric bootstrap. Appendix D.2 (please see the supplementary materials online) shows the simulation steps and provides the absolute biases and root-MSEs.

Figure 1 presents the boxplots of the relative biases (RB_{drt}) and the relative root-MSEs RRE_{drt}) of $mse(\widehat{\mu}_{drt})$ (Analytic), and $mse^*(\widehat{\mu}_{drt})$, $d = 1, \ldots, D$, $r = 1, 2$, $t = 1, 2, 3$ for $B = 50, 100, 200, 400$ bootstrap replicates. The same analysis is observed as the discussed one of table 6.

Table 6. Relative Performance Measures $AARB^{a}$ and $ARRE^{a}$, $a=0,1$

Measure	$B=50$	$B=100$	$B=200$	$B=400$
$AARB^{0}$	3.6553	3.6553	3.6553	3.6553
$AARB^{1}$	3.9675	3.8750	3.7966	3.7685
$ARRE^{0}$	11.8722	11.8722	11.8722	11.8722
$ARRE^{1}$	23.4473	18.8434	15.9973	14.3562

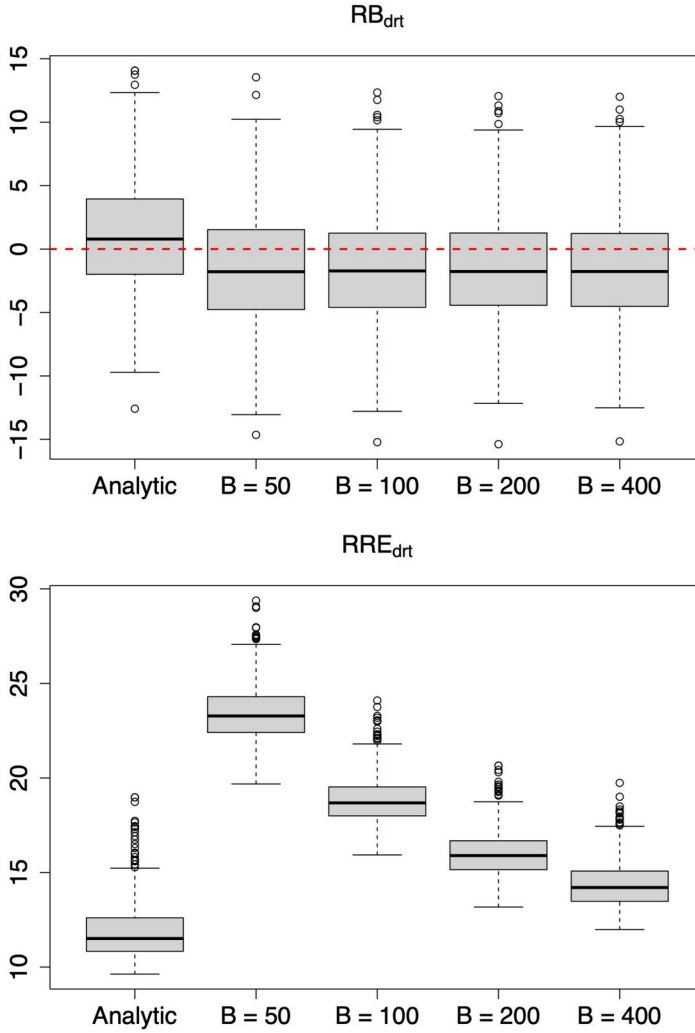

Figure 1. Boxplots of RB^{a}_{drt} and RRE^{a}_{drt}, $a=0,1$ for $B=50,100,200,400$.

To assess the computational efficiency of the proposed estimators, computational times were monitored using a MacBook Pro equipped with an 8-core Apple M3 chip and 16 GB of RAM. Computing the analytic MSE estimator proved extremely efficient, taking approximately 2.35 seconds. In contrast, the parametric bootstrap estimator took an average of 1.68 seconds per bootstrap run. Thus, for $D = 100$ and $B = \{50, 100, 200, 400\}$, the total computation times for each Monte Carlo iteration are approximately 1.4, 2.8, 5.6, and 11.25 minutes, respectively.

5. CASE STUDY: POVERTY PROPORTION IN SPAIN

This section applies the new statistical methodology to SLCS data from 2019 to 2021, publicly available on the INE website. The primary objective is to estimate poverty proportions (i.e., the proportion of individuals falling below the poverty line) within distinct crosses of Spanish provinces and age groups (domains), sexes (subdomains), and years (sub-subdomains), by using predictors based on the FH3var model defined in (4). Our analysis covers $D = 208$ crosses of provinces by age group (domains), $R = 2$ sexes and $T = 3$ years. The response variable, y_{drt}, is the Hájek direct estimator of the poverty proportion, while the error variance σ^2_{drt} is the estimated design-based variance of the direct estimator, as defined in (3). The hierarchical structure of the data makes the FH3var model suitable for this analysis. The FH3var model introduces a multilevel framework in which variability is specified at different levels of aggregation, while the standard FH model assumes homoscedasticity within domains and overlooks nested dependencies, leading to inefficient estimation of the model variances $\text{var}(y_{drt})$.

Figure 2 presents six maps of the direct estimates of poverty proportions for the first age group across the Spanish provinces of men (left) and women (right) from 2019 (top) to 2021 (bottom). This socio-economic context motivates the use of the FH3var-type models to capture the different variability of the target variable by province-age group, sex, and year. To address the precision limitations of direct estimators, we apply model-based statistical methodology to incorporate auxiliary information that explains the behavior of the target variable and captures the variability within threefold clusters. We take data from the Spanish Labour Force Survey (SLFS) to construct a data file of aggregated auxiliary variables, including demographic characteristics, socioeconomic status, education, and immigration factors.

The SLFS adopts an annual survey format with a rotating panel design, featuring a sample composed of four independent subsamples, each forming a four-year panel. The sample is renewed annually within one of the panels. Subsample selection utilizes a two-stage design with initial unit

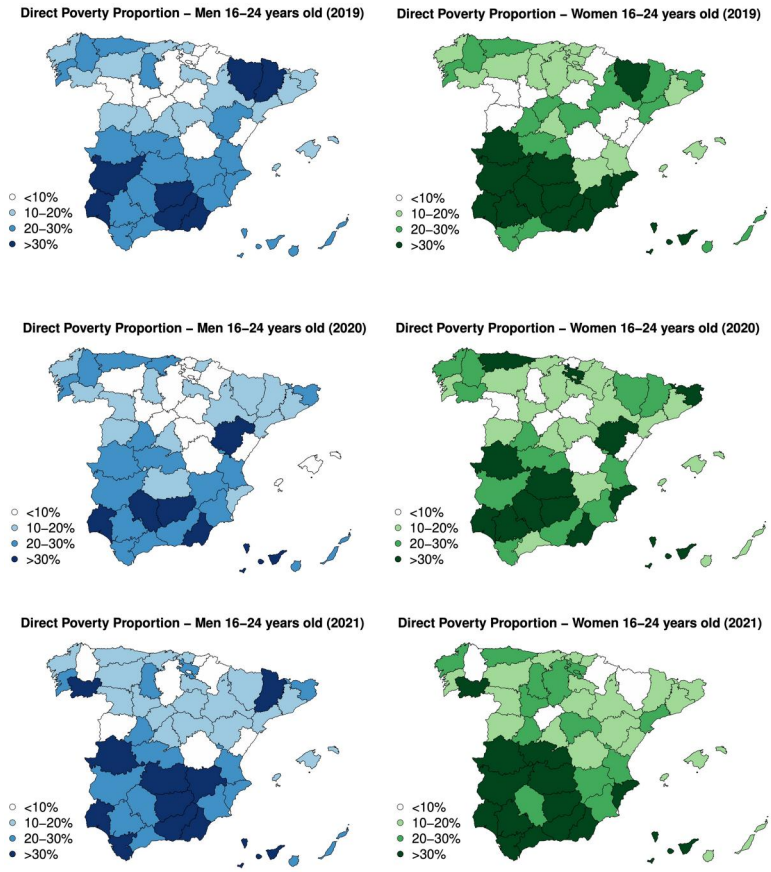

Figure 2. Maps of the Direct Estimates of the Poverty Proportion for Men (Left) and Women (Right) Between 16 and 44 Years Old, From 2019 (Up) and 2021 (Bottom).

stratification involving census sections in the first stage and main family dwellings in the second stage. The quarterly SLFS sample size exceeds that of an annual SLCS sample. Moreover, for enhanced estimate precision, we collectively take data from the four quarters spanning 2019–2021 to compute auxiliary variables, employing the formula of the Hájek direct estimator for domain means. Constructed auxiliary variables include proportions of individuals categorized by Education (*edu*0: less than primary, *edu*1: primary, *edu*2: basic secondary, *edu*3: bachelor and/or university), Labor situation (*lab*1: employed, *lab*2: unemployed, *lab*3: inactive), Nationality (*nac*1: Spanish, *nac*2: Not Spanish) and Professional status (*st*1: employer or independent worker, *st*2: public sector salaried employee, *st*3: public sector salaried employee, *st*4: private salaried

employee, *st*5: others). To maintain the proportion summation to one, reference categories (*edu*0, *lab*3, *nac*2, *st*5) are excluded from the data set of aggregated auxiliary variables.

