TEST (2026) 35:285–316
https://doi.org/10.1007/s11749-025-00998-2

ORIGINAL PAPER

Variable selection for Fay–Herriot models: a cooperative game theory approach

E. Cabello[1] · **J. C. Gonçalves-Dosantos**[1] · **D. Morales**[1] · **J. Sánchez-Soriano**[1]

Received: 25 June 2025 / Accepted: 21 November 2025 / Published online: 12 December 2025

Abstract
This paper presents a novel approach to variable selection in small area estimation, focusing on the Fay–Herriot model. Traditional methods, such as those based on Akaike and Kullback symmetric divergence information criteria, often rely on stepwise selection and focus on the complete model, without individually examining the influence of auxiliary variables. The Shapley value of cooperative game theory is proposed to measure the average importance of auxiliary variables. The Shapley value evaluates all possible combinations of auxiliary variables, ensuring an efficient average influence of the selected combination. We study its properties mathematically and investigate its performance through simulation experiments, showing consistent identification of the true generating variables even under challenging conditions. An application to the 2022 Spanish Living Conditions Survey illustrates the method's usefulness in selecting the model on which to base predictors of poverty proportions in Spanish provinces by sex.

Keywords Fay-Herriot models · Information criteria · Shapley value · Influence measure · Small area estimation · Living condition surveys

Mathematics Subject Classification 62E30 · 62J12 · 91A80

1 Introduction

Small area estimation (SAE) is crucial for producing reliable statistics for domains with limited sample sizes, such as geographic regions or demographic subgroups. Model-

Supported by the CIPROM/2024/34 grant, funded by the Conselleria de Educación, Cultura, Universidades y Empleo, Generalitat Valenciana, by the Spanish grants PID2022-136878NB-I00, PID2021-12403030NB-C31 and PID2022-137211NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe/EU".

✉ E. Cabello
ecabello@umh.es

[1] Centro de Investigación Operativa,Universidad Miguel Hernández de Elche,Elche, Spain

based approaches, particularly the area-level Fay-Herriot (FH) model, are widely employed to improve precision by incorporating complex variance structures and auxiliary information. The monograph Rao and Molina (2015) provides the theoretical foundation and a comprehensive treatment of the methods and applications of SAE. A central aspect of constructing effective SAE models lies in model selection, as the choice of auxiliary variables significantly affects estimation accuracy, interpretability and robustness. Choosing the correct explanatory variables helps balance between capturing the underlying structure of the data and avoiding over-fitting. While numerous studies have addressed model selection in SAE, the specific problem of assigning importance to individual variables has not yet been explored.

Traditional model selection techniques in SAE typically rely on information criteria: Akaike's Information Criterion (AIC), Kullback's Information Criterion (KIC) and Bayesian Information Criterion (BIC), among others, are the most popular. Notable contributions include the use of bias-corrected AIC by Hurvich and Tsai (1989), bootstrap-corrected variants by Shang and Cavanaugh (2008) and Marhuenda et al. (2014), and extensions to spatio-temporal and semi-parametric models by Kubokawa (2011) and Lombardía et al. (2018; 2021). Additionally, the work by Lahiri and Suntornchost (2015) studies penalized likelihood methods for model selection in linear mixed models (LMM), including the FH model. These approaches aim to identify optimal subsets of variables by minimizing loss functions over a set of candidate models. However, rather than quantifying the marginal influence of auxiliary variables, these criteria provide a pointwise evaluation of the overall model fit.

This gap in the literature highlights the need for new methodologies that evaluate the marginal contributions of each variable, which could lead to robust and effective SAE models. We address this problem by introducing and analysing empirically an influence measure based on the Shapley value for cooperative games (Shapley 1953), which quantifies the average contribution of each auxiliary variable by considering all possible subsets of them. The connection between cooperative game theory, and in particular the Shapley value, and some areas of statistical science is well established. The Shapley value is widely applied in machine learning (Rozemberczki et al. 2022), feature selection, explainability, multi-agent reinforcement learning, set pruning and data valuation. Furthermore, the Shapley value has previously been applied in linear fixed-effects models to study the relative importance of auxiliary variables in the presence of multicollinearity and their contributions to the R-squared coefficient, as analysed by Lipovetsky and Conklin (2001); Israeli (2007), and Mishra (2016), among others. However, the connection between game theory and SAE has not yet been studied.

The strength of SAE models lies not only in their ability to generate accurate predictions of small area parameters, but also in their ability to identify socioeconomic patterns relevant to policymaking. These models allow the behaviour of a target variable to be explained by auxiliary variables, and the marginal contribution of each explanatory variable can shed additional light on the interrelationships in the data. We illustrate these ideas with an application to SAE estimation of poverty proportions, using data from the 2022 Spanish Living Conditions Survey (SLCS).

The estimation of poverty indicators in small areas has received considerable attention from researchers in recent years. Without wishing to be exhaustive, we cite some

relevant contributions with applications to the European surveys of income and living conditions. From the perspective of unit-level models, we can cite the empirical best predictors based on LMMs introduced by Molina and Rao (2010); Marhuenda et al. (2017) and Guadarrama et al. (2021). Robust estimation procedures have been addressed by Tzavidis et al. (2008; 2015), Marchetti et al. (2012) and Marchetti and Secondi (2016). In the case of area-level models, Esteban et al. (2012); Marhuenda et al. (2013), Morales et al. (2015) introduced predictors based on LMMs. However, Boubeta et al. (2016; 2017), López-Vizcaíno et al. (2013; 2015) or Chandra et al. (2017) employed generalized LMMs , with Poisson or multinomial distributions. Specifically, the volumes by Pratesi (2016) and Betti and Lemmi (2013) focus on small-area estimation of poverty indicators.

In the context of SAE and LMMs, the use of the Shapley value remains unexplored. The novelty of our approach lies in four key aspects: (1) to propose an influence measure based on the Shapley value tailored to model selection criteria in the FH model, (2) to introduce a model selection algorithm, (3) to characterize axiomatically the influence measure and (4) to validate the approach through both simulation experiments and a real-world application using the SLCS 2022 data to estimate sex-disaggregated poverty proportions in Spanish provinces.

The paper is organized as follows. Section 2 describes the Fay–Herriot model and the AIC and KIC families of information criteria. Section 3 describes a measure of variable influence based on the Shapley value. Additionally, an axiomatic characterization is included. Section 4 presents simulation experiments to investigate the behaviour of the Shapley value to detect the model generating auxiliary variables and the proposed Shapley's algorithm for model selection. Section 5 applies the proposed methodology to data from the SLCS of the last quarter of 2022 in Spanish provinces. Section 6 summarizes the main conclusions of this work.

The paper includes five appendices. Appendix A describes the procedure to simulate the synthetic data used in Examples 1, 2 and 3. Appendix B provides the proof of the theorem establishing the desired properties of the influence measure based on the Shapley value. Appendix C presents two propositions regarding the properties of the Shapley value estimator. Appendix D describes the parametric bootstrap approach to estimate the mean squared errors of the model-based predictors. Appendix E contains the results of Simulation 1 for the Shapley value estimator.

2 Fay–Herriot model and information criteria

2.1 Fay–Herriot model

The FH model is the basic area-level model in SAE, especially useful when individuallevel auxiliary information and information on the dependent variable are not accessible, but area-level aggregates of the auxiliary variables can be obtained from administrative sources. It represents a special case of an LMM featuring a domain-specific random intercept. Its principal strength lies in the capacity to integrate direct survey estimates with auxiliary information available at the domain level, thereby borrowing strength among areas. The model improves the precision of direct estimates by

smoothing the target variable towards a synthetic value, which results from a weighted combination of survey data and auxiliary variables.

Let us consider a finite population U partitioned into D domains $U_1, \ldots, U_D$, i.e. $U = \cup_{d=1}^{D} U_d$, $U_{d_1} \cap U_{d_2} = \emptyset$, $d_1, d_2 = 1, \ldots, D, d_1 \neq d_2$. Let y_d be the direct estimate of the unknown domain quantity μ_d, $d = 1, \ldots, D$. Let $P = \{1, \ldots, |P|\}$ be the set of subscripts of $|P|$ potential domain-level explanatory variables, where $|\cdot|$ denotes the cardinality of a set. Let x_{di}, $i \in P$, be the value that the explanatory variable x_i takes in the domain $U_d, d = 1, \ldots, D$. For a subset $S = \{i_1, \ldots, i_{|S|}\} \subseteq P$, of explanatory variables, the FH model M_S is defined in two stages. The sampling model is

$$y_d = \mu_d + e_d, \quad d = 1, \ldots, D, \tag{2.1}$$

where $e_1, \ldots, e_D$ are independent and normally distributed, with $e_d \sim N(0, \sigma_{ed}^2)$, where $\sigma_{ed}^2 > 0$ is known. In applied settings, when unit-level data is available, these variances can be estimated from the survey. For example, using the direct estimator of the variance (see Morales et al. (2021), chap. 2, p. 24) and the sampling weights under the sampling design. This link from design-based to model-based is standard in official statistics, allowing the FH model to exploit unit-level information indirectly while maintaining an area-level formulation.

The linking model assumes that

$$\mu_d = \boldsymbol{x}_{Sd}\boldsymbol{\beta}_S + u_{Sd}, \quad d = 1, \ldots, D, \tag{2.2}$$

where $\boldsymbol{x}_{Sd}$ is a row vector containing the aggregated (population) values of $|S|$ explanatory variables for area d, $\boldsymbol{\beta}_S$ is the column vector of regression coefficients and $u_{S1}, \ldots, u_{SD}$ are independent and identically distributed as $N(0, \sigma_{Su}^2)$, with variance $\sigma_{Su}^2 > 0$ unknown, and independent of the sampling errors e_d's. The unknown parameter vector of model M_S is $\boldsymbol{\theta}_S = (\boldsymbol{\beta}_S', \sigma_{Su}^2)'$. We assume that there exists an $S = T$, with $T \subseteq P$, for which the FH model M_T is 'true'; that is, the M_T model generates y_d, $d = 1, \ldots, D$. For $S \neq T$, the equation (2.2) does not hold and, therefore, the model M_S is not correct.

Combining (2.1) and (2.2), the FH model M_S can be expressed as a single model in the form

$$y_d = \boldsymbol{x}_{Sd}\boldsymbol{\beta}_S + u_{Sd} + e_d, \quad d = 1, \ldots, D, \tag{2.3}$$

or in matrix form as $\boldsymbol{y} = \boldsymbol{X}_S\boldsymbol{\beta}_S + \boldsymbol{u}_S + \boldsymbol{e}$, where

$$\boldsymbol{y} = \underset{1\le d\le D}{\text{col}}(y_d), \ \boldsymbol{X}_S = \underset{1\le d\le D}{\text{col}}(\boldsymbol{x}_{Sd}), \ \boldsymbol{u}_S = \underset{1\le d\le D}{\text{col}}(u_{Sd}), \ \boldsymbol{e} = \underset{1\le d\le D}{\text{col}}(e_d).$$

Then, $\boldsymbol{V}_{Su} = \text{var}(\boldsymbol{u}_S) = \sigma_{Su}^2\boldsymbol{I}_D, \boldsymbol{V}_e = \text{var}(\boldsymbol{e}) = \text{diag}(\sigma_{e1}^2, \ldots, \sigma_{eD}^2), \boldsymbol{V}_S = \text{var}(\boldsymbol{y}) = \boldsymbol{V}_{Su} + \boldsymbol{V}_e = \text{diag}(\sigma_{Su}^2 + \sigma_{e1}^2, \ldots, \sigma_{Su}^2 + \sigma_{eD}^2)$, and $\text{cov}(\boldsymbol{y}, \boldsymbol{u}_S) = \boldsymbol{V}_{Su}$.

If σ^2_{Su} is known, the BLUP of μ_d is $\tilde{\mu}_{Sd} = \boldsymbol{x}_{Sd}\tilde{\boldsymbol{\beta}}_S + \tilde{u}_{Sd}$, where

$$\tilde{\boldsymbol{\beta}}_S = (\boldsymbol{X}_S{}'\boldsymbol{V}_S^{-1}\boldsymbol{X}_S)^{-1}\boldsymbol{X}_S{}'\boldsymbol{V}_S^{-1}\boldsymbol{y} \quad \text{and} \quad \tilde{u}_{Sd} = \frac{\sigma^2_{Su}}{\sigma^2_{Su}+\sigma^2_{ed}}\big(y_d - \boldsymbol{x}_{Sd}\tilde{\boldsymbol{\beta}}_S\big), \quad d = 1, \ldots, D,$$

are the best linear unbiased estimator (BLUE) and predictor (BLUP) of $\boldsymbol{\beta}_S$ and u_{Sd}. In matrix form, the BLUP of the vector $\boldsymbol{u}_S$ is given by

$$\tilde{\boldsymbol{u}}_S = \boldsymbol{V}_{Su}\boldsymbol{V}_S^{-1}\big(\boldsymbol{y} - \boldsymbol{X}_S\tilde{\boldsymbol{\beta}}_S\big).$$

Thus, the BLUP of μ_d can be written as

$$\begin{aligned}\tilde{\mu}_{Sd} &= \boldsymbol{x}_{Sd}\tilde{\boldsymbol{\beta}}_S + \tilde{u}_{Sd} = \boldsymbol{x}_{Sd}\tilde{\boldsymbol{\beta}}_S + \frac{\sigma^2_{Su}}{\sigma^2_{Su}+\sigma^2_{ed}}\big(y_d - \boldsymbol{x}_{Sd}\tilde{\boldsymbol{\beta}}_S\big)\\ &= \frac{\sigma^2_{Su}}{\sigma^2_{Su}+\sigma^2_{ed}}\,y_d + \frac{\sigma^2_{ed}}{\sigma^2_{Su}+\sigma^2_{ed}}\,\boldsymbol{x}_{Sd}\tilde{\boldsymbol{\beta}}_S. \end{aligned} \tag{2.4}$$

The empirical BLUE (EBLUE) of $\boldsymbol{\beta}_S$ and the empirical BLUP (EBLUP) of $\boldsymbol{u}_S$ are obtained by substituting an estimator $\hat{\sigma}^2_{Su}$ of σ^2_{Su}, i.e.

$$\hat{\boldsymbol{\beta}}_S = (\boldsymbol{X}_S'\hat{\boldsymbol{V}}_S^{-1}\boldsymbol{X}_S)^{-1}\boldsymbol{X}_S'\hat{\boldsymbol{V}}_S^{-1}\boldsymbol{y}, \quad \hat{\boldsymbol{u}}_S = \hat{\boldsymbol{V}}_{Su}\hat{\boldsymbol{V}}_S^{-1}\big(\boldsymbol{y} - \boldsymbol{X}_S\hat{\boldsymbol{\beta}}_S\big), \tag{2.5}$$

where $\hat{\boldsymbol{V}}_{Su} = \hat{\sigma}^2_{Su}\boldsymbol{I}_D$ and $\hat{\boldsymbol{V}}_S = \hat{\boldsymbol{V}}_{Su} + \boldsymbol{V}_e = \underset{1\le d\le D}{\text{diag}}\,(\hat{\sigma}^2_{Su} + \sigma^2_{ed})$. Under the model (2.3), the EBLUP of μ_d is similarly obtained from the BLUP (2.4), i.e.

$$\hat{\mu}_{Sd} = \frac{\hat{\sigma}^2_{Su}}{\hat{\sigma}^2_{Su}+\sigma^2_{ed}}\,y_d + \frac{\sigma^2_{ed}}{\hat{\sigma}^2_{Su}+\sigma^2_{ed}}\,\boldsymbol{x}_{Sd}\hat{\boldsymbol{\beta}}_S. \tag{2.6}$$

The model parameters are estimated using maximum likelihood (ML). This choice is consistent with the use of AIC and KIC information criteria, which are defined in terms of the maximized log-likelihood and therefore require ML-based fitting. A description of the fitting algorithm and the asymptotic distributions of the estimators under ML can be found in Morales et al. (2021) (chap. 16, pp. 432-434). The p-values for the regression coefficients are obtained from a Wald test based on the ML fit, using the inverse Fisher information matrix. The FH model is fitted using the function `eblupFH` from the `sae` package in `R`, see Molina and Marhuenda (2015). Throughout the manuscript, and with a slight abuse of notation, we use S indistinctly to refer either to the subset of explanatory variables of the model M_S, or to the subindex set corresponding to those variables. The family of all possible FH models (2.3), for any subset $S \subseteq P$ of explanatory variables is given by $\mathcal{M} = \{M_S : S \in \mathcal{P}(P)\backslash\emptyset\}$, where $\mathcal{P}(P)$ denotes the power set of set P. There are $2^{|P|} - 1$ possible model candidates which can be fitted to the data.

2.2 Information criteria

In practice, an information criterion (IC) is used to find the model M_S, $S \subseteq P$, that minimizes an estimate of the criterion. For a model M_S, the ICs are generally of the form $g(S) = Q(\hat{\boldsymbol{\theta}}_S) + \alpha(d_S)$, where $\hat{\boldsymbol{\theta}}_S$ is the vector of estimated model parameters, Q is the loss function and α is the penalty function that takes into account the model complexity d_S. For candidate models M_S and M_R, such that $R \subseteq S \subseteq P$, the loss function satisfies $Q(\hat{\boldsymbol{\theta}}_S) \leq Q(\hat{\boldsymbol{\theta}}_R)$. Among the ICs most used in LMMs, we find the expected Kullback's divergence (AIC family) and the expected Kullback's symmetric divergence (KIC family). Given $\boldsymbol{x}_{Pd}$, $d = 1, \ldots, D$, these ICs are functions from $\mathcal{P}(P)$ to $\mathbb{R}$.

2.2.1 AIC family

Let us assume that the data-generating model (true model) fulfils equation (2.3) for $T \subseteq P$. Omitting the subscript T, we denote the true parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_u^2)$, where $\sigma_u^2 > 0$ and $\beta_i \neq 0$ for all $i \in T$. We refer to $f(\boldsymbol{y}|\boldsymbol{\theta})$ as the p.d.f. of the true generating model and $f(\boldsymbol{y}|\boldsymbol{\theta}_S)$ as the approximating p.d.f. associated with some $S \subseteq P$. The Kullback–Leibler divergence between $f(\boldsymbol{y}|\boldsymbol{\theta})$ and $f(\boldsymbol{y}|\boldsymbol{\theta}_S)$ with respect to $f(\boldsymbol{y}|\boldsymbol{\theta})$,

$$I(\boldsymbol{\theta}, \boldsymbol{\theta}_S) = E_\theta \left[\log \frac{f(\boldsymbol{y}|\boldsymbol{\theta})}{f(\boldsymbol{y}|\boldsymbol{\theta}_S)} \right],$$

reflects the separation between the true p.d.f. $f(\boldsymbol{y}|\boldsymbol{\theta})$ and the approximating p.d.f. $f(\boldsymbol{y}|\boldsymbol{\theta}_S)$. Let $\hat{\boldsymbol{\theta}_S}$ be the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}_S$ under model M_S. The most common estimator of the expected Kullback-Leibler divergence $E_{\boldsymbol{\theta}}[I(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}_S})]$ is the AIC statistic

$$\text{AIC} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_S) + 2(|S| + 1).$$

As AIC is not an unbiased estimator of $E_{\boldsymbol{\theta}}[I(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}_S})]$, Hurvich and Tsai (1989) introduced the bias-corrected AIC for small sample sizes,

$$\text{AICc} = \text{AIC} + \frac{2|S|^2 + 2|S|}{D - |S| - 1},$$

and Marhuenda et al. (2014) proposed the bias-corrected bootstrap variants

$$\text{AIC}_{b1} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_S) + \frac{1}{B}\sum_{b=1}^{B} -2\log \frac{f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_S^{*(b)})}{f(\boldsymbol{y}_S^{*(b)}|\hat{\boldsymbol{\theta}}_S^{*(b)})},$$

$$\text{AIC}_{b2} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_S) + \frac{2}{B}\sum_{b=1}^{B} -2\log \frac{f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_S^{*(b)})}{f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_S)},$$

where $\{\hat{\boldsymbol{\theta}}_S^{*(b)}, b = 1, \ldots, B\}$ and $\{\boldsymbol{y}_S^{*(b)}, b = 1, \ldots, B\}$ are sets of B bootstrap replicates of $\hat{\boldsymbol{\theta}}_S$ and $\boldsymbol{y}$, respectively.

2.2.2 KIC family

This section considers the Kullback symmetric divergence criterion between $f(\boldsymbol{y}|\boldsymbol{\theta})$ and $f(\boldsymbol{y}|\boldsymbol{\theta}_S)$,

$$J(\boldsymbol{\theta}, \boldsymbol{\theta}_S) = E_{\boldsymbol{\theta}}\left[\log\frac{f(\boldsymbol{y}|\boldsymbol{\theta})}{f(\boldsymbol{y}|\boldsymbol{\theta}_S)}\right] + E_{\boldsymbol{\theta}_S}\left[\log\frac{f(\boldsymbol{y}|\boldsymbol{\theta}_S)}{f(\boldsymbol{y}|\boldsymbol{\theta})}\right],$$

that reflects the separation between the true p.d.f. $f(\boldsymbol{y}|\boldsymbol{\theta})$ and the approximating p.d.f. $f(\boldsymbol{y}|\boldsymbol{\theta}_S)$. The most common estimator of the expected Kullback symmetric divergence, $E_{\boldsymbol{\theta}}[J(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}_S})]$, is the KIC statistic

$$\text{KIC} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_S) + 3(|S| + 1).$$

Marhuenda et al. (2014) introduced the bias-corrected bootstrap variants

$$\text{KICc} = \text{AICc} + \hat{B}_{2*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^*),\ \ \text{KIC}_{b1} = \text{AIC}_{b1} + \hat{B}_{2*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^*),$$
$$\text{KIC}_{b2} = \text{AIC}_{b2} + \hat{B}_{2*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^*),$$

where

$$\hat{B}_{2*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^*) = \sum_{d=1}^{D}\log(\hat{\sigma}_{Su}^2 + \sigma_{ed}^2) + \frac{1}{B}\sum_{b=1}^{B}B_{3*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^{*(b)}) - \frac{1}{B_1}\sum_{b=1}^{B}B_{4*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^{*(b)}),$$

$$\hat{B}_{3*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^*) = \sum_{d=1}^{D}\left\{\frac{\hat{\sigma}_{Su}^{2*(b)} + \sigma_{ed}^2}{\hat{\sigma}_{Su}^2 + \sigma_{ed}^2} + \frac{(\boldsymbol{x}_{Sd}(\hat{\boldsymbol{\beta}}_S^{*(b)} - \hat{\boldsymbol{\beta}}_S))^2}{\hat{\sigma}_{Su}^2 + \sigma_{ed}^2}\right\},$$

$$\hat{B}_{4*}(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_S^*) = \sum_{d=1}^{D}\left\{\log(\hat{\sigma}_{Su}^{2*(b)} + \sigma_{ed}^2) + \frac{(y_{Sd}^{*(b)} - \boldsymbol{x}_{Sd}\hat{\boldsymbol{\beta}}_S^{*(b)})^2}{\hat{\sigma}_{Su}^{2*(b)} + \sigma_{ed}^2}\right\},$$

and $\hat{\boldsymbol{\theta}}_S^{*(b)}$, $\hat{\sigma}_{Su}^{2*(b)}$, $\hat{\boldsymbol{\beta}}_S^{*(b)}$ and $y_{Sd}^{*(b)}$, $d = 1, \ldots, D$, $b = 1, \ldots, B$, are the bootstrap replicates of the MLEs of $\hat{\boldsymbol{\theta}}_S$, $\hat{\sigma}_{Su}^2$ and $\hat{\boldsymbol{\beta}}_S$, and of the target variable y_d, $d = 1, \ldots, D$, respectively.

3 The cooperative game for variable selection

This section introduces an influence measure for the FH model, which quantifies the average marginal contribution of each auxiliary variable to overall model performance. This measure enables a ranking of variables and provides a robust approach for model selection. First, let us define an influence measure in this context.

Table 1 Regression parameters and p-values for the maximal model

Variable	EBLUE-ML	SE	p-value
X_1	0.803	0.231	0.000
X_2	0.953	0.194	0.000
X_3	0.190	0.210	0.364

Definition 1 A measure of influence for a model M_S, with $S \subseteq P$, is a rule I^g that assigns a vector $I^g(M_S) = (I^g_1(M_S), \ldots, I^g_{|S|}(M_S)) \in \mathbb{R}^{|S|}$ providing an evaluation of the influence that each auxiliary variable of S has on the target variable and given an IC g.