The explanatory variables in our study have measurement errors that are smaller than the sampling errors. This is because the SLFS sample sizes used to calculate the $\boldsymbol{x}_{drt}$'s are, on average, 10 times larger than the corresponding SLCS sample sizes used to calculate the y_{drt}'s. We computed the annual values by averaging the four quarterly SLFS for each year. This strategy was employed to minimize sampling variability in the explanatory variables as much as possible, ensuring that the primary source of uncertainty in the model arises from the target variable rather than the predictors. Extending our research to incorporate methodologies proposed by Ybarra and Lohr (2008) and Burgard et al. (2020, 2022) within the context of the FH3var model (4) would require extensive efforts beyond the scope of our current research objectives. Similarly, other alternatives discussed in the literature, such as spatial Fay–Herriot models that explicitly account for geographical correlation (Pratesi and Salvati 2008), robust approaches designed to handle outliers (Sinha and Rao 2009), or multilevel time-series models based on state-space frameworks (Boonstra and Van Den Brakel 2022), represent distinct modelling paradigms. While these approaches offer robust tools for spatial dependencies and time-series data, their application would differ fundamentally from the model approach adopted here. Our work prioritises structural modelling of hierarchical heteroscedasticity and therefore, focuses on model comparison for the FH3var family. This section serves as an illustrative application of the statistical methodology outlined in section 3.

5.1 Parameter Estimation and Model Selection

This section mainly focuses on the findings derived from applying the FH3var23rt fitting model. The selection of this model is based on its outperformance in terms of significant regression parameters and variance components, as well as a residual analysis (section 3.2). The detailed results of the remaining variants of the FH3var model (FH3, FH3var2r3t, and FH3var), presented in section 3, are provided in appendix E.1 (please see the supplementary materials online).

To fit the FH3var model and submodels to the poverty proportions y_{drt}, those auxiliary variables that were not significant were removed. The model identified two significant auxiliary variables: *lab2* (proportion of the unemployed population) and *st4* (proportion of private salaried employees).

Table 7 presents the REML estimates of the regression parameters ($\boldsymbol{\beta}$), the standard error (SE), the *P*-values and the logical statements of whether (T) or not (F) the zero belongs to the 90% asymptotic CI.

Table 7. Regression Parameters of the FH3 Model and Submodels

	FH3				FH3var2r3t			
Variable	$\boldsymbol{\beta}$	SE	p-value	$0 \in$ CI	$\boldsymbol{\beta}$	SE	p value	$0 \in$ CI
Intercept	0.1427	0.0091	.000	F	0.1416	0.0090	.000	F
*lab*2	0.5866	0.0825	.000	F	0.5660	0.0815	.000	F
*st*4	−0.0939	0.0267	.000	F	−0.0939	0.0269	.000	F
	FH3var23rt				FH3var			
Variable	$\boldsymbol{\beta}$	SE	p-value	$0 \in$ CI	$\boldsymbol{\beta}$	SE	p-value	$0 \in$ CI
Intercept	0.1442	0.0091	.000	F	0.1415	0.0089	.000	F
*lab*2	0.5788	0.0825	.000	F	0.5766	0.0815	.000	F
*st*4	−0.0880	0.0267	.000	F	−0.1016	0.0269	.000	F

Under linear models, the sign of the coefficients $\boldsymbol{\beta}$ reveals information about the relationship between the target variable and the auxiliary variable, assuming the remaining ones are fixed. From table 7, we observe a positive relationship between poverty proportion and unemployed people proportion *lab2*, which is in line with the socio-economic theory, since job loss reduces household incomes and raises the risk of poverty. By contrast, the negative coefficient for *st4* indicates that areas with a larger private-sector workforce tend to exhibit lower poverty. This could be explained by the facts that: (1) stronger labor demand and higher employment intensity in the private sector increase the probability of work and reduce income risk, (2) a sectoral composition with more tradable and higher-productivity activities generally raises average salaries and pulls up the lower tail of the income distribution.

All the p-values are lower than $\alpha = 0.1$, both for the intercept and the considered auxiliary variable. The 90% asymptotic CIs of the regression parameters show that all the regression parameters differ significantly from zero.

Table 8 presents the estimates of the model variances, the standard errors, the p-values, the lower (Low.) and upper (Upp.) bounds of the 90% asymptotic CI and the logical statements of whether (T) or not (F) the zero belongs to the 90% asymptotic CI. Two out of six variances are not significant. Table E.1 of appendix E (please see the supplementary materials online) shows the same results for the rest of the models. All of them have at least two non-significant model variances, except the FH3 model. Although two of the eight variance parameters were not found to be statistically significant, the presence of significant heteroscedasticity at the third hierarchical level (with four of the six components proving

Table 8. Variance Components Estimates Under the FH3var23rt Model

Parameter	Estimate	SE	p-value	Low.	Upp.	$0 \in$ CI
σ_1^2	0.00590	0.00071	.00000	0.00474	0.00707	F
σ_2^2	0.00038	0.00015	.00957	0.00014	0.00061	F
σ_{311}^2	0.00062	0.00030	.04286	0.00012	0.00112	F
σ_{312}^2	0.00065	0.00033	.04576	0.00011	0.00119	F
σ_{313}^2	0.00001	0.00030	.97860	−0.00048	0.00050	T
σ_{321}^2	0.00065	0.00030	.03099	0.00016	0.00115	F
σ_{322}^2	0.00030	0.00034	.38034	−0.00026	0.00087	T
σ_{323}^2	0.00061	0.00031	.04694	0.00010	0.00111	F

significant) strongly justifies adopting this model specification. This confirms that a standard homoscedastic FH3 structure would fail to capture the underlying variability of the data.

Figure 3 (left) plots the direct estimates and the EBLUPs based on the FH3var23rt model (FH3var23rt-EBLUPs) of the poverty proportions by crosses of provinces, age groups, sexes, and years. Figure 3 (right) plots the direct estimates and the FH3var23rt-EBLUPs of poverty proportions for all sub-subdomains of 2022, sorted by sample size.

In figure 3 (left), the dashed green line is the line $y = x$, thereby indicating the similarity between the direct estimates and model-based predictions. Most of the cloud points are situated around the line $y = x$, so the behavior of the FH3var23rt-EBLUPs is very similar to the direct estimates, except for those points that lie below this line and far to the right. These last model-based predictions show a more noticeable smoothing towards the synthetic estimator $\widehat{\mu}_{drt}^{synth} = x_{drt}\widehat{\boldsymbol{\beta}}$, that is, model mean, due to the small sample size for those sub-subdomains. A remainder of these observations present unrealistic values, between $40 - 50\%$, in the direct estimates of the poverty proportions.

5.1.1 Diagnosis measures.

This section performs the diagnosis of the selected FH3var23rt model by applying the measures introduced in section 3.2. Figure 4 (left) plots the raw residuals $r_{drt} = y_{drt} - \widehat{\mu}_{drt}$ of the fitted model. Figure 4 (right) shows the raw residuals sorted, in ascending order, by the variances of the direct estimator. It shows that the greater the variance of the direct estimators is, the higher the model residuals are.

Figure 5 (left) presents the studentized marginal residuals $\tilde{r}_{ch,drt}^{m}$, defined in (16), using boxplots. These are grouped by age group, sex, and year and are all within a suitably short range close to 0 (red line). However, they present a slightly right-skewed distribution.

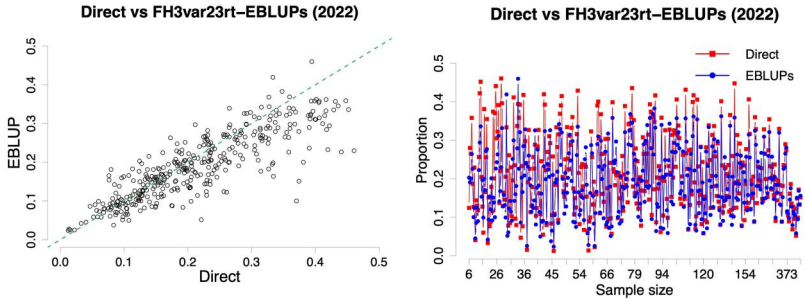

Figure 3. Direct Estimates Versus FH3var23rt-EBLUPs of Poverty Proportions by Crosses of Provinces, Age Groups, Sexes, and Years (Left), and by Crosses of Provinces, Age Groups, Sexes, and Year 2022, Sorted By Sample Size (Right).

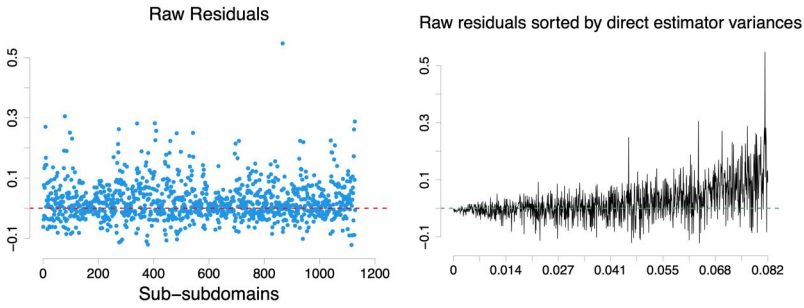

Figure 4. Raw Residuals of the FH3var23rt Model by Sub-Subdomain (Left) and Sorted by the Direct Estimator Variances (Right).

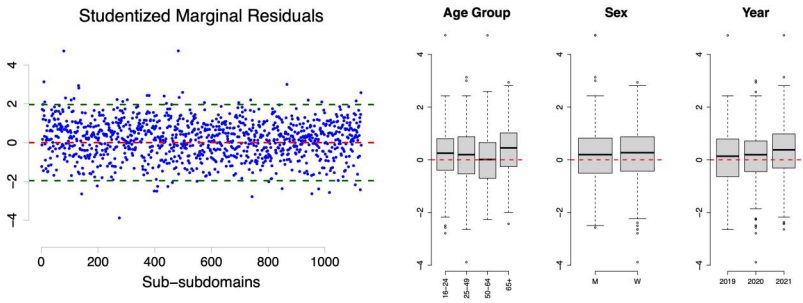

Figure 5. Boxplots of the studentized Marginal Residuals of the FH3var23rt Model Grouped by Age Groups, Sex and Year (Left) and Scatter Plot Across all the Domains (Right).

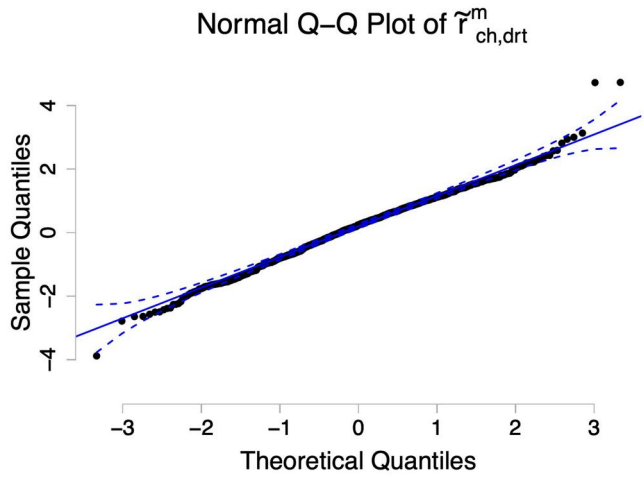

Figure 6. Normal *Q*–*Q* Plot of the Studentized Marginal Residuals of the FH3var23rt Model.

Figure 5 (right) shows the scatter plot of the studentized marginal residuals with the corresponding confidence limits at ± 2 (black lines). The band captures most of the studentized residuals.

Figure 6 shows the normal Q–Q plot of the studentized residuals with a good fitting; that is, it presents a linear trend according to the quantiles of the $N(0,1)$ distribution. Furthermore, with a confidence level of 95% most of the points belong to the confidence region, with only a small number of points falling outside. The graphical analysis of figures 5 and 6 concludes that the normality assumption may be acceptable.

Figure 7 (left) plots the Cook's distances, defined in (14), of the FH3var23rt-EBLUPs and highlights that the most influential provinces due to extreme residual points are Ávila (VI), Alicante (A), Badajoz (BA), Cáceres (CC), Burgos (BU), Ciudad Real (CR), Castellón (CS), Guadalajara (GU), Huelva (H), Jaén (J), Ourense (OU), and Toledo (TO), with this latter presenting the highest value.

Figure 7 (left) displays the two thresholds defined in (13), the traditional $4/D$ (red line) and the adjusted one $4/(D-p)$ (black line). The abbreviation PR_a denotes the cross of province acronym PR with age group a, $a = 1, \ldots, 4$. Due to high fixed-effects leverage value, people between [45,54] years old in Toledo (TO_2) are the most influential. Figure 7 (right) plots the fixed-effects versus the random-effects average leverages points, $L_{f,d}$ and $L_{r,d}$ defined in (12), of the domains (provinces $\times$ age group), with the corresponding thresholds. After averaging by subsubdomains and subdomains, Barcelona $\times [65, \infty]$, Madrid $\times [16, 44]$, and

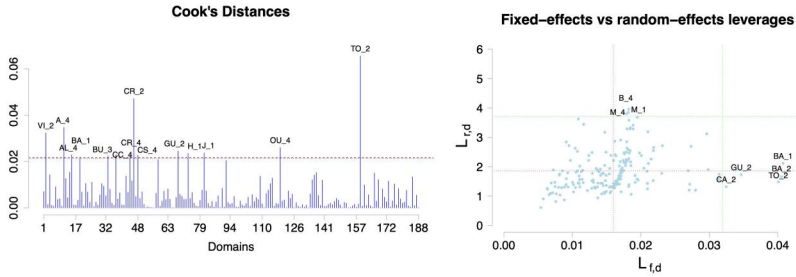

Figure 7. Cook's Distances (left) and Fixed-Effects Versus Random-Effects Average Leverages Points of the Major Domain (Right) of the FH3var23rt Model, With Province Abbreviations.

Madrid $\times [65, \infty]$ (B_4, M_1, M_4) are on the border of the second-level threshold as leverage points for the random-effects. Toledo $\times [45, 54]$, Badajoz $\times [16, 44]$, and Badajoz $\times [45, 54)$ (TO_2, BA_1 and BA_2) are the most influential on fixed-effects, exceeding the second cut-off value. This analysis confirms that people aged 45 and 54 in Toledo are the most influential points in the fitting of the FH3var23rt model.

Figure 8 plots the studentized marginal residuals, the fixed-effects leverage, and their cut-off thresholds of the FH3var23rt model. The labels PR_$a.k.t$ of the figure indicate: the province abbreviation, the age group ($a = 1, 2, 3, 4$), the sex ($k = M, W$), and the year ($t = 1, 2, 3$), respectively. We observe clear differences between men and women and within men. Specifically, women's fixed-effects leverages tend to be more concentrated, and they do around 0, than men's. In turn, men's fixed-effects leverages at the province and sex crosses are more dispersed for the year 2019 compared to the years 2020 and 2021. Differences can also be seen in the estimation of the EBLUPs under the FH3var23rt model, generally resulting in underestimates for women (J_1.W.3, AL_3.W.1, VI_2.W.3, and SG_2.W.2) and overestimates for men (CR_2.M.2). Table E.2 of the supplementary material online provides the abbreviations of the Spanish provinces, in alphabetical order.