Among the possible tools for introducing influence measures, this section considers the Shapley value from cooperative game theory. To reinforce the theoretical foundation of this measure, we will present it axiomatically by identifying a set of desirable properties and proving the existence of a unique measure that satisfies them. As will be demonstrated below, only two properties are sufficient to fully characterize our measure. We begin by introducing some fundamental concepts from game theory. A cooperative game, or just a game, is a pair (P, v) where P is a non-empty finite set, and v is a function from $\mathcal{P}(P)$ to $\mathbb{R}$ with $v(\emptyset) = 0$. We say that P is the player set of the game and v is the characteristic function, and we usually identify (P, v) with its characteristic function v. Definition 2 associates with the FH model (2.3) a cooperative game, for a given IC g.

Definition 2 Given the maximal FH model M_P (including all potential auxiliary variables P) and an IC g, the (M_P, g) cooperative game is the pair $(P, v_g^{M_P})$, where $v_g^{M_P}(S) = g(S)$ for all $S \subseteq P$.

In the (M_P, g) game, the value of the IC g is calculated for every subset $S \subseteq P$ of explanatory variables on the corresponding submodel M_S. Consider the following example to illustrate this.

Example 1 We aim to model a linear relationship between a response variable Y and three explanatory variables X_1, X_2 and X_3[1] using the FH model (2.3). First, a maximal model is fitted, including all the auxiliary variables to assess their significance. Table 1 displays the results of the linear regression coefficients (EBLUE-ML), the standard errors (SE) and the p-values. The EBLUE-ML estimators, following Equation (2.5), are obtained using the ML estimator $\hat{\sigma}^2_{Pu}$ of σ^2_{Pu}. In addition, Table 2 presents the KIC computed for each subset $S \subseteq P = \{1, 2, 3\}$, that is, the game $(P, v^{M_P}_{KIC})$. The auxiliary variables, X_1 and X_2, which were identified in Table 1 as significant of the maximal model, form the model with the lowest KIC (game value) in Table 2.

Returning to game theory, the main goal in a cooperative game is to allocate the value $v(P) = v_g^{M_P}(P)$ of the grand coalition among all players in a fair way. A well-known

[1] Data and its description can be found in the following GitHub repository: https://github.com/small-area-estimation/Variable-Selection-for-Fay-Herriot-Models-A-Cooperative-Game-Theory-Approach#.

Table 2 KIC for each subset of explanatory variables

Model	S	$v_{KIC}^{M_P}(S)$
X_1	$\{1\}$	216.24
X_2	$\{2\}$	206.01
X_3	$\{3\}$	227.16
$X_1 + X_2$	$\{1, 2\}$	**191.84**
$X_1 + X_3$	$\{1, 3\}$	211.47
$X_2 + X_3$	$\{2, 3\}$	202.28
$X_1 + X_2 + X_3$	$\{1, 2, 3\}$	194.98

The lowest KIC is shown in bold

allocation rule is the Shapley value. The Shapley value is a map Φ that associates with every cooperative game (P, v) a vector $\Phi(v) \in \mathbb{R}^p$ that satisfies $\sum_{i \in P} \Phi_i(v) = v(P)$ and provides a fair allocation of $v(P)$ to players in P. The explicit formula of the Shapley value for every cooperative game (P, v) and every $i \in P$ is

$$\Phi_i(v) = \sum_{S \subseteq P \setminus i} \frac{(|P| - |S| - 1)!\,|S|!}{|P|!}(v(S \cup i) - v(S)), \tag{3.1}$$

where "!" denotes the factorial function, defined for non-negative integers.

Since its introduction by Shapley (1953), the Shapley value has proven to be one of the most important allocation rules in cooperative game theory and has applications in many practical problems (see, for instance, Algaba et al. (2019)). In our context, given an FH model (2.3) and its corresponding cooperative game, we define the following influence measure.

Definition 3 Given the maximal FH model M_P, we define an influence measure based on the Shapley value and a given IC g as $I^{\Phi,g}(M_P) = \Phi(v_g^{M_P}) \in \mathbb{R}^p$.

Note that the i-th component of the Shapley value for the cooperative game $(P, v_g^{M_P})$ quantifies the contribution of explanatory variable $i \in P$ to the associated IC g. Criteria such as AIC and KIC families measure information loss; therefore, a lower Shapley value for an explanatory variable indicates that, on average, its inclusion in the model leads to a greater reduction in information loss. To illustrate this, we return to Example 1 and compute the proposed influence measure.

Example 2 We compute the Shapley influence measure for the explanatory variables X_1, X_2 and X_3 introduced in Example 1, obtaining the following values, respectively: 64.37, 54.66 and 75.95, which add up to 194.98. These results indicate that including X_1 and X_2 in the model leads to a lower average information loss compared to X_3. In particular, X_2 yields the smallest value, suggesting it is the most influential variable with respect to the KIC criterion.

Let us now define the properties of interest for an influence measure. The first property we require for an influence measure is that it be efficient, that is, that it distributes the total influence $g(P)$, given by an IC g evaluated in the FH model M_P,

among all the explanatory variables of the set P. Given a FH model M_P and a IC g, an influence measure is (g, P)-*Efficiency* iff

$$\sum_{1 \leq i \leq |P|} I_i^g (M_P) = g(P). \tag{3.2}$$

The second property is a fairness property that treats all the explanatory variables in the same way. If one of two variables is disregarded, the resulting impact on the influence of the other is equivalent to the impact of disregarding the latter on the former. Note that the marginal loss or gain of influence caused by the inclusion or exclusion of one variable on another is a result of the existing dependence between the two. Given a FH model M_P and an IC g, an influence measure satisfies *Balanced Influence* iff for each explanatory variables $i, j \in P$, with $i \neq j$, it holds that

$$I_i^g (M_P) - I_i^g \left(M_{P \setminus \{j\}}\right) = I_j^g (M_P) - I_j^g \left(M_{P \setminus \{i\}}\right). \tag{3.3}$$

Theorem 1 states that the Shapley value is the unique influence measure that satisfies the above two properties. The proof is included in the Appendix B.

Theorem 1 *There exists a unique influence measure for the FH model M_P satisfying equations (3.2) and (3.3), i.e, the properties of (g, P)-Efficiency and Balanced Influence. This measure is given by $I^{\Phi,g}(M_P)$.*

Using the Shapley measure of influence, a ranking of the most influential variables in the model is obtained. This ranking can be used to construct an algorithm for model selection. Given a maximal FH model M_P and an IC g, Algorithm 1 selects a subset of auxiliary variables. Example 3 applies Algorithm 1 to select the auxiliary variables of Examples 1 and 2.

Algorithm 1 Shapley algorithm for FH model selection.

1: Calculate $(P, v_g^{M_P})$ and $I^{\Phi,g}(M_P)$. ▷ Initialization.
2: Sort the vector $I^{\Phi,g}(M_P)$ in ascending order, i.e, $I^{\Phi,g}(M_P)_{sort} = (I^{\Phi,g}_{[1]}(M_P), \ldots, I^{\Phi,g}_{[P]}(M_P))$, where $[i]$ denotes the index of the i-th smallest element in the original vector $I^{\Phi,g}(M_P)$, corresponding to the i-th auxiliary variable in the sorted order.
3: **for** $i \in P$
4: Calculate the IC $g(\{[1], \ldots, [i]\})$ for the nested model $M_{\{[1],\ldots,[i]\}}$.
5: **end for**
6: Select the model $M_{\{[1],\ldots,[i]\}}$ such that all explanatory variables are significant at α level and with the lowest $g(\{[1], \ldots, [i]\})$. ▷ Finalisation.

Example 3 We rank the variables in ascending order according to the Shapley influence measure based on the KIC criterion. Moreover, we compute the KIC value for the nested models, i.e., by sequentially adding the auxiliary variables with lower influence values to the model. Table 3 presents the nested models, the corresponding KIC values, and the p-values of the explanatory variables from Examples 1 and 2. As we can see

Table 3 KIC-values and p-values for nested models ordered by Shapley influence measure

Model	KIC	p-values
X_2	216.25	0.000
$X_2 + X_1$	191.89	0.000, 0.000
$X_2 + X_1 + X_3$	194.08	0.000, 0.000, 0.364

in Table 3, the selected model would be the one formed by the explanatory variables X_1 and X_2.

The calculation of the Shapley influence measure requires evaluating the IC of any possible submodel M_S, $S \subseteq P$. At first glance, one might argue that if we can calculate the IC for all possible submodels, then we could simply select the best one directly, making the ranking of variable influences seemingly unnecessary. However, it is important to note that both tasks, although related, address different objectives. While model selection seeks the best combination of variables, influence measurement quantifies the individual contribution of each variable to the model's performance.

Moreover, evaluating all submodels is computationally infeasible when the number of explanatory variables is large, because of the exponential growth in the number of possible models. To address this, and inspired by Castro et al. (2009), we propose a sampling-based algorithm that allows us to efficiently estimate the influence measure in polynomial time, without exhaustively evaluating every submodel. Algorithm 2 outlines the details of this procedure.

Importantly, by ranking variables according to their estimated influence, we also obtain a practical approach to model selection, providing a natural path to identify high-performing submodels without full enumeration. The algorithm consists of selecting m samples uniformly at random from all possible permutations of the set of variables, $\Pi(P)$. For each sample, the marginal contribution of each variable with respect to the variables preceding it is computed. The resulting estimator, $\hat{I}^{\Phi,g}$, is defined as the mean of the marginal contributions across the m samples. It is important to note that the proposed estimator is unbiased and consistent (see Propositions 1 and 2, Appendix C). These properties are essential to ensure the accuracy and reliability of the estimator.

Algorithm 2 Estimation $I^{\Phi,g}(M_P)$.

1: Determine m. Fix $cont = 0$ and $\hat{I}^{g}(M_P) = 0$. ▷ Initialization.
2: **while** $cont < m$ **do**
3: Take $\sigma \in \Pi(P)$ with probability $\frac{1}{|P|!}$.
4: **for** $i \in P$
5: Calculate $L_i^{\Phi,g}(\sigma) = g(\{\sigma(1), \ldots, \sigma(k)\} \cup \{i\}) - g(\{\sigma(1), \ldots, \sigma(k)\})$, where $\sigma(k+1) = i$.
6: Calculate $\hat{I}_i^{\Phi,g}(M_P) = \hat{I}_i^{\Phi,g}(M_P) + L_i^{\Phi,g}(\sigma)$.
7: **end for**
8: $cont = cont + 1$.
9: **end while**
10: $\hat{I}^{\Phi,g}(M_P) = \frac{1}{m}\hat{I}^{\Phi,g}(M_P)$. ▷ Finalisation.

While Examples 1, 2, and 3 illustrate that Shapley's algorithm for model selection and the KIC criterion can lead to the same model choice in controlled settings, such alignment cannot be expected in real-data applications. We point out two main facets that must be taken into account in the analysis of real data. First, the variables most highly ranked by the Shapley value are those with the highest average contribution across all model configurations; it does not necessarily follow that they will constitute the best model for a given IC. Second, ICs such as AIC and KIC are typically carried out by heuristic or stepwise procedures, which are either model or search strategy-dependent and might not necessarily locate the global optimum. These observations highlight the importance of not overrating AIC and KIC estimators, which may be subject to high sampling variability, and considering an alternative approach that, while not a model selection tool itself, provides a robust perspective on the role of explanatory variables in model selection.

4 Simulation experiments

This section examines the performance of the Shapley influence measure for variable selection in the FH model. As ICs, we take the AIC and the KIC, along with their bias-corrected variants, AICc, AIC_{b1}, AIC_{b2}, KICc, KIC_{b1}, and KIC_{b2}, described in Section 2.2. The target variable is generated from an FH model M_T with $|T| = 5$ auxiliary variables serving as the true data-generating auxiliary variables. Alternative models M_S are considered under two conditions: underfitted models, where $1 \leq |S| < 5$, and overfitted models, where $5 < |S| \leq 10$.

We pursue two main objectives. The first objective is to evaluate the ability of the Shapley influence measure to correctly identify the true data-generating auxiliary variables. Specifically, we assess whether the most relevant auxiliary variables in the true model M_T consistently rank among the top five most influential variables according to the influence measure. The second objective is to examine the performance of Shapley's algorithm introduced in Section 3 under usual statistical significance levels ($\alpha = 0.05, 0.1$).

Following a simulation experiment inspired by Marhuenda et al. (2014), we analyse four different scenarios where the importance of the true explanatory variables is systematically decreased. The true model parameters are denoted as $\boldsymbol{\beta}_T \in$

$\{\boldsymbol{\beta}_{T0}, \boldsymbol{\beta}_{T1}, \boldsymbol{\beta}_{T2}, \boldsymbol{\beta}_{T3}\}$, where $\boldsymbol{\beta}_{T0} = (1, 1, 1, 1, 1)'$, $\boldsymbol{\beta}_{T1} = (1, 1, 1, 1, 0.25)'$, $\boldsymbol{\beta}_{T2} = (1, 1, 1, 0.25, 0.25)'$, $\boldsymbol{\beta}_{T3} = (1, 1, 0.25, 0.25, 0.25)'$.

The auxiliary variables are simulated as $x_{dj} \sim U(0, 5)$, $d = 1, \ldots, D$, $j = 1, \ldots, |P|$, where U denotes uniform distribution and $|P| = 10$. As the range of values of the x-variables is $(0, 5)$, the beta parameters play the role of weights in the model linear equation. The set of subindexes of the generating auxiliary variables is $T = \{1, \ldots, 5\} \subset P = \{1, \ldots, 10\}$. The variance of the random effects is $\sigma^2_{Tu} = 1$. The sampling error variances are $\sigma^2_{ed} = \lambda_0 + \frac{(\lambda_1 - \lambda_0)(d-1)}{D-1}$, with $\lambda_0 = 0.8$ and $\lambda_1 = 1.2$ for each $d = 1, \ldots, D$. In the case of the estimator of the Shapley influence measure, m denotes the sample size.

The simulation experiment is conducted with $K = 2000$ Monte Carlo iterations and with the following steps.

1. Repeat K times ($k = 1, \ldots, K$)
   - 1.1. For $d = 1, \ldots, D$, generate $\boldsymbol{x}^{(k)}_{Pd} = (x^{(k)}_{d1}, \ldots, x^{(k)}_{d10})$, with $x^{(k)}_{di} \sim U(0, 5)$, $i = 1, \ldots, 10$.
   - 1.2. For $d = 1, \ldots, D$, generate, $u^{(k)}_{Td} \sim N(0, \sigma^2_{Tu})$, $e^{(k)}_{d} \sim N(0, \sigma^2_{ed})$, $y^{(k)}_{d} = \boldsymbol{x}^{(k)}_{Td}\boldsymbol{\beta}_T + u^{(k)}_{Td} + e^{(k)}_{d}$, where $\boldsymbol{x}^{(k)}_{Td} = (x^{(k)}_{d1}, \ldots, x^{(k)}_{d5})$.
   - 1.3. For all $S \subseteq P = \{1, \ldots, 10\}$, calculate
     - 1.3.1 the ML estimates $\hat{\boldsymbol{\theta}}^{(k)}_S = (\hat{\boldsymbol{\beta}}'^{(k)}_S, \hat{\sigma}^{2(k)}_{Su})'$, and
     - 1.3.2 the characteristic function $v_g^{M^{(k)}_P}(S) = g^{(k)}(S)$, where $g^{(k)}(S)$ denotes the value of the information criterion g applied to model $M^{(k)}_S$, i.e, g evaluated at the ML estimate $\hat{\boldsymbol{\theta}}^{(k)}_S$.
   - 1.4. Apply Shapley's algorithm for model selection.
     - 1.4.1. Calculate $I^{\Phi,g,(k)} = I^{\Phi,g}(M^{(k)}_P)$.
     - 1.4.2. Sort the vector $I^{\Phi,g,(k)}$ in ascending order, i.e,

$$I^{\Phi,g,(k)}_{sort} = \left(I^{\Phi,g}_{[1]}(v_g^{M^{(k)}_P}), \ldots, I^{\Phi,g}_{[10]}(v_g^{M^{(k)}_P})\right),$$

       where $[i]$ is the index of the i-th auxiliary variable in the sorted vector $I^{\Phi,g,(k)}_{sort}$.
     - 1.4.3. For $i = 1, \ldots, 10$, calculate the IC $g^{(k)}(\{[1], \ldots, [i]\})$ for the nested model $M_{\{[1],\ldots,[i]\}}$.
     - 1.4.4. For $\alpha = 0.05, 0.10$, select the model $M^{(k)}_{\{[1],\ldots,[i]\}}$ such that all explanatory variables are significant at α level and have the lowest $g^{(k)}(\{[1], \ldots, [i]\})$.
2. For $i = 1, \ldots, 10$, calculate the proportion of Monte Carlo iterations that the i-th variable is classified among the top five variables according to the Shapley influence measure, i.e.

$$q_i = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\left(I^{\Phi,g}_{[i]}(v_g^{M^{(k)}_P}) \leq I^{\Phi,g}_{[5]}(v_g^{M^{(k)}_P})\right),$$

where $\mathbb{1}(\cdot)$ is the indicator function.

3. For $\alpha = 0.05, 0.1, i = 1, \ldots, 10$, calculate the proportion of Monte Carlo iterations that the model $M^{(k)}_{\{[1],\ldots,[i]\}}$ is selected as the best model, according to Step 1.4.4.

4.1 Simulation 1. Variable Importance

To assess the ability of the Shapley influence measure to identify influential auxiliary variables, this section calculates the set of variables that rank among the top five based on their influence, for $D = 50$ domains, four regression parameter vectors, and eight ICs. Tables 4 and 5 report the proportion of times each auxiliary variable appears in this top-five ranking across $K = 2000$ Monte Carlo iterations. These proportions indicate how consistently a variable is identified as influential, averaged over all possible models. The bias-corrected bootstrap variants have been calculated with $B = 200$ replicates and $D = 50$ domains. In the ideal scenario $\boldsymbol{\beta}_{T0}$, the Shapley influence measure successfully identifies all five generating variables with 100% accuracy across all criteria. As we move to more challenging scenarios $\boldsymbol{\beta}_{T1}$, $\boldsymbol{\beta}_{T2}$ and $\boldsymbol{\beta}_{T3}$, the detection proportion decreases, particularly for the variables with lower coefficients. However, Tables 4 and 5 show that, even in the worst scenarios, the percentage detection of the less relevant generating auxiliary variables, x_3, x_4 and x_5, varies between 50% and 70%. This suggests that selection based on the Shapley influence measure is effective, even when the influence of the auxiliary variable is weak.

Tables 4 and 5 show that the marginal Shapley influence measure associated with the least influential variables tends to increase as the overall number of low-impact variables in the model grows. This behaviour occurs because, by incorporating more low-weight variables, the mean marginal contributions of each auxiliary variable to the target variable become more detectable, increasing the capacity of the influence measure to identify the true data-generating explanatory variables. Simulation 1 empirically demonstrates that our proposed influence measure remains constant regardless of the IC selected.

Note that the regression parameter estimates are consistent under the FH model (2.3). Therefore, increasing the number of domains D (typically, $D = 100, 150$ in real-world applications) results in better model fit and better performance of the influence measures. Thus, to test the ability of the Shapley influence measure to detect strongly predictive explanatory variables, we have decided to present simulation results for $D = 50$ domains.

Table 12 of Appendix E presents the results of the same simulation experiment, but using the estimator of the Shapley influence measure described in Algorithm 2 with $m = 300$ samples. Since the Shapley estimator is computed via Monte Carlo sampling of random permutations (equivalently, of the induced coalitions), using approximately 30% of the total, it is expected that the corresponding proportions of iterations in which each auxiliary variable is ranked among the top five are slightly lower. However, the results of the different scenarios are similar to those obtained using the Shapley influence measure.

Table 4 Proportion of iterations in which the i-th auxiliary variable is ranked among the top five according to the Shapley influence measure, in the scenarios $\boldsymbol{\beta}_{T0}$ (top), $\boldsymbol{\beta}_{T1}$ (bottom), $D = 50$

	$\boldsymbol{\beta}_{T0}$							
i	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
5	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
6	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	$\boldsymbol{\beta}_{T1}$							
i	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.512	0.512	0.519	0.520	0.525	0.515	0.514	0.517
6	0.090	0.090	0.084	0.084	0.087	0.087	0.089	0.087
7	0.097	0.097	0.098	0.099	0.098	0.099	0.099	0.101
8	0.103	0.103	0.098	0.100	0.095	0.098	0.101	0.098
9	0.099	0.099	0.100	0.099	0.097	0.103	0.100	0.101
10	0.099	0.099	0.101	0.098	0.098	0.098	0.097	0.096

4.2 Simulation 2. Shapley algorithm for model selection

Beyond quantifying variable importance, the Shapley influence measure can be leveraged to construct a model selection procedure, described in Section 3. Tables 6 and 7 summarize the outcomes of the full model selection procedure based on the Shapley algorithm described in Section 3. For each Monte Carlo iteration $k = 1, \ldots, K$, the ranking of variables determined by the Shapley influence measure varies. Consequently, the nested models evaluated in each iteration differ in structure, even under the same IC. This situation introduces two important limitations: substantial computational costs and the impossibility of establishing a global ranking of models over the K iterations.

It is worth recalling that, in this part of the simulation, the true set $T = \{1, 2, 3, 4, 5\}$ has cardinality $|T| = 5$. Specifically, the variables x_{dj}, $j = 1, \ldots, 5$, are the true data-generating auxiliary variables. To provide a comprehensive summary, the columns 1, 2, 3, 4 and 5 of Tables 6 and 7 report the proportion of iterations in which the selected

Table 5 Proportion of iterations in which the i-th auxiliary variable is ranked among the top five according to the Shapley influence measure, in the scenarios $\boldsymbol{\beta}_{T2}$ (top), $\boldsymbol{\beta}_{T3}$ (bottom), $D = 50$

	$\boldsymbol{\beta}_{T2}$							
i	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.633	0.633	0.646	0.649	0.646	0.649	0.651	0.644
5	0.624	0.624	0.633	0.631	0.633	0.622	0.624	0.629
6	0.131	0.131	0.128	0.130	0.128	0.128	0.130	0.129
7	0.141	0.141	0.140	0.141	0.138	0.141	0.142	0.139
8	0.162	0.162	0.152	0.152	0.153	0.153	0.153	0.151
9	0.140	0.140	0.135	0.134	0.136	0.137	0.136	0.139
10	0.169	0.169	0.166	0.163	0.166	0.170	0.164	0.169
	$\boldsymbol{\beta}_{T3}$							
i	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.694	0.694	0.694	0.689	0.696	0.690	0.686	0.688
4	0.707	0.707	0.714	0.714	0.710	0.712	0.709	0.708
5	0.701	0.701	0.702	0.700	0.706	0.700	0.697	0.705
6	0.166	0.166	0.167	0.167	0.167	0.167	0.168	0.170
7	0.184	0.184	0.178	0.183	0.180	0.181	0.183	0.181
8	0.186	0.186	0.185	0.188	0.187	0.185	0.189	0.186
9	0.177	0.177	0.177	0.175	0.174	0.179	0.181	0.179
10	0.185	0.185	0.183	0.184	0.180	0.186	0.187	0.183

model M_S is such that $S \subset T$ with $|S| = j$ with $j = 1, \ldots, 5$. This is to say, M_S contains exclusively true data-generating auxiliary variables, from one to five. The column 0 reports the proportion of iterations in which the selected model M_S is such that $S \subsetneq T$. This is to say, M_S includes at least one non-generating auxiliary variable. When the first five columns are added up, that is, the true model M_T with $|T| = 5$, or the underfitted "quasi-correct" models M_S with $S \subset T$ and $|S| = 1, 2, 3, 4$, across all ICs and simulation scenarios, the results indicate that the algorithm effectively identifies models containing the true explanatory variables in approximately 65% to 90% of the iterations.