Table 9 reports several selection methods, ranging from the AIC and the KIC criterion to the efficiency measure ε and the REML log-likelihood. To emphasise that these measures have been calculated using the REML fitting algorithm, they are denoted by this subscript (Log-like$_{REML}$, AIC$_{REML}$, KIC$_{REML}$, and ε_{REML}). The Log-like$_{REML}$, and the ε_{REML} criteria agree, reporting the best model as the FH3var23rt model, but the AIC and the KIC criteria indicate that the FH3var23rt model is the second best. The KIC selection criterion (Marhuenda et al. 2014) balances the goodness-of-fit of the model and the "level of complexity" of the model

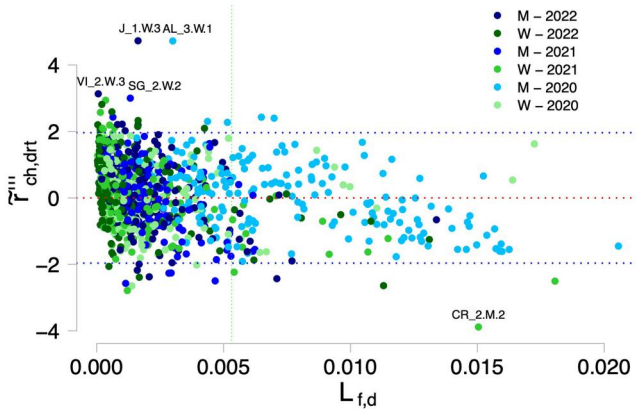

Figure 8. Fixed-Effects Leverages of the "Reduced" Model and Studentized Marginal Residuals, Together with Their Cut-Off Thresholds.

Table 9. FH3var Models Comparison Criteria

Model	Log-like$_{REML}$	AIC$_{REML}$	KIC$_{REML}$	ε_{REML}
FH3	2299.99	**−4587.98**	**−4581.98**	0.8191
FH3var2r3t	2301.07	−4584.15	−4575.15	0.8214
FH3var23rt	**2301.80**	−4581.61	−4570.61	**0.8218**
FH3var	2293.23	−4562.47	−4550.47	0.7927

NOTE.—The highest values of Log-like$_{REML}$ and ε, and the lowest values of AIC$_{REML}$ and KIC$_{REML}$, are shown in bold.

itself, where the complexity index is given by the Hilbert-Schmidt norm of the model covariance matrix, divided by the square of the model error variance (Danafar et al. 2015). For the FH3 model, due to the coincidence of high representativeness given by the REML log-likelihood and reduced complexity by a smaller number of parameters, the AIC and the KIC outcomes are explained. It is important to note that estimators AIC and KIC are theoretically derived and optimized under the maximum likelihood (ML) framework. Since our estimation is performed using REML likelihood to reduce bias in variance components, the AIC and KIC estimates should be interpreted with caution and are provided primarily for reference purposes. The efficiency index ε is based on the mean of the correlations between the model errors and the conditional residuals of the area-level model. The higher the correlation between the two, the better the approximation of the model error given by the residual. In particular, when the residuals approach the model errors (that is, whose information can be revealed by a high value of ε), the unknown area parameters under

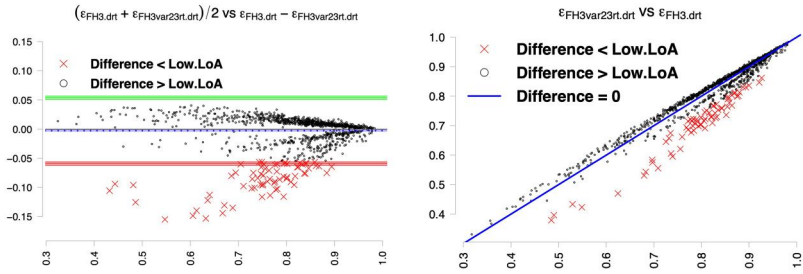

Figure 9. Bland–Altman Plot and Comparison Plot Between ε_{FH3} and $\varepsilon_{FH3var23rt}$.

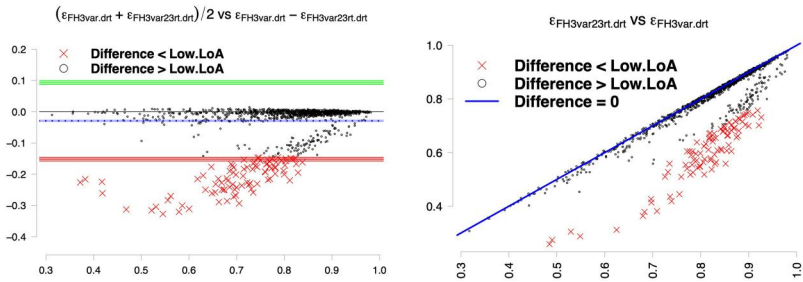

Figure 10. Bland–Altman Plot and Comparison Plot Between ε_{FH3var} and $\varepsilon_{FH3var23rt}$.

investigation are well approximated by the small area predictors. This situation is best represented by the highest values of the covariance between errors and residuals as the component of the numerator of the efficiency index ε.

Figures 9 and 10 show the Bland–Altman plot (left) based on the joint evaluation of the models FH3 and FH3var23rt, and the models FH3var and FH3var23rt, respectively. Together with the comparison chart (right), they provide a useful graphical evaluation to compare the models' performances. The Bland–Altman plot compares two measurement techniques and assesses the agreement between them (Bland and Altman 1999). The x-axis of the Bland–Altman plot displays the average of the efficiency measures ε of both compared models for each sub-subdomain, whereas the y-axis shows the difference. A systematic discrepancy between the models' performances would be reflected in a bias or trend, highlighting that the correlations between the residuals and model errors do not agree. Assuming the differences are approximately normally distributed, it is expected that 95% of them must lie between the limits of agreement

(LoA) $\overline{d} \pm 1.96 s_d$ (green and red bands), where $\overline{d}$ (blue dashed line) is the bias, estimated by the mean of differences, and s_d is the standard deviation of the differences, that is,

$$\begin{aligned}\overline{d} &= \frac{1}{DRT}\sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{t=1}^{T}(\varepsilon_{M1,drt} - \varepsilon_{M2,drt}), s_d \\ &= \left(\frac{1}{DRT-1}\sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{t=1}^{T}(\varepsilon_{M1,drt} - \varepsilon_{M2,drt})^2\right)^{1/2},\end{aligned}$$

where $\varepsilon_{M1,drt}, \varepsilon_{M2,drt}$ denote the efficiency measure of sub-subdomain *drt* under model $M1$ and $M2$, respectively. The bias in figures 9 and 10 indicates that the efficiency measures of the compared models are not equal. In particular, about 6.1% (69 out of 1128) and 8.1% (92 out of 1128) of the differences $\varepsilon_{FH3,drt} - \varepsilon_{FH3var23rt,drt}$ and $\varepsilon_{FH3var,drt} - \varepsilon_{FH3var23rt,drt}$, respectively, lie below (cross marks) the lower limit of agreement (Low.LoA) in both figures.

On the other hand, the comparison chart displays the efficiency measures $\varepsilon_{FH3var23rt,drt}$ on the x-axis against those of $\varepsilon_{FH3var,drt}$ and $\varepsilon_{FH3,drt}$ on the y-axis. Scatter points lie below the identity line ($y = x$, blue line), as the FH3var23rt model consistently provides higher measures of efficiency. This trend indicates a higher correlation of the conditional residuals with the sampling errors under the FH3var23rt model, thus in turn justifying its choice as the most efficient and appropriate model.

We conducted additional normality tests on the distribution of the differences in the efficiency measures to explore the differences between the models further. Table 10 reports the results of the Kolmogorov–Smirnov (K.Smirnov), Shapiro–Wilk (S.Wilk), and Anderson–Darling (A.Darling) tests, alongside the corresponding contrast statistic values (Stat.) and the *p*-values. All normality tests consistently reject the null hypothesis of normality. The rejection of the normality assumption indicates asymmetry and the presence of outliers in the distribution of differences between the FH3, FH3var, and FH3var23rt models. This reveals a systematic lack of agreement between the models' efficiency measures. This normality analysis and the results shown by figures 9 and 10 confirm that the FH3var23rt model is more efficient.

Figure 11 plots the efficiencies ε_{drt} of the FH3var23rt model versus the MSE ratios for the years 2020 (left) and 2021 (right), respectively. This figure shows to which provinces the model attributes more efficiency in the estimation. In our case study, we got $\boldsymbol{\epsilon} = 0.8218$ for the FH3var23rt model, the vertical blue bar in the figure. Those provinces and age groups that are less efficient in terms of ε_{drt} and low value of g_{3drt} are Alicante $\times [45, 54]$ (A_2), Alicante $\times [55, 64]$ (A_3), Almería $\times [65, \infty]$ (AL_4), Badajoz $\times [54, 65]$ (BA_3), Burgos $\times [16, 44]$ (BU_1), Cádiz $\times [16, 44]$

Table 10. Normality Tests on the Efficiency Measures Differences Distribution

	FH3var vs FH3var23rt		FH3 vs FH3var23rt	
Test	Stat.	*p*-value	Stat.	*p*-value
S.Wilk	0.5647	<.0001	0.8026	<.0001
K.Smirnov	0.4920	<.0001	0.4875	<.0001
A.Darling	204.280	<.0001	66.749	<.0001

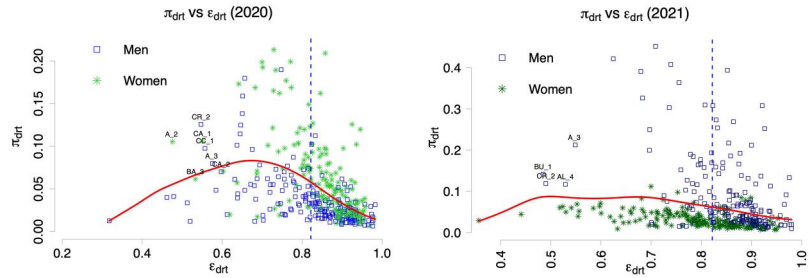

Figure 11. π_{drt} **vs** ε_{drt} **With Province** $\times$ **Age Group Abbreviation.**

(CA_1), Cádiz $\times [45, 54]$ (CA_2), Cáceres $\times [16, 44]$ (CC_1), and Ciudad Real $\times [45, 54]$ (CR_2).

Figure 11 shows that the most efficient predictions are those associated with sub-subdomains in the efficiency region, located to the right of the blue vertical line and below the red curved line, the local polynomial regression estimate (Wand and Jones 1995). These small area predictors are those that simultaneously present a high value of ε_{drt}—and therefore contribute to the reduction of the MSE—and a low value of π_{drt}—that means that the sum of $g_{1drt}(\widehat{\boldsymbol{\theta}})$ and $g_{2drt}(\widehat{\boldsymbol{\theta}})$ is a good approximation of MSE. Since there is a large accumulation of sub-subdomains in the efficiency region, table 11 shows the provinces that give rise to sub-subdomains belonging to the efficiency region. The column "#(Areas)" indicates the number of sub-subdomains of the efficiency region that are constructed as crosses of a given province with the categories of the variables age group, sex, and year. The "Provinces" column shows the labels of the provinces that generate efficient subdomains. For each province and year, there are up to 8 sub-subdomains, due to the 4 age groups and the 2 sexes. The most efficient provinces in terms of ε_{drt} and low value of g_{3drt} are Alicante (AL), Cádiz (CA), Cáceres (CC), Ceuta (CE), Córdoba (CO), Ciudad Real (CR), Jaén (J), Ourense (OU), Tarragona (T), and Santa Cruz de Tenerife (TF). In general,

Table 11. Most Efficient Provinces in Terms of ε_{drt} and Low Value of g_{3drt} in 2020 and 2021

#(Areas)	2020 Provinces	#(Areas)	2021 Provinces
8	AL, CC, CE, J, TF	8	J, OU, TF, VA
7	CR, ML, OU, TO	7	AB, AL, CE, CO, CR, H, T
6	AV, CA, CO, GC, GR, H, HU, T, TE, ZA	6	AV, CC, GI, L, VI
5	LU, MA	5	BU, GC, GR, IB, SE, TO, ZA
4	A, AB, CU, GI, L, PO, SG	4	CU, GU, HU, LE, LU, MA, P, SG

these correspond to groups with smaller sample sizes that benefit substantially from "borrowing strength" through the model.

5.2 Error Measures and Prediction

Model analysis and validation are the first steps in the model-based prediction approach. However, obtaining good estimates is as important as providing measures to quantify estimation error. This section assesses numerical and geographical information on model estimates and error estimation. Figure 12 plots the analytical approximation of the root-MSE of $\widehat{\overline{Y}}_{drt}^{eblup}$, that is, $(mse(\widehat{\overline{Y}}_{drt}^{eblup}))^{1/2}$, described in (8), against the benchmark $rmse(\widehat{\overline{Y}}_{drt}^{dir}) \approx \sigma_{drt}$. Figure 12 (left) and (right) compare these estimates for all the crosses of provinces, age groups, sexes, and years, sorted by direct estimator variances on the right side. The root-MSEs estimates for the FH3var23rt-EBLUPs are notably lower than those of the direct estimator, which is even more pronounced as the variance of the direct estimator increases.

Consistent with the evaluation in Simulation 3, computation times were evaluated using the same hardware specifications (MacBook Pro, Apple M3 8-core CPU, 16GB RAM). For the FH3var23rt model, the analytic MSE approximation required 3.42 seconds. On the other hand, the parametric bootstrap estimator averaged 1.68 seconds per iteration, resulting in a total computation time of 14 minutes for $B = 500$ replicates. This substantial difference further justifies the use of the analytical solution, particularly for applications involving large datasets.

Figure 13 maps the geographical distribution, at the provincial level, for the year 2021 of the FH3var23rt-EBLUPs of poverty proportions, by sex and age groups. These maps show an evident distributional discrepancy in

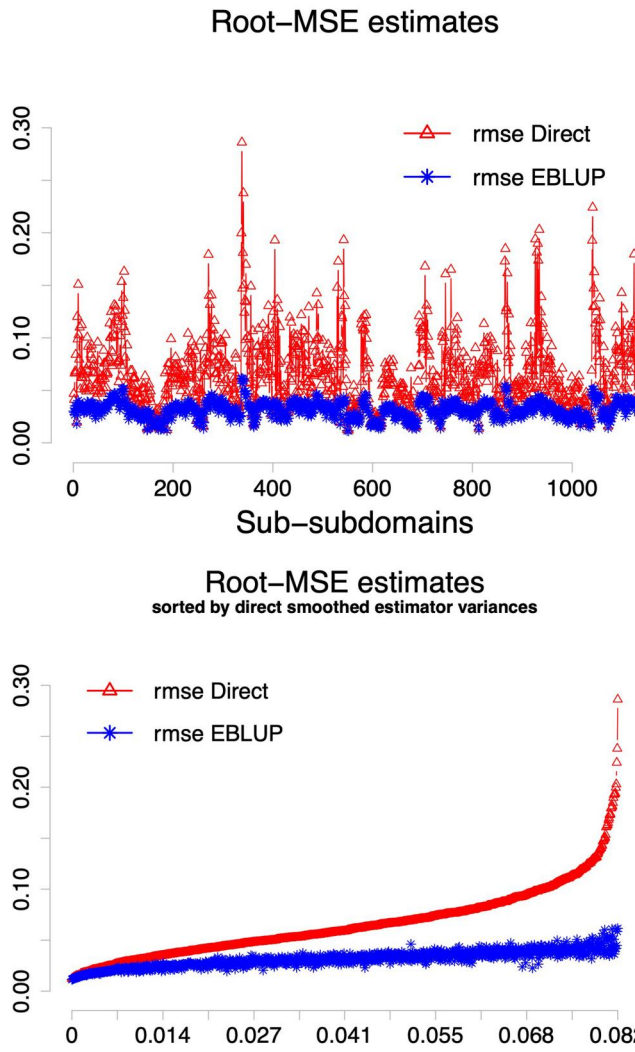

Figure 12. Root-MSE of the Direct Estimators Verus FH3var23rt-EBLUPs of the Poverty Proportions by Province Crossed With Age Groups, Sex, and Year (Top) Sorted by Direct Estimator Variances (Bottom).

the estimated poverty proportions between men and women, young and elderly people, and between the north and the south.