Under the scenario $\boldsymbol{\beta}_{T0}$, for $\alpha = 0.05$ or $\alpha = 0.1$, the algorithm performs consistently well across all ICs. Specifically, the proportion of times the selected model contains all five true variables exceeds 82% under the KIC-based criteria and remains above 78% for AIC-based variants. The intermediate scenarios $\boldsymbol{\beta}_{T1}$ and $\boldsymbol{\beta}_{T2}$ tend to decrease the rate of success identifying the true model M_T with $|T| = 5$, or the underfitted "quasi-correct" models. Despite this, the percentages of correct identification

Table 6 Proportions of times that Algorithm 1 selects FH models M_S such that $S \subset T$ and $|S| = j$, $j = 1, \ldots, 5$. The column 0 is for $S \subsetneq T$. The type I error probability is $\alpha = 0.05$

	$\boldsymbol{\beta}_{T0}$						$\boldsymbol{\beta}_{T1}$					
$\alpha = 0.05$	1	2	3	4	5	0	1	2	3	4	5	0
AIC	0.00	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.50	0.35	0.15
AICc	0.00	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.50	0.34	0.15
AIC_{b1}	0.00	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.50	0.34	0.15
AIC_{b2}	0.00	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.50	0.35	0.16
KIC	0.00	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.50	0.35	0.15
KICc	0.00	0.00	0.00	0.00	0.90	0.10	0.00	0.00	0.00	0.54	0.32	0.13
KIC_{b1}	0.00	0.00	0.00	0.00	0.90	0.10	0.00	0.00	0.00	0.54	0.32	0.14
KIC_{b2}	0.00	0.00	0.00	0.00	0.90	0.10	0.00	0.00	0.00	0.54	0.32	0.13
	$\boldsymbol{\beta}_{T2}$						$\boldsymbol{\beta}_{T3}$					
$\alpha = 0.05$	1	2	3	4	5	0	1	2	3	4	5	0
AIC	0.00	0.00	0.21	0.43	0.14	0.21	0.00	0.05	0.32	0.33	0.06	0.25
AICc	0.00	0.00	0.22	0.44	0.14	0.20	0.00	0.05	0.33	0.32	0.06	0.24
AIC_{b1}	0.00	0.00	0.22	0.43	0.14	0.21	0.00	0.04	0.33	0.32	0.06	0.25
AIC_{b2}	0.00	0.00	0.21	0.43	0.14	0.22	0.00	0.04	0.32	0.33	0.06	0.25
KIC	0.00	0.00	0.21	0.43	0.14	0.21	0.00	0.05	0.32	0.33	0.06	0.25
KICc	0.00	0.00	0.25	0.44	0.13	0.18	0.00	0.05	0.36	0.31	0.05	0.23
KIC_{b1}	0.00	0.00	0.26	0.43	0.13	0.18	0.00	0.05	0.37	0.30	0.05	0.23
KIC_{b2}	0.00	0.00	0.24	0.44	0.12	0.19	0.00	0.05	0.35	0.32	0.05	0.23

by the true model vary between 14% and 35%. The sum of proportions in columns 1-5 gives percentages greater than 70% of identifying models that do not make the mistake of including non-generating explanatory variables.

On the other hand, notable differences are observed between the AIC and KIC-based estimators, particularly as the type I error probability α increases. The proportion of correct model selections differs by approximately 5 to 10 percentage points, with KIC-based criteria exhibiting superior performance. This highlights the greater sensitivity of the AIC family to the choice of significance level and its comparatively lower robustness in detecting the true data-generating variables. Finally, the scenario $\boldsymbol{\beta}_{T3}$ (only two of the five auxiliary variables have high weights) presents relatively low percentages of detection of the correct model with $|T| = 5$ generating auxiliary variables, which vary between 5% and 7% in all criteria. However, the underfitted "quasi-correct" models with $|S| = 3, 4$ auxiliary variables are chosen more than 50% of the time. This behaviour indicates a tendency towards parsimonious models.

In contrast to the simulation results reported by Marhuenda et al. (2014), it is important to highlight a key methodological difference in our framework. The model selection algorithm proposed here does not assume prior knowledge of the true data-generating process; rather, it builds models sequentially based on variable importance derived from the Shapley influence measure. As such, the lower selection rates

Table 7 Proportions of times that Algorithm 1 selects FH models M_S such that $S \subset T$ and $|S| = j$, $j = 1, 2, 3, 4, 5$. The column 0 is for $S \not\subseteq T$. The type I error probability is $\alpha = 0.1$

	$\boldsymbol{\beta}_{T0}$						$\boldsymbol{\beta}_{T1}$					
$\alpha = 0.1$	1	2	3	4	5	0	1	2	3	4	5	0
AIC	0.00	0.00	0.00	0.00	0.78	0.22	0.00	0.00	0.00	0.40	0.35	0.25
AICc	0.00	0.00	0.00	0.00	0.81	0.19	0.00	0.00	0.00	0.43	0.34	0.23
AIC_{b1}	0.00	0.00	0.00	0.00	0.81	0.19	0.00	0.00	0.00	0.44	0.33	0.23
AIC_{b2}	0.00	0.00	0.00	0.00	0.81	0.19	0.00	0.00	0.00	0.43	0.34	0.23
KIC	0.00	0.00	0.00	0.00	0.82	0.18	0.00	0.00	0.00	0.43	0.35	0.23
KICc	0.00	0.00	0.00	0.00	0.85	0.15	0.00	0.00	0.00	0.51	0.32	0.17
KIC_{b1}	0.00	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.50	0.33	0.17
KIC_{b2}	0.00	0.00	0.00	0.00	0.86	0.14	0.00	0.00	0.00	0.52	0.32	0.16
	$\boldsymbol{\beta}_{T2}$						$\boldsymbol{\beta}_{T3}$					
$\alpha = 0.1$	1	2	3	4	5	0	1	2	3	4	5	0
AIC	0.00	0.00	0.17	0.37	0.17	0.28	0.00	0.03	0.21	0.34	0.07	0.35
AICc	0.00	0.00	0.19	0.39	0.17	0.26	0.00	0.03	0.24	0.32	0.06	0.34
AIC_{b1}	0.00	0.00	0.19	0.39	0.17	0.25	0.00	0.03	0.24	0.33	0.07	0.33
AIC_{b2}	0.00	0.00	0.19	0.39	0.17	0.26	0.00	0.03	0.23	0.33	0.07	0.34
KIC	0.00	0.00	0.20	0.38	0.16	0.25	0.00	0.03	0.24	0.34	0.06	0.33
KICc	0.00	0.00	0.24	0.41	0.14	0.20	0.00	0.05	0.28	0.32	0.05	0.29
KIC_{b1}	0.00	0.00	0.25	0.41	0.13	0.21	0.00	0.05	0.29	0.32	0.05	0.30
KIC_{b2}	0.00	0.00	0.23	0.43	0.14	0.20	0.00	0.05	0.28	0.32	0.05	0.30

observed for the full generating model with $|T| = 5$ variables, particularly in scenarios with lower weights, are to be expected. This finding reflects the inherent difficulty of identifying low-impact variables without structural assumptions and illustrates the strength of the Shapley-based approach in uncovering influential auxiliary variables in a data-driven setting, or, at the very least, in providing reliable underestimates that remain close to the true generating model.

The results obtained by the simulation experiment for Shapley's algorithm for model selection suggest that the KIC family of estimators is particularly well-suited, with the KIC_{b2} estimator producing better performance using at least $B = 200$ bootstrap iterations. Furthermore, we propose a moderate type I error ($\alpha \leq 0.1$) to prevent both selecting models involving non-significant explanatory variables and too stringent thresholds. The balance between statistical significance and efficiency supports the use of KIC_{b2} in practice, particularly where interpretability of models and explanatory variable importance are most relevant.

5 Application to SLCS 2022

This section applies the proposed new methodology to establish a classification of the best explanatory variables and models for predicting poverty proportions in Span-

ish provinces. We take data from the SLCS 2022. The dependent variable in the Fay–Herriot model is the direct estimator of the province−sex poverty proportion, calculated from the SLCS 2022 unit-level data under the sampling design. The primary aim of the SLCS is to systematically produce statistical information regarding income, living conditions, poverty levels, and social exclusion. This annual survey is designed to provide detailed statistical analyses and classifications at the autonomous community level. However, due to smaller sample sizes at the cross province-sex, direct estimators are challenged to provide reliable estimates at this level. Consequently, the use of model-based predictors ensures a better approach to overcome these limitations by allowing the inclusion of auxiliary variables, correlation structures and information from other domains.

The target population is made up of people aged 16 and over, with permanent residence in Spain. For this reason, respondents under 16 are excluded, as they fall below the legal minimum working age. Consequently, the size of the survey data file is reduced to n = 50,000 working-age respondents, approximately. Then, respondents were classified into domains defined by the variables province of residence and sex (1: male, 2: female). As regards territorial division, Spain has 52 provinces, including the autonomous cities of Ceuta and Melilla. There are $D = 52 \cdot 2 = 104$ domains, defined by the crosses of province and sex, respectively.

The auxiliary variables are taken from the Spanish Labour Force survey (SLFS) of 2022. The SLFS adopts a quarterly survey format with a rotating panel design. Sample selection uses a two-stage design with stratification involving census sections in the first stage and main family dwellings, i.e., those used as a permanent or habitual residence throughout or for most of the year, in the second stage. Each quarterly SLFS sample size exceeds that of the annual SLCS sample. For the FH model selection, the auxiliary variables are the domain-level proportions, calculated using the formula of the Hájek direct estimator for domain means (Morales et al. 2021, see][p.23), of the categories of the unit-level variables listed below. The formula is applied to the data of the four 2022 SLFS samples together, which represents a sample size much larger than the 2022 SLCS sample size. The domain-level auxiliary variables are

Age group, categorized into five groups ($age1$-$age5$): (0, 15], [16, 24], [25, 49], [50, 64], and [65, ∞), respectively.
Citizenship, divided into ($nac1$-$nac2$): Spanish ($nac1$) and non-Spanish ($nac2$).
Education, split into four levels ($edu0$-$edu3$): $edu0$ for less than primary education, $edu1$ for primary education, $edu2$ for secondary education, and $edu3$ for including short-cycle tertiary education, bachelor's degrees, master's degrees and doctoral degrees.
Labour status, grouped into four levels ($lab0 - lab3$): $lab0$ ($\leq$ 15 years), $lab1$ (employed), $lab2$ (unemployed), and $lab3$ (inactive).

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 domain-level auxiliary variables are much larger than the corresponding SLCS sample sizes used to calculate the target variables. Extending our research to incorporate methodologies proposed by Ybarra and Lohr (2008) and Burgard et al. (2020) within

Table 8 Top five models with lowest IC in terms of the AIC and KIC variants

Model	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
M_{S1}	**-352.60**	**-346.60**	**-353.39**	-353.85	**-352.32**	**-342.66**	**-343.12**	**-341.59**
M_{S2}	-350.13	-345.13	-348.76	-350.59	-349.61	-339.58	-341.40	-340.42
M_{S3}	-351.40	-344.40	-352.34	**-354.03**	-351.79	-339.31	-341.00	-338.76
M_{S4}	-351.09	-344.09	-351.79	-352.89	-350.46	-338.39	-339.49	-337.06
M_{S5}	-350.70	-343.70	-348.95	-351.56	-348.18	-338.22	-340.83	-337.44

The lowest ICs are shown in bold

the context of the FH model (2.3) would require extensive efforts beyond the scope of our current research objectives.

Since the domain proportions of the categories of each unit variable sum to 1 and $age1$=$lab0$, we exclude $age1$, $nac2$, $edu0$, $lab0$ and $lab3$ from the final set P of possible auxiliary variables. Finally, the $|P| = 10$ auxiliary variables in P are $age2$, $age3$, $age4$, $age5$, $nac1$, $edu1$, $edu2$, $edu3$, $lab1$, $lab2$.

First, we conduct a model selection analysis utilizing the ICs described in Section 2.2. Using the classic AIC and KIC measures for selecting an FH model, we obtain the five best subsets $S \subseteq P$ of auxiliary variables. They are

$$
\begin{aligned}
S1 &= \{age3, age4, edu1, edu3, lab2\},\\
S2 &= \{age3, age4, edu3, lab2\},\\
S3 &= \{age3, age4, edu1, edu2, edu3, lab2\},\\
S4 &= \{age3, age4, nac1, edu1, edu3, lab2\},\\
S5 &= \{age3, age4, edu1, edu3, lab1, lab2\}.
\end{aligned}
$$

Table 8 displays the best five models, $M_{S1} - M_{S5}$, and their AIC and KIC bootstrap-corrected variants, calculated with $B = 200$ iterations. This table identifies the models that minimize information measures. The model M_{S1}, incorporating the explanatory variables $age3$, $age4$, $edu1$, $edu3$, $lab2$, achieves the best performance in 7 out of the 8 information measures proposed, and it is considered the reference model.

Table 9 classifies the ten explanatory variables according to the contributions, assigned by the Shapley influence measure, to the ICs on the maximal model M_P. The columns of Table 9 contains the vectors $I^{\Phi,g}(M_P) = \left(I_1^{\Phi,g}(M_P), \ldots, I_{10}^{\Phi,g}(M_P)\right)$, cf. Definition 3, for all the considered ICs g. The labels of the columns identify the IC. As discussed in Section 3, our interest is focused on minimizing some IC. Therefore, those variables with lower values are those with higher mean contribution. The three most important variables are $lab2$ (proportion of unemployed people), $edu3$ (proportion of people with high levels of education) and $edu2$ (proportion of people with a secondary level of education). In line with the results of the simulation experiment, the ranking remains stable across the different measures, which consolidates the robustness of the proposed methodology.

Note that the most important variables, as classified by our influence measure, are based on the average contribution to information loss for a given IC. As a result, they may not always appear in models with the lowest value of a given IC or neces-

Table 9 Classification of the ten auxiliary variables based on the Shapley influence measure for the ICs AIC, KIC, and their bootstrap-corrected variants

Variable	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
lab2	−62.46	−61.36	−62.50	−58.95	−62.33	−60.26	−56.72	−60.10
edu3	−43.26	−42.16	−42.88	−43.73	−42.94	−40.63	−41.48	−40.70
edu2	−37.34	−36.24	−37.05	−38.78	−37.06	−34.73	−36.54	−34.74
age3	−37.14	−36.04	−36.94	−36.87	−36.80	−34.69	−34.56	−34.56
age4	−33.30	−32.20	−33.05	−34.75	−33.11	−30.93	−32.62	−30.99
age2	−32.29	−31.19	−32.24	−33.75	−32.00	−30.01	−31.52	−29.78
nac1	−30.61	−29.51	−29.90	−30.64	−30.18	−27.61	−28.34	−27.88
lab1	−26.96	−25.86	−27.03	−27.56	−26.89	−24.90	−25.43	−24.76
age5	−23.28	−22.17	−23.02	−23.68	−22.97	−20.81	−21.47	−20.76
edu1	−20.48	−19.38	−20.41	−21.40	−20.27	−18.26	−19.25	−18.12

sarily lead to the model minimizing the IC. However, the Shapley influence measure remains useful for identifying high-quality auxiliary variables and understanding their contribution to the chosen IC.

We apply the Shapley Algorithm 1 for FH model selection, with the IC KIC_{b2}, to set P of explanatory variables. The algorithm successively evaluates 10 models, starting with the maximal model $M10 = M_P$ containing the available $|P| = 10$ explanatory variables. The remaining models are denoted by $M1, \ldots, M9$ since they depend on $1, \ldots, 9$ explanatory variables. On the other hand, we calculate KIC_{b2} for all the $2^{10} - 1 = 1023$ possible models. and we order them, according to the value of KIC_{b2}, from lowest (rank 1) to highest (rank 1023). The lowest value is reached in the model M_{S1} with a value $KIC_{b2} = -341.59$, as shown by Table 8. For every model M, the relative inefficiency (in %) is $RI(M) = 100\left(1 - KIC_{b2}(M)/KIC_{b2}(M_{S1})\right)$. Table 10 presents the ten nested models, $M1$-$M10$, evaluated by Algorithm 1, and the corresponding ranks and relative inefficiencies. With an absolute difference of just 3.55% or less between the relative inefficiencies, the models $M4$, $M5$ and $M6$ have the lowest relative inefficiencies. This highlights the strong identifiability of the Shapley influence measure within the FH models and the IC KIC_{b2}.

Table 11 shows the regression coefficients and the corresponding p-values for the auxiliary variables of the models M_{S1}, $M6$, $M5$ and $M4$, respectively. Model M_{S1} has all significant auxiliary variables, but the signs of the coefficients of $edu1$ (negative) and $age4$ (positive) are not coherent with the socio-economic theory. From a socio-economic perspective, a higher proportion of people with low levels of education is typically associated with a greater risk of poverty, i.e a positive association with poverty levels is expected for $edu1$. Similarly, the elderly population are comparatively well protected by consolidated career trajectories, higher earning capacity, and accumulated assets, which jointly reduce their poverty risk. Therefore, a negative sign would be expected for $age4$ or $age5$. On the other hand, while models $M5$ and $M6$ exhibit the lowest relative inefficiency (1.56% and 2.35%, respectively), both contain a non-significant variable (p-value > 0.49). As a result, we do not consider them to be

Table 10 Classification of the models evaluated by the Shapley Algorithm 1 with the KIC_{b2} IC and their relative inefficiency in %

Model	Auxiliary variables	Rank	$RI(M)$
$M1$	$lab2$	724	17.81
$M2$	$lab2, edu3$	729	17.87
$M3$	$lab2, edu3, edu2$	169	4.98
$M4$	$lab2, edu3, edu2, age3$	72	3.55
$M5$	$lab2, edu3, edu2, age3, age4$	10	1.56
$M6$	$lab2, edu3, edu2, age3, age4, age2$	21	2.35
$M7$	$lab2, edu3, edu2, age3, age4, age2, nac1$	111	4.12
$M8$	$lab2, edu3, edu2, age3, age4, age2, nac1, lab1$	151	4.79
$M9$	$lab2, edu3, edu2, age3, age4, age2, nac1, lab1, age5$	198	5.47
$M10$	$lab2, edu3, edu2, age3, age4, age2, nac1, lab1, age5, edu1$	166	4.94

Table 11 Regression coefficients and p-values for the auxiliary variables of models M_{S1}, $M6$, $M5$ and $M4$

Model	Coeff./p-value	$lab2$	$edu3$	$edu2$	$age3$	$age4$	$age2$	$edu1$
M_{S1}	$\hat{\beta}_i$	1.43	−0.60		0.52	1.04		−0.28
	p-value	0.000	0.000		0.000	0.000		0.035
$M6$	$\hat{\beta}_i$	1.41	−0.56	0.06	0.58	0.73	−0.11	
	p-value	0.000	0.000	0.518	0.005	0.000	0.784	
$M5$	$\hat{\beta}_i$	1.39	−0.55	0.07	0.54	0.73		
	p-value	0.000	0.000	0.494	0.000	0.000		
$M4$	$\hat{\beta}_i$	1.71	−0.33	0.28	0.51			
	p-value	0.000	0.000	0.000	0.000			

suitable models. As previously pointed out, this is not necessarily a contradiction, as the Shapley value evaluates the average marginal contribution across all possible submodels, which may lead to including some non-significant variables.

Therefore, among the models presented in Table 11, Model $M4$ is the most appropriate, since it achieves good performance in terms of relative inefficiency (3.55%), all its variables are statistically significant, and its coefficients are coherent with the socio-economic theory.

5.1 Model validation

We validate the selected model by combining residual diagnostics with an assessment of the shrinkage factors of the $M4$ model. First, we examine standardized area-level residuals to check the normality assumption. Second, we study the magnitude and dispersion of the model's weights, i.e., the shrinkage factors or associated regression weights in $M4$. For this model, we have $S = \{lab2, edu3, edu2, age3\}$. To simplify

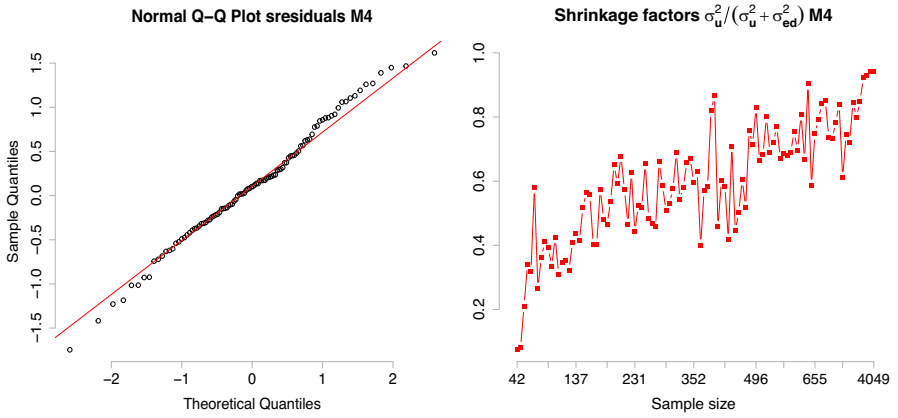

Fig. 1 Normal Q-Q plot of the sresiduals under the $M4$ model (left) and scatter plot of the shrinkage factors of $M4$ versus the sample sizes (right)

notation in what follows, we drop the subscript S on all estimators and write $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_u^2$, $\hat{u}_d$ and $\hat{\mu}_d$ for the corresponding estimators/predictors under $M4$.

The raw residuals and standardized area-level residuals (sresiduals) of the basic FH model are

$$\hat{e}_d = y_d - \hat{\mu}_d, \;\; \hat{e}_d^{st} = \hat{e}_d/\sigma_{ed}, \quad d = 1, \ldots, D.$$

Additionally, following Equation (2.6), the shrinkage factors are given by

$$\gamma_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_{ed}^2}, \quad d = 1, \ldots, D.$$

Note that large values in σ_{ed}^2 (small n_d) imply values close to zero. This is to say, more weight on the synthetic estimator $\boldsymbol{x}_d\hat{\boldsymbol{\beta}}$. Conversely, small values in σ_{ed}^2 (large n_d) imply values close to one; this is to say, more weight on the direct estimator y_d.

Figure 1 (left) shows the normal Q-Q plot of the sresiduals under the $M4$ model. Since the residuals lie close to the reference line in the Q–Q plot, we find no evidence against normality. On the other hand, Figure 1 (right) presents the scatter plot of the shrinkage factors of $M4$ versus the sample sizes. We can observe that as we increase the sample size, the shrinkage factors converge to one. This graphic confirms that the FH $M4$ model assigns higher weights to the direct estimator y_d when the sample size is large and higher weights to the synthetic estimator $\boldsymbol{x}_d\hat{\boldsymbol{\beta}}$ when the sample size is very small.

5.2 Prediction and error measures

This section presents maps of the predicted poverty proportions by sex across Spanish provinces, using model $M4$, to enhance spatial understanding of poverty distribution in

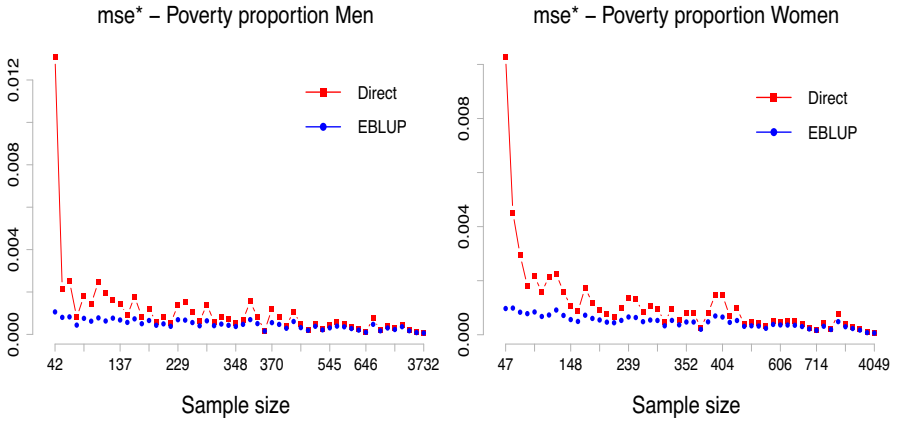

Fig. 2 MSE bootstrap estimates of the direct estimators and the EBLUPs of the province poverty proportions for men (left) and women (right), sorted by sample size

Spain. As measures of prediction error, the mean square errors (MSEs) are computed by the parametric bootstrap approach described in Appendix D with $B = 500$ iterations.