The maps highlight a notably higher incidence of poverty among the youngest and the oldest groups, with female subgroups being particularly vulnerable. Moreover, the spatial pattern of poverty shifts with age: while younger workers present a stronger southern concentration, the

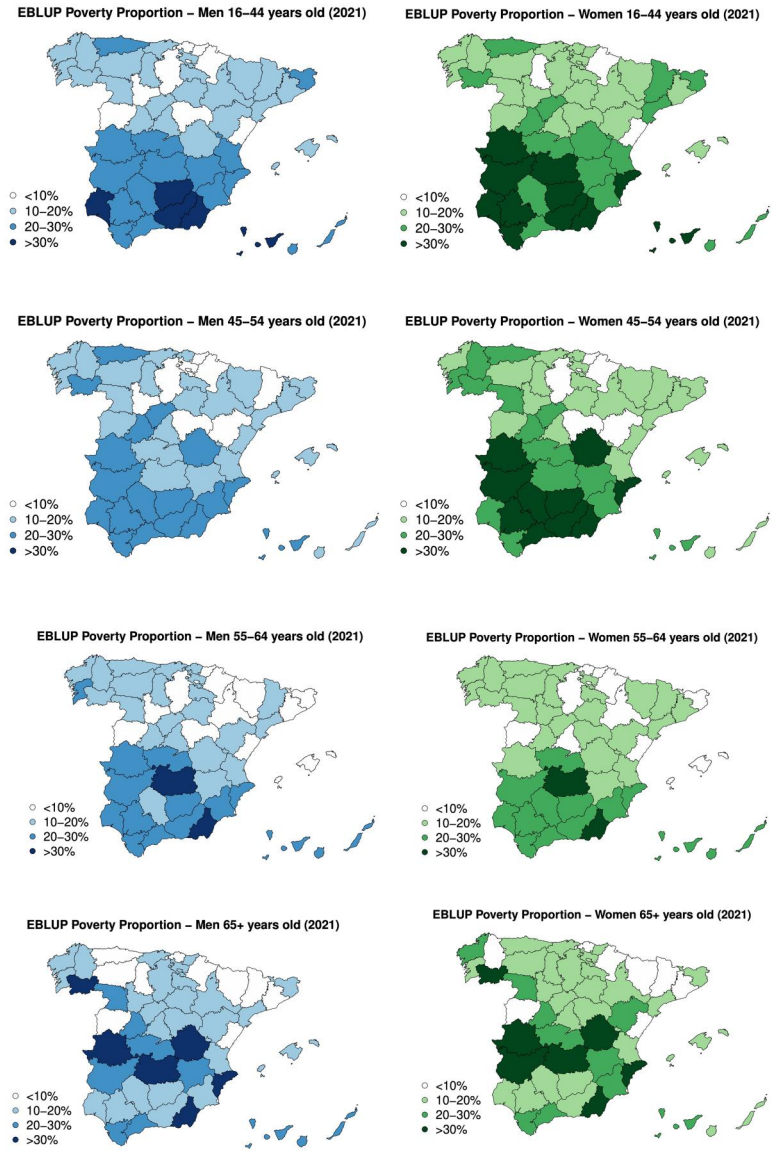

Figure 13. Maps of the FH3var23rt-EBLUPs of the Poverty Proportions in 2021, Cross-Classified by Sex (Left: Men, Right: Women) and Age Group (Top to Bottom: 16–44, 45–54, 55–64, and 65+).

distribution among the elderly becomes more heterogeneous, with a greater incidence of poverty in the central zone. Appendix E.2 (please see the supplementary materials online) contains the remaining maps for the

Spanish provinces crossed by age group in 2019 and 2020. No substantial changes in the spatial distribution of poverty are appreciated.

To provide a comprehensive national overview and ensure complete coverage, the out-of-sample domains (zero direct variance estimates) were predicted using the synthetic estimator $\widehat{\mu}_{drt}^{synth} = \boldsymbol{x}_{drt}\widehat{\boldsymbol{\beta}}$. To associate a measure of uncertainty with these predictions, we followed the approximation for the MSE of synthetic estimators described in Rao and Molina (2015, pp. 126–127). This approach accounts for both the uncertainty in the estimation of regression coefficients and the total variability of the unobserved random effects: $mse(\widehat{\mu}_{drt}^{synth}) \approx \boldsymbol{x}'_{drt}\widehat{\text{var}}(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_{drt} + \widehat{\sigma}_1^2 + \widehat{\sigma}_2^2 + \widehat{\sigma}_{3rt}^2$. The use of this MSE estimator enables the calculation of coefficients of variation (CVs) for these domains, thereby providing uncertainty measures for the entire set of EBLUPs.

Figure 14 maps the CVs of FH3var23rt-EBLUPs by sub-subdomains for the year 2021. Appendix E.2 (please see the supplementary materials online) provides the spatial distribution of the CVs for the years 2019 and 2020. The north zone of Spain, both east and west, presents the highest CV values due to the small sample size in these geographical domains. Specifically, people between 55–64 and +65 years old are the groups with lower sample sizes. Sex does not appear to be an influential factor in the distribution of CVs, indicating that both men and women are homogeneously represented in each age group.

Table 12 presents the quantiles of the estimated CVs (in %) of the direct estimators and the EBLUPs by sex and age group, that is,

$$\widehat{\text{CV}}(\widehat{\overline{Y}}_{drt}^{dir}) = 100\frac{\sqrt{\sigma_{drt}^2}}{\widehat{\overline{Y}}_{drt}^{dir}}, \widehat{\text{CV}}(\widehat{\overline{Y}}_{drt}^{eblup}) = 100\frac{\sqrt{mse(\widehat{\overline{Y}}_{drt}^{eblup})}}{\widehat{\overline{Y}}_{drt}^{eblup}}, d = 1, \ldots, D,$$

$r = 1, 2, a = 1, 2, 3, 4.$

Table 12 shows that CVs of EBLUPs are lower than CVs of direct estimators.

Table 13 displays the direct estimates, the FH3var23rt-EBLUPs and the coefficients of variation (CV) averaged by sex $(r = 1, 2)$ and age group $(a = 1, 2, 3, 4)$, that is,

$$\widehat{\overline{Y}}_{a,r.}^{dir} = \frac{1}{DT}\sum_{d\in\mathcal{D}_a}\sum_{t=1}^{T}\widehat{\overline{Y}}_{drt}^{dir}, \overline{\text{CV}}\left(\widehat{\overline{Y}}_{a,r.}^{dir}\right) = \frac{1}{DT}\sum_{d\in\mathcal{D}_a}\sum_{t=1}^{T}\widehat{\text{CV}}\left(\widehat{\overline{Y}}_{drt}^{dir}\right),$$

$$\widehat{\overline{Y}}_{a,r.}^{eblup} = \frac{1}{DT}\sum_{d\in\mathcal{D}_a}\sum_{t=1}^{T}\widehat{\overline{Y}}_{drt}^{eblup}, \overline{\text{CV}}\left(\widehat{\overline{Y}}_{a,r.}^{eblup}\right) = \frac{1}{DT}\sum_{d\in\mathcal{D}_a}\sum_{t=1}^{T}\widehat{\text{CV}}\left(\widehat{\overline{Y}}_{drt}^{eblup}\right).$$

The averages of the estimated poverty proportions, at the national level by age groups, are 21.2%, 19.9%, 17.4%, and 17.6%. In other words, the most