Figure 2 plots the MSE bootstrap estimates of the direct estimators, $mse^*(y_d)$, and the EBLUPs, $mse^*(\hat{\mu}_d)$, of the province poverty proportions for men (left) and women (right), sorted by sample size. The largest differences between the MSE bootstrap estimates of the direct estimators and the EBLUPs are found in domains with very small sample sizes - Soria ($n_d^M = 42$, $n_d^W = 47$); Zamora ($n_d^M = 54$, $n_d^W = 56$); Ávila ($n_d^M = 95$, $n_d^W = 101$); and Teruel ($n_d^M = 99$, $n_d^W = 109$) - where n_d^M and n_d^W denote the sample sizes in province d for men and women, respectively. A small difference between the MSE bootstrap estimates of the direct estimators and the EBLUPs can occur even with a small sample when the direct estimate is already close to the regression fit (e.g., because the observed poverty proportion lies near the overall synthetic prediction, or its design-based variance is unusually small). This is observed in the province of Teruel for men (Figure 2 left, 4th observation), which presents a shrinkage factor close to 0.6 (Figure 1 right, 6th observation). In fact, for this province, the design-based variance is comparatively small despite its limited sample size; up to 22 provinces have a smaller design-based variance. Therefore, EBLUP inherits a substantial portion of the sample variability of the direct estimator, leading to estimates close to the MSE.

Figure 3 plots the EBLUPs of poverty proportions for men (left) and women (right). Provinces in the southern and central zones of the peninsula exhibit higher poverty proportions than those in the northern zone. Moreover, although the spatial distribution of poverty proportions is broadly similar for men and women across Spanish provinces, the female population generally exhibits higher values. Figure 4 maps the estimated CVs of the EBLUPs under the FH model for men (left) and women (right). The predictions of the FH model are more accurate for women than men since the sample sizes for women are generally larger than for men. Only three provinces are above the threshold of 30% (Guipúzcoa, Burgos and Navarra for men) while the rest remain below it. This corresponds to suitable predictions under the model-based approach.

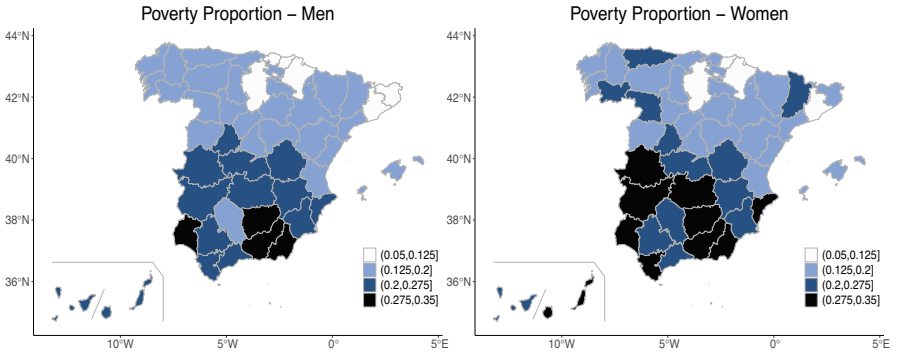

Fig. 3 EBLUPs of province poverty proportions of men (left) and women (right)

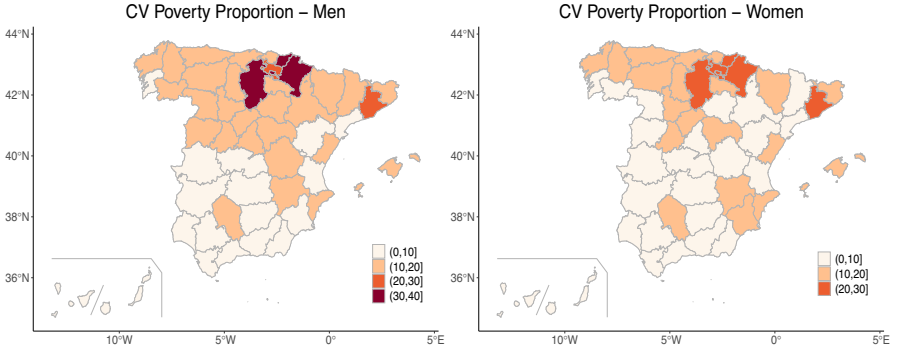

Fig. 4 CVs of EBLUPs of province poverty proportions for men (left) and women (right)

6 Conclusions

This paper introduces a new methodology for variable selection in the FH model, based on the Shapley value from cooperative game theory. The proposed influence measure allows for a principled ranking of auxiliary variables by quantifying their average marginal contribution to a chosen IC. Unlike the traditional models selection procedures based only on the ICs (AIC or KIC variants), the Shapley influence measure provides a more robust evaluation by averaging all possible auxiliary variable subsets.

A key component of our contribution is a Shapley-based model selection algorithm that sequentially constructs nested models according to variable rankings and evaluates them using ICs and type I error probability levels in regression parameter significance tests. The algorithm is evaluated through a simulation study, which also examines the ability of the Shapley influence measure to identify influential explanatory variables for the response variable. Unlike previous studies that assess model performance assuming knowledge of the true data-generating structure, our procedure builds models sequentially based on the data, without relying on prior information. Consequently, the selection proportions of the full generating model ($|T| = 5$) tend to be slightly lower, particularly in scenarios where some true variables have low

coefficients. However, the algorithm tends to recover informative underfitted models ($S \subset T, |S| = 3, 4$) most of the time, even under challenging conditions. The comparative analysis also shows that the KIC family of criteria, particularly KIC_{b2}, performs best when at least $B = 200$ bootstrap iterations are used. We recommend using moderate type I error probabilities ($\alpha \leq 0.1$) to mitigate the risk of selecting models that include non-significant variables, rather than imposing overly strict thresholds. Moreover, the results from the simulation experiment confirm that the Shapley influence measure successfully identifies the most influential variables, with stable performance across multiple ICs. The classification frequency of true explanatory variables remains high even under unfavourable scenarios.

The real data application to the SLCS 2022 further validates the utility of the proposed approach. The Shapley-based ranking provides meaningful insights into variable importance, supporting model choices that agree with the established theory and real-world evidence. Compared to classical selection techniques, the models identified by Shapley's algorithm tend to be more parsimonious while still capturing the key explanatory variables in the data.

In sum, our study underscores the advantages of combining traditional ICs with Shapley-based variable classification to improve model interpretability and accuracy in SAE applications. This hybrid approach allows for a robust and parsimonious selection of auxiliary variables. An interesting direction for future research is the extension of the proposed methodology to multivariate small area estimation models.

Appendices

A Synthetic Data Setup

This appendix provides the data-generating process of the synthetic data used in Examples 1, 2 and 3. The process mimics the simulation experiments from Section 4.

The true model parameters are $\boldsymbol{\beta} = (1, 1, 0.25)'$. The auxiliary variables are simulated as $x_{dj} \sim U(0, 3), d = 1, \ldots, D, j = 1, \ldots, |P|$, where U denotes uniform distribution, $D = 50$ and $|P| = 3$. The variance of the random effects is $\sigma_u^2 = 1$. The sampling error variances are $\sigma_{ed}^2 = \lambda_0 + \frac{(\lambda_1 - \lambda_0)(d-1)}{D-1}$, with $\lambda_0 = 0.8$ and $\lambda_1 = 1.2$ for each $d = 1, \ldots, D$. We construct the auxiliary variable vectors $\boldsymbol{x}_d = (x_{d1}, x_{d2}, x_{d3})$ and the target vectors as $y_d = \boldsymbol{x}_d\boldsymbol{\beta} + u_d + e_d$, where $u_d \sim N(0, \sigma_u^2)$ and $e_d \sim N(0, \sigma_{ed}^2)$.

B Proof of Theorem 1

This appendix provides the proof of Theorem 1, given in Section 3. This proof establishes that the influence measure based on the Shapley value is the unique measure satisfying the properties of (g, P)-efficiency and balanced influence.

Proof *Existence.* First, we prove that $I^{\Phi,g}$ satisfies (g, P)-efficiency. The Shapley value, see Shapley (1953), satisfies the efficiency property, that is, for any cooperative

game (P, v), it holds that $\sum_{j\in P} \Phi_j(v) = v(P)$. Applying this result, we have

$$\sum_{j\in P} I_j^{\Phi,g}(M_P) = \sum_{j\in S} \Phi_j(v^{M_P}) = v^{M_P}(P) = g(P).$$

Second, let us see that $I^{\Phi,g}$ satisfies balanced influence. The Shapley value satisfies the property of balanced contributions, see Myerson (1980). This property says that for any cooperative game (P, v), it holds that

$$\Phi_i(v) - \Phi_i(v_{-j}) = \Phi_j(v) - \Phi_j(v_{-i}),$$

for any $i, j \in P$, where v_{-j} is the cooperative game restricted to set of players $P\backslash\{j\}$. In our setting, we have

$$\begin{aligned} I_i^{\Phi,g}(M_P) - I_i^{\Phi,g}\left(M_{P\backslash\{j\}}\right) &= \Phi_i(v^{M_S}) - \Phi_i(v^{M_{S\backslash\{j\}}}) \\ I_j^{\Phi,g}(M_S) - I_j^{\Phi,g}\left(M_{S\backslash\{j\}}\right) &= \Phi_j(v^{M_S}) - \Phi_j(v^{M_{S\backslash\{i\}}}). \end{aligned}$$

For every $j \in S$; $v^{M_{S\backslash\{j\}}}$ is a cooperative game with set of players $S\backslash\{j\}$. For all $R \subseteq S\backslash\{j\}$, we have

$$v^{M_{S\backslash\{j\}}}(R) = g(R) = v^{M_S}(R).$$

By other way, for all $R \subseteq S\backslash\{j\}$ it holds that $v^{M_S}(R) = v_{-j}^{M_S}(R)$. Then, $v^{M_{S\backslash\{j\}}} = v_{-j}^{M_S}$ and

$$I_i^{\Phi,g}(M_S) - I_i^{\Phi,g}\left(M_{S\backslash\{j\}}\right) = \Phi_i(v^{M_S}) - \Phi_i(v^{M_{S\backslash\{j\}}}) = \Phi_i(v^{M_S}) - \Phi_i(v_{-j}^{M_S}) \tag{B.1}$$

$$I_j^{\Phi,g}(M_S) - I_j^{\Phi,g}\left(M_{S\backslash\{j\}}\right) = \Phi_j(v^{M_S}) - \Phi_j(v^{M_{S\backslash\{i\}}}) = \Phi_j(v^{M_S}) - \Phi_j(v_{-i}^{M_S}). \tag{B.2}$$

Since Myerson (1980) proved that the Shapley value satisfies balanced contribution, equations (B.1) and (B.2) are the same.

Uniqueness. Suppose that now there is exist another influence measure $I(M_P) \neq I^{\Phi,g}(M_P)$ that satisfies both properties. If $|P| = 1$ by (g, P)-efficiency,

$$I(M_P) = g(P) = I^{\Phi,g}(M_P).$$

Let be $|P| \geq 2$. Assume that M_P is minimal in the sense that $I\left(M_{P\backslash\{j\}}\right) = I^{\Phi,g}\left(M_{P\backslash\{j\}}\right)$ for any $j \in P$. Consider $i, j \in P$ with $i \neq j$. Since $I(M_P)$ and $I^{\Phi,g}(M_P)$ satisfies balanced influence, then

$$\begin{aligned} I_i(M_P) - I_i\left(M_{P\backslash\{j\}}\right) &= I_j(M_P) - I_j\left(M_{P\backslash\{i\}}\right) \\ I_i^{\Phi,g}(M_P) - I_i^{\Phi,g}\left(M_{P\backslash\{j\}}\right) &= I_j^{\Phi,g}(M_P) - I_j^{\Phi,g}\left(M_{P\backslash\{j\}}\right) \end{aligned}$$

and by the minimality of M_P

$$I_i(M_P) - I_j(M_P) = I_i^{\Phi,g}(M_P) - I_j^{\Phi,g}(M_P)$$

or

$$I_i(M_P) - I_i^{\Phi,g}(M_P) = I_j(M_P) - I_j^{\Phi,g}(M_P).$$

Since the above equation does not depend on i, it is equivalent to $I_i(M_P) - I_i^{\Phi,g}(M_P) = c \in \mathbb{R}$ for any $i \in P$. Therefore,

$$\sum_{i \in P} \left(I_i(M_P) - I_i^{\Phi,g}(M_P) \right) = |P|c$$

and by (g, P)-efficiency, we have that $c = 0$. This implies that $I(M_P) = I^{\Phi,g}(M_P)$, and the proof is concluded. □

C Propositions

Propositions 1 and 2 are given and proved in Castro et al. (2009).