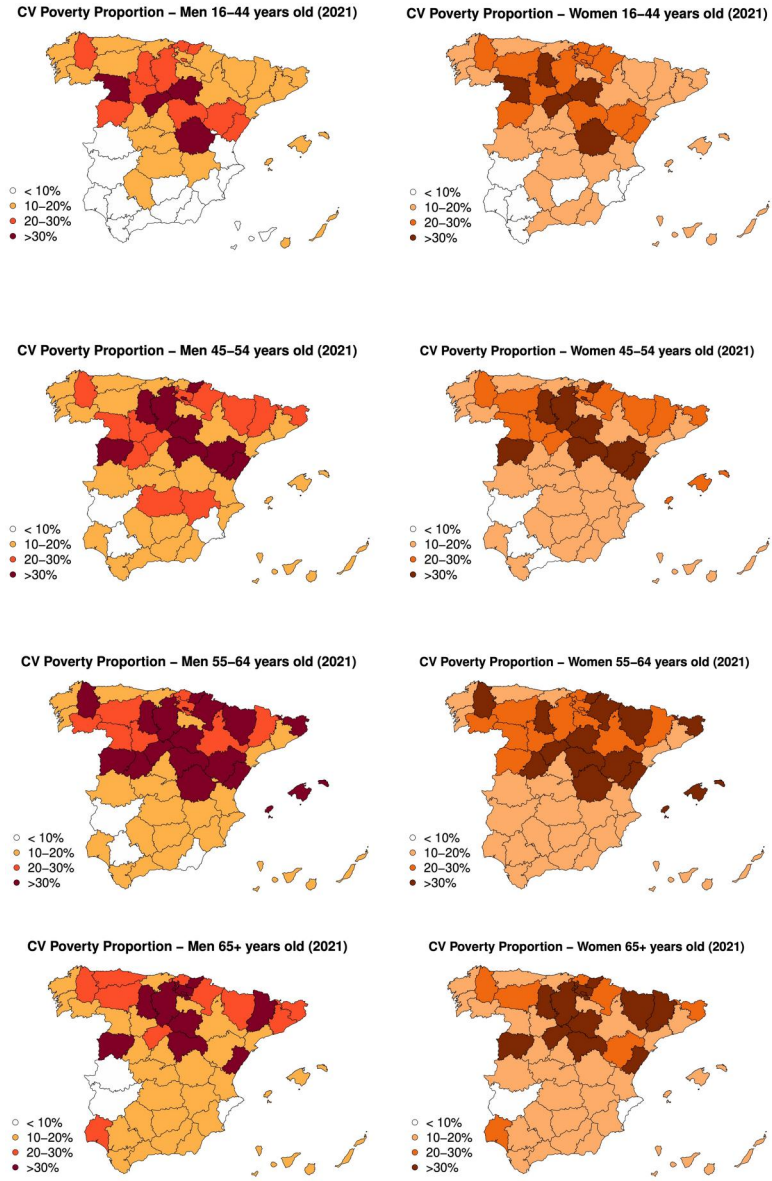

Figure 14. CVs' Maps of the FH3var23rt-EBLUPs of the Poverty Proportions in 2021, Cross-Classified by Sex (Left: Men, Right: Women) and Age Group (Top to Bottom: 16–44, 45–54, 55–64, and 65+).

Table 12. Quantiles of the CVs Estimates (in %) of the Direct Estimators and the EBLUPs by Sex and Age Group

	Age group	1		2		3		4	
	Sex	1	2	1	2	1	2	1	2
$\widehat{\mathrm{CV}}(\widehat{\overline{Y}}_{drt}^{dir})$	q_0	9.51	10.33	13.69	12.17	11.27	16.91	14.18	10.55
	$q_{0.25}$	19.62	18.51	25.99	24.78	26.30	27.44	27.63	21.63
	$q_{0.5}$	27.36	24.16	35.78	31.95	37.30	37.49	35.45	29.52
	$q_{0.75}$	42.27	36.27	50.58	48.83	50.27	53.02	51.77	40.77
	q_1	103.25	102.51	101.58	99.14	100.91	102.15	103.82	100.83
$\widehat{\mathrm{CV}}(\widehat{\overline{Y}}_{drt}^{eblup})$	q_0	7.44	7.83	9.02	8.49	8.91	10.72	8.62	8.81
	$q_{0.25}$	11.28	10.50	14.60	13.18	15.06	15.55	15.19	14.58
	$q_{0.5}$	15.64	14.86	19.95	17.62	20.34	20.31	19.74	17.84
	$q_{0.75}$	25.67	21.88	26.83	25.23	35.35	32.83	29.50	27.90
	q_1	53.78	47.03	71.59	82.44	74.82	55.37	88.93	63.12

Table 13. Averaged Direct Estimates, EBLUPs and CVs (in %) Estimates by Sex and Age Group

Sex	Age group	$\widehat{\overline{Y}}_{a,r.}^{dir}$	$\widehat{\overline{Y}}_{a,r.}^{eblup}$	$\overline{\mathrm{CV}}(\widehat{\overline{Y}}_{a,r.}^{dir})$	$\overline{\mathrm{CV}}(\widehat{\overline{Y}}_{a,r.}^{eblup})$
1	1	0.2081	0.2008	33.3769	19.4003
2	1	0.2296	0.2230	29.7522	17.8029
1	2	0.2281	0.1872	41.6890	23.9176
2	2	0.2302	0.2105	38.5930	21.3317
1	3	0.1947	0.1727	41.0232	25.6138
2	3	0.1760	0.1756	43.1747	24.9453
1	4	0.1728	0.1709	42.3042	25.0406
2	4	0.2010	0.1822	34.5354	22.9202

affected age groups are the ones between 16 and 54 years old. In addition, the women belonging to the younger age group (16–44) have the highest proportions of poverty among all. Differences in the averaged coefficients of variation are striking between the direct estimators and the FH3var23rt-EBLUPs, with $\overline{\mathrm{CV}}$ values increasing with age group. Specifically, the proposed model achieves reductions in the average coefficient of variation ranging from 11.5% to 18% across age groups and sex compared to the direct estimates.

6. CONCLUSIONS

This article presents and tests the threefold Fay–Herriot with unequal variance random effects model (FH3var) and its submodels for small area

estimation. A key contribution of this work lies in the derivation and implementation of this full model family, extending beyond the standard homoscedastic FH3 specification. We offer applied statisticians and practitioners a flexible modelling that captures hierarchical dependencies across three levels of aggregation to improve the precision and reliability of small area estimates.

Simulation results empirically verify the expected theoretical properties of unbiasedness and consistency of model parameter estimators. In terms of predictive performance, the new models demonstrate robust behavior across different scenarios. Under homoscedastic conditions, the loss of precision of the MSE compared to the standard FH3 model is negligible, indicating that the computational cost of estimating additional variance parameters does not compromise accuracy. Conversely, in heteroscedastic scenarios, the proposed models outperform the standard approach, although the gains in efficiency are modest. This finding emphasizes the importance of balancing predictive accuracy with model interpretability. The FH3var family enables practitioners to rigorously capture underlying heteroscedasticity without sacrificing the explanatory power of the models. Regarding the performance of the MSE estimators, the analytical approach is found to be highly accurate and stable. However, the bootstrap-based estimator provides accurate results but requires at least $B = 400$ iterations.

The diagnosis of the four variants of the FH3var model has allowed us to choose the optimal model. A decisive factor in this selection was the presence of statistically significant variance components in the more complex configurations, which confirmed the existence of heteroscedasticity that the standard model fails to capture. This structural evidence complements the utility of the index ε, which measures the efficiency of the best linear predictor. Although selection criteria may require a careful assessment of the hierarchical structure, the index ε is based solely on the correlation between model errors and residuals. Moreover, the Bland–Altman analysis of the differences in this index has made it possible to identify the optimal model as statistically significant, suggesting a substantial difference in the estimation of area parameters across the models.

The application to SLCS 2019–2021 data reveals the practical advantage of the FH3var-type models in poverty estimation. Regarding the auxiliary information used in this application, it is important to acknowledge that the use of survey estimates as explanatory variables introduces measurement error. While the standard recommendation is to utilize administrative or census records, such data are not always available. We addressed this limitation by averaging the four quarterly SLFS for each year, a strategy that reduces the sampling variability of the auxiliary variables and minimizes the potential impact on fixed effects estimation. Consequently, the model successfully smooths direct estimators while preserving meaningful spatial and temporal distributions. The threefold hierarchical setup

allows for more reliable estimates by provinces and demographic subgroups, especially in domains with small sample sizes. The comparative analysis of the CV estimates strongly suggests this improvement. Specifically, across all age groups and sexes, the proposed model achieves an average reduction in CVs of 10–20 percentage points compared to the direct estimates. This substantial gain in precision demonstrates that the FH3var framework successfully reduces estimation uncertainty to reliable levels for official statistics. Overall, the results show that an appropriate variance structure improves both accuracy and interpretability.

SUPPLEMENTARY MATERIALS

Supplementary materials are available online at academic.oup.com/jssam.

DATA AVAILABILITY

The data analyzed in this study were provided by the Spanish Statistical Office (Instituto Nacional de Estadística, INE). Due to confidentiality agreements and privacy restrictions associated with the survey unit-level data, the datasets are not publicly available. Data access may be requested directly from INE, subject to their specific data protection protocols and licensing agreements.

REFERENCES

Benavent, R., and Morales, D. (2021), "Small Area Estimation Under a Temporal Bivariate Area-Level Linear Mixed Model with Independent Time Effects," *Statistical Methods and Applications*, 30, 195–222.

Betti, G., and Lemmi, A. (2013), *Poverty and Social Exclusion: New Methods of Analysis* (1st ed.), London, UK: Routledge.

Bland, J. M., and Altman, D. G. (1999), "Measuring Agreement in Method Comparison Studies," *Statistical Methods in Medical Research*, 8, 135–160.

Boonstra, H. J., and Van Den Brakel, J. (2022), "Multilevel Time-Series Models for Small Area Estimation at Different Frequencies and Domain Levels," *The Annals of Applied Statistics*, 16, 2314–2338.

Burgard, J. P., Esteban, M. D., Morales, D., and Pérez, A. (2020), "A Fay-Herriot Model When Auxiliary Variables Are Measured With Error," *TEST*, 29, 166–195.

Burgard, J. P., Krause, J., and Morales, D. (2022), "A Measurement Error Rao-Yu Model for Regional Prevalence Estimation over Time Using Uncertain Data Obtained from Dependent Survey Estimates," *TEST*, 31, 204–234.

Cabello, E., Esteban, M. D., Morales, D., and Pérez, A. (2025), "Small Area Estimation of Poverty Proportions Under Heteroscedastic Fay-Herriot Models," *Journal of Applied Statistics*, 1–23, Online publication. https://doi.org/10.1080/02664763.2025.2568679.

Cai, S., and Rao, J. N. K. (2022), "Selection of Auxiliary Variables for Three-Fold Linking Models in Small Area Estimation: A Simple and Effective Method," *Stats*, 5, 128–138.

Cai, S., Rao, J. N. K., Dumitrescu, L., and Chatrchi, G. (2020), "Effective Transformation-Based Variable Selection Under Two-Fold Subarea Models in Small Area Estimation," *Statistics in Transition New Series*, 21, 68–83.

Calvin, J. A., and Sedransk, J. (1991), "Bayesian and Frequentist Predictive Inference for the Patterns of Care Studies," *Journal of the American Statistical Association*, 86, 36–48.

Danafar, S., Fukumizu, K., and Gomez, F. (2015), "Kernel-Based Information Criterion," *Computer and Information Science*, 8, 10–10.

Datta, G. S., Lahiri, P., Maiti, T., and Lu, K. L. (1999), "Hierarchical Bayes Estimation of Unemployment Rates for the U.S. states," *Journal of the American Statistical Association*, 94, 1074–1082.

Datta, G. S., Lahiri, P., and Maiti, T. (2002), "Empirical Bayes Estimation of Median Income of Four-Person Families by State Using Time Series and Cross-Sectional Data," *Journal of Statistical Planning and Inference*, 102, 83–97.

Demidenko, E., and Stukel, T. A. (2005), "Influence Analysis for Linear Mixed-Effects Models," *Statistics in Medicine*, 24, 893–909.

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, 2840–2855.

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, 269–277.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2010), "Small Area Estimation Under Fay-Herriot Models with Nonparametric Estimation of Heteroskedasticity," *Statistical Modelling*, 10, 2, 215–239.

Ghosh, M., Nangia, N., and Kim, D. (1996), "Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach," *Journal of the American Statistical Association*, 91, 1423–1431.

Harrell, F. E. Jr (2015), *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, Cham, Switzerland: Springer.

Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J. M., and Thiébaut, R. (2007), "Robustness of the Linear Mixed Model to Misspecified Error Distribution," *Computational Statistics & Data Analysis*, 51, 5142–5154.

Krenzke, T., Mohadjer, L., Li, J., Erciulescu, A., Fay, R. E., Ren, W., VanDeKerckhove, W., Li, L., and Rao, J. N. K. (2020), *Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report; Technical Report*. Washington, D.C.: Institute of Education Sciences, National Center for Education Statistics.

Marcis, L., Morales, D., Pagliarella, M. C., and Salvatore, R. (2023), "Three-Fold Fay-Herriot Model for Small Area Estimation and Its Diagnostics," *Statistical Methods and Applications*, 32, 1563–1609.

Marhuenda, Y., Molina, I., and Morales, D. (2013), "Small Area Estimation with Spatio-Temporal Fay-Herriot Models," *Computational Statistics & Data Analysis*, 58, 308–325.

Marhuenda, Y., Morales, D., and del Carmen Pardo, M. (2014), "Information Criteria for Fay-Herriot Model Selection," *Computational Statistics & Data Analysis*, 70, 268–280.

Morales, D., Pagliarella, M. C., and Salvatore, R. (2015), "Small Area Estimation of Poverty Indicators Under Partitioned Area-Level Time Models," *SORT-Statistics and Operations Research Transactions*, 39, 19–34.

Morales, D., Esteban, M. D., Pérez, A., and Hobza, T. (2021), *A Course on Small Area Estimation and Mixed Models*, Cham, Switzerland: Springer.

Nobre, J. S., and Singer, J. M. (2007), "Residual Analysis for Linear Mixed Models," *Biometrical Journal*, 49, 863–875.

Pfeffermann, D., and Burck, L. (1990), "Robust Small Area Estimation Combining Time Series and Cross-Sectional Data," *Survey Methodology*, 16, 217–237.

Pratesi, M. (Ed.). (2016), *Analysis of Poverty Data by Small Area Estimation*, Chichester, UK: Wiley.

Rao, J. N. K., and Molina, I. (2015), *Small Area Estimation* (2nd ed.), Hoboken: Wiley
Rao, J. N. K., and Yu, M. (1994), "Small Area Estimation by Combining Time Series and Cross Sectional Data," *Canadian Journal of Statistics*, 22, 511–528.
Pratesi, M., and Salvati, N. (2008), "Small Area Estimation: The EBLUP Estimator Based on Spatially Correlated Random Area Effects," *Statistical Methods and Applications*, 17, 113–141.
Sinha, S. K., and Rao, J. N. K. (2009), "Robust Small Area Estimation," *Canadian Journal of Statistics*, 37, 381–399.
Singh, B., Shukla, G., and Kundu, D. (2005), "Spatio-Temporal Models in Small Area Estimation," *Survey Methodology*, 31, 183–195.
Torabi, M., and Rao, J. N. K. (2014), "On Small Area Estimation Under a Sub-Area Level Model," *Journal of Multivariate Analysis*, 127, 36–55.
Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman and Hall.
Ybarra, L. M. R., and Lohr, S. L. (2008), "Small Area Estimation When Auxiliary Information is Measured With Error," *Biometrika*, 95, 919–931.
Wooldridge, J. M. (2016), *Introductory Econometrics: A Modern Approach*, Boston, MA: South-Western Cengage Learning.
You, Y. (2021), "Small Area Estimation Using Fay-Herriot Area Level Model with Sampling Variance Smoothing and Modeling," *Survey Methodology*, 47, 361–371.
You, Y., and Rao, J. N. K. (2000), "Hierarchical Bayes Estimation of Small Area Means Using Multi-Level Models," *Survey Methodology*, 26, 173–181.
Zewotir, T., and Galpin, J. (2007), "A Unified Approach on Residuals, Leverages and Outliers in the Linear Mixed Model," *TEST*, 16, 58–75.

Journal of Survey Statistics and Methodology, 2026, 00, 1–43
https://doi.org/10.1093/jssam/smag013
Article