Proposition 1 *The estimator $\hat{I}_i^{\Phi,g}(M_P)$ is unbiased and its variance is $var[\hat{I}_i^{\Phi,g}] = \frac{\sigma^2}{m}$, where $\sigma^2 = \frac{1}{|P|!}\sum_{\sigma \in \Pi(P)} \left(x_i(\sigma) - I_i^{\Phi,g}(M_P) \right)$, $x_i(\sigma) = g(M_{\{\sigma(1),\ldots,\sigma(k)\} \cup \{i\}}) - g(M_{\{\sigma(1),\ldots,\sigma(k)\}})$ and $\sigma(k+1) = i$.*

Proposition 2 *The estimator $\hat{I}^{\Phi,g}(M_P)$ is consistent in probability.*

D Bootstrap estimation of the MSE

Following González-Manteiga et al. (2008; 2010), we propose a parametric bootstrap-based procedure to estimate the MSEs of the EBLUPs $\hat{\mu}_{Sd}$. The steps are the following

1. Fit the FH model M_S to $(\boldsymbol{x}_{Sd}, y_d)$, $d = 1, \ldots, D$. Calculate the REML estimators $\hat{\sigma}_{Su}^2$ and $\hat{\boldsymbol{\beta}}_S$.
2. Repeat B times $(b = 1, \ldots, B)$:
   (a) For $d = 1, \ldots, D$, generate $u_{Sd}^{*(b)} \sim N(0, \hat{\sigma}_{Su}^2)$ and $e_d^{*(b)} \sim N(0, \sigma_{ed}^2)$.
   (b) Calculate the bootstrap direct estimates

$$y_{Sd}^{*(b)} = \boldsymbol{x}_{Sd}\hat{\boldsymbol{\beta}}_S + u_{Sd}^{*(b)} + e_d^{*(b)}. \quad d = 1, \ldots, D. \tag{D.1}$$

   (c) Fit the FH model M_S to the bootstrap data $(\boldsymbol{x}_{Sd}, y_{Sd}^{*(b)})$, $d = 1, \ldots, D$, calculate the estimators $\hat{\sigma}_{Su}^{2*(b)}$, $\hat{\boldsymbol{\beta}}_S^{*(b)}$ of the model parameters. Calculate the bootstrap target parameter $\mu_{Sd}^{*(b)} = \boldsymbol{x}_{Sd}\hat{\boldsymbol{\beta}}_S + u_{Sd}^{*(b)}$ and the EBLUP $\hat{\mu}_{Sd}^{*(b)}$, $d = 1, \ldots, D$.

Table 12 Proportion of iterations of i-th variable in the top five variables classification using the Shapley value estimator in scenarios $\boldsymbol{\beta}_{T0}$ and $\boldsymbol{\beta}_{T1}$, $\boldsymbol{\beta}_{T2}$ and $\boldsymbol{\beta}_{T3}$

	$\boldsymbol{\beta}_{T0}$							
i	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
1	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.95
2	0.97	0.97	0.96	0.97	0.96	0.96	0.96	0.96
3	0.96	0.96	0.97	0.96	0.95	0.97	0.96	0.96
4	0.96	0.96	0.96	0.96	0.97	0.96	0.97	0.96
5	0.96	0.96	0.96	0.96	0.95	0.96	0.96	0.96
6	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04
7	0.04	0.04	0.03	0.04	0.04	0.04	0.03	0.03
8	0.03	0.04	0.05	0.04	0.04	0.04	0.04	0.04
9	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05
10	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.04
	$\boldsymbol{\beta}_{T1}$							
i	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
1	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
2	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99
3	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
4	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
5	0.33	0.33	0.33	0.35	0.33	0.35	0.34	0.34
6	0.14	0.15	0.14	0.14	0.14	0.14	0.14	0.15
7	0.14	0.14	0.14	0.14	0.13	0.13	0.15	0.14
8	0.15	0.12	0.15	0.13	0.14	0.15	0.13	0.13
9	0.15	0.15	0.14	0.14	0.16	0.14	0.15	0.14
10	0.13	0.16	0.14	0.13	0.15	0.13	0.14	0.15
	$\boldsymbol{\beta}_{T2}$							
i	AIC	KIC	$AICc$	AIC_{b1}	AIC_{b2}	KIC_c	KIC_{b1}	KIC_{b2}
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	0.45	0.48	0.47	0.47	0.45	0.46	0.46	0.46
5	0.46	0.46	0.46	0.43	0.45	0.45	0.44	0.46
6	0.23	0.23	0.21	0.23	0.22	0.21	0.22	0.24
7	0.22	0.19	0.21	0.23	0.23	0.23	0.20	0.21
8	0.22	0.22	0.23	0.21	0.22	0.21	0.23	0.22
9	0.21	0.21	0.22	0.23	0.23	0.23	0.24	0.21
10	0.22	0.21	0.21	0.20	0.21	0.22	0.23	0.22

Table 12 continued

	$\boldsymbol{\beta}_{T0}$							
i	AIC	KIC	AICc	AIC_{b1}	AIC_{b2}	KICc	KIC_{b1}	KIC_{b2}
	$\boldsymbol{\beta}_{T3}$							
i	AIC	KIC	$AICc$	AIC_{b1}	AIC_{b2}	KIC_c	KIC_{b1}	KIC_{b2}
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0.54	0.55	0.53	0.55	0.55	0.54	0.54	0.56
4	0.56	0.55	0.56	0.54	0.53	0.57	0.54	0.53
5	0.55	0.54	0.56	0.55	0.56	0.55	0.56	0.54
6	0.27	0.26	0.26	0.26	0.25	0.26	0.27	0.27
7	0.26	0.28	0.28	0.29	0.28	0.27	0.28	0.28
8	0.28	0.28	0.27	0.27	0.27	0.27	0.27	0.28
9	0.27	0.26	0.26	0.27	0.28	0.26	0.27	0.26
10	0.28	0.28	0.28	0.27	0.27	0.28	0.27	0.27

3. Output: For $d = 1, \ldots, D$, calculate

$$mse^*(\hat{\mu}_{Sd}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mu}_{Sd}^{*(b)} - \mu_{Sd}^{*(b)}\right)^2, \quad mse^*(y_d) = \frac{1}{B}\sum_{b=1}^{B}\left(y_{Sd}^{*(b)} - \mu_{Sd}^{*(b)}\right)^2.$$

E Simulation 3. Shapley value estimator

Table 12 presents the results of Simulation 1 for the Shapley value estimator. The estimator is calculated using the Algorithm 2 with $m = 300$ explanatory variable subsets, which represent approximately 30% of the total possible configurations. Analogously to the Shapley value (Tables 4 and 5), the estimator demonstrates an acceptable detection rate of the true data-generating variables across all scenarios. This rate tends to decrease progressively in scenarios where the explanatory variables exhibit lower levels of importance.

Acknowledgements The authors thank the Associate Editor and the two anonymous reviewers for their constructive remarks, which significantly improved the quality of this article.

References

Algaba E, Fragnelli V, Sánchez-Soriano J (2019) Handbook of the Shapley value. CRC Press, Boca Raton

Betti G, Lemmi A (2013) Poverty and social exclusion in 3d: Multidimensional, longitudinal and small area estimation. In: Poverty and Social Exclusion, pp 301–315

Boubeta M, Lombardía MJ, Morales D (2016) Empirical best prediction under area-level poisson mixed models. TEST 25:548–569

Boubeta M, Lombardía MJ, Morales D (2017) Poisson mixed models for studying the poverty in small areas. Comput Stat Data Anal 107:32–47

Burgard JP, Esteban MD, Morales D, Pérez A (2020) A Fay-Herriot model when auxiliary variables are measured with error. TEST 29:166–195

Castro J, Gómez D, Tejada J (2009) Polynomial calculation of the Shapley value based on sampling. Comput Oper Res 36(5):1726–1730

Chandra H, Salvati N, Chambers R (2017) Small area prediction of counts under a non-stationary spatial model. Spatial Stat 20:30–56

Esteban MD, Morales D, Pérez A, Santamaría L (2012) Small area estimation of poverty proportions under area-level time models. Comput Stat Data Anal 56(10):2840–2855

González-Manteiga W, Lombarda M, Molina I, Morales D, Santamaría L (2008) Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. Comput Stat Data Anal 52(12):5242–5252

González-Manteiga W, Lombarda M, Molina I, Morales D, Santamaría L (2010) Small area estimation under Fay-Herriot models with non-parametric estimation of heteroscedasticity. Stat Model 10(2):215–239

Guadarrama M, Morales D, Molina I (2021) Time stable empirical best predictors under a unit-level model. Comput Stat Data Anal 160:107226

Hurvich CM, Tsai C-L (1989) Regression and time series model selection in small samples. Biometrika 76(2):297–307

Israeli O (2007) A Shapley-based decomposition of the r-square of a linear regression. J Econ Inequal 5:199–212

Kubokawa T (2011) Conditional and unconditional methods for selecting variables in linear mixed models. J Multivar Anal 102(3):641–660

Lahiri P, Suntornchost J (2015) Variable selection for linear mixed models with applications in small area estimation. Sankhya B 77:312–320

Lipovetsky S, Conklin M (2001) Analysis of regression in game theory approach. Appl Stoch Model Bus Ind 17(4):319–330

Lombardía M.J, López-Vizcaıno E, Rueda C (2018). Selection of small area estimators

Lombardía MJ, López-Vizcaíno E, Rueda C (2021) Selection model for domains across time: application to labour force survey by economic activities. TEST 30(1):228–254

López-Vizcaíno E, Lombardía MJ, Morales D (2013) Multinomial-based small area estimation of labour force indicators. Stat Model 13(2):153–178

López-Vizcaíno E, Lombardía MJ, Morales D (2015) Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. J R Stat Soc Ser A Stat Soc 178(3):535–565

Marchetti S, Secondi L (2016) Estimates of household consumption expenditure at provincial level in Italy by using small area estimation methods

Marchetti S, Tzavidis N, Pratesi M (2012) Non-parametric bootstrap mean squared error estimation for m-quantile estimators of small area averages, quantiles and poverty indicators. Comput Stat Data Anal 56(10):2889–2902

Marhuenda Y, Molina I, Morales D (2013) Small area estimation with spatio-temporal Fay-Herriot models. Comput Stat Data Anal 58:308–325

Marhuenda Y, Molina I, Morales D, Rao J (2017) Poverty mapping in small areas under a twofold nested error regression model. J R Stat Soc Ser A Stat Soc 180(4):1111–1136

Marhuenda Y, Morales D, Pardo MC (2014) Information criteria for Fay-Herriot model selection. Comput Stat Data Anal 70:268–280

Mishra SK (2016) Shapley value regression and the resolution of multicollinearity. Available at SSRN 2797224

Molina I, Marhuenda Y (2015) SAE: an R package for small area estimation. R J 7(1):81–98

Molina I, Rao JN (2010) Small area estimation of poverty indicators. Can J Stat 38(3):369–385

Morales D, Esteban MD, Pérez A, Hobza T (2021) A course on small area estimation and mixed models, Springer
Morales D, Pagliarella MC, Salvatore R (2015) Small area estimation of poverty indicators under partitioned area-level time models. SORT 39(1):19–34
Myerson RB (1980) Conference structures and fair allocation rules. Internat J Game Theory 9:169–182
Pratesi M (2016) Analysis of poverty data by small area estimation
Rao JN, Molina I (2015) Small area estimation. Wiley, Hoboken
Rozemberczki B, Watson L, Bayer P, Yang H-T, Kiss O, Nilsson S, Sarkar R (2022) The shapley value in machine learning. arXiv preprint arXiv:2202.05594
Shang J, Cavanaugh JE (2008) Bootstrap variants of the akaike information criterion for mixed model selection. Comput Stat Data Anal 52(4):2004–2021
Shapley LS (1953) A value for n-person games. Contribution to the Theory of Games, 2
Tzavidis N, Ranalli MG, Salvati N, Dreassi E, Chambers R (2015) Robust small area prediction for counts. Stat Methods Med Res 24(3):373–395
Tzavidis N, Salvati N, Pratesi M, Chambers R (2008) M-quantile models with application to poverty mapping. Stat Methods Appl 17:393–411
Ybarra L, Lohr S (2008) Small area estimation when auxiliary information is measured with error. Biometrika 95:919–931