*XVII Congreso Galego de Estatística e Investigación de Operacións*
*Ourense, 23 - 25 de outubro de 2025*

# INFERENCE AND DIAGNOSIS IN M-QUANTILE MODELS WITH APPLICATIONS TO SMALL AREA ESTIMATION

María Bugallo[1] and Domingo Morales[1]

[1]Center of Operations Research, Miguel Hernández University of Elche, Spain.

## ABSTRACT

This research develops methodological contributions to M-quantile (MQ) regression models in the context of small area estimation, with particular emphasis on inference and diagnostics. The authors establish the consistency of the area-specific MQ coefficients, which play a role analogous to random effects in mixed models by capturing between-area variability. They also address the analysis of the residuals, providing an approximation to their distribution and drawing connections with mixed models. Furthermore, they estimate the distribution of optimal robustness parameters for bias correction via bootstrap, which enables the construction of a statistical test for the identification of atypical areas. Through simulation studies, the new methodology is shown to be effective in detecting outliers and assessing model fit. An application to Spanish income data demonstrates the practical utility of the contributions.

**Keywords**: Consistent estimator; residual analysis; optimal bias correction; outlier detection; robustness parameter; bootstrap inference.

## 1. INTRODUCTION

National statistical offices design surveys to achieve cost-effective and accurate estimates at a given level of aggregation. However, producing disaggregated statistics is essential for informed decision-making (Rao and Molina, 2015), which often requires additional information or more sophisticated prediction tools. Small area estimation (SAE) addresses these challenges when domain-level sample sizes are too small for reliable direct estimation (Morales et al., 2021). SAE typically relies on linear and generalized linear mixed models (LMM/GLMM), using either unit-level or area-level data, to borrow strength across domains via auxiliary information and some structural assumptions. Unit-level models –our focus in this research– commonly use empirical best linear unbiased predictors (EBLUP) or empirical best predictors (EBP) derived under restrictive parametric assumptions, where normally distributed random effects account for between-area variability. Numerous recent contributions to unit-level mixed models have been made, many of which are reviewed in Bugallo et al. (2024).

In contrast, M-quantile (MQ) regression (Breckling and Chambers, 1988; Chambers and Tzavidis, 2006) offers a robust alternative, avoiding strong distributional assumptions while delivering reliable predictions. MQ-based SAE has been successfully applied to predict indicators such as poverty rates and economic variables (e.g., Salvati et al., 2012; Marchetti et al., 2018). Despite their advantages, robust methods can lead to biased predictions, particularly when the robustness mechanism is not optimally calibrated. Bias-corrected MQ predictors incorporate a second influence function governed by a robustness parameter, whose optimal, data-driven selection was recently addressed by Bugallo et al. (2025).

In our ongoing unpublished manuscript (Bugallo and Morales, 2025), we build on the work of Bugallo et al. (2025) by examining the distribution of optimal robustness parameters and introducing a new test for detecting atypical areas. In addition, we propose a novel approximation to the distribution of the MQ residuals and establish the consistency of the area-specific MQ coefficients, which play a role analogous to random effects in mixed models. Our approach enables formal outlier detection in MQ models, which has been relatively unexplored. Simulations and an application to income data from the 2022 Spanish Living Conditions Survey (SLCS) illustrate the effectiveness of the proposed methodology.

The remainder of the document is structured as follows. Section 2 reviews the MQ models for small area linear prediction. Section 3 establishes consistency results. Section 4 deals with the residual analysis and the bootstrap inference. Section 5 introduces a new test for area-level outlier detection. Sections 6 and 7 present simulation and empirical results related to differences in average income between provinces in Andalucía (Spain), using data from the 2022 SLCS. Section 8 concludes with some final remarks.

## 2. M-QUANTILE MODELS FOR SMALL AREA LINEAR PREDICTION

Let $U$ be a finite population of size $N$, hierarchically partitioned into $D$ non-overlapping small areas $U_d$, each of size $N_d$, for $d = 1, \ldots, D$. The samples drawn from the population and from each small area are denoted by $s$ and $s_d$, with respective sample sizes $n$ and $n_d$. It is assumed that the indexes in $U_d$ are ordered such that the first $n_d$ units of each domain correspond to the sample units in $s_d$, and the remaining $N_d - n_d$ units correspond to the non-sampled units in $r_d = U_d \setminus s_d$. The vector of $p \geq 1$ unit-level auxiliary variables $\boldsymbol{x}_{dj}$ is assumed to be known for all population units, whereas the target variables $y_{dj}$ are absolutely continuous and observed only for the units included in the sample subsets.

For $0 < q < 1$, the two-level MQ models are specified as

$$y_{dj} = \boldsymbol{x}'_{dj}\boldsymbol{\beta}_\psi(q) + e_{\psi,dj}(q), \quad d = 1, \ldots, D,\ j = 1, \ldots, N_d, \tag{1}$$

where the MQ function (Breckling and Chambers, 1988) of order $q$ for $y_{dj}$, given $\boldsymbol{x}_{dj}$, is

$$Q_q(y_{dj}; \sigma_q, \psi \mid \boldsymbol{x}_{dj}) = \boldsymbol{x}'_{dj}\boldsymbol{\beta}_\psi(q) \tag{2}$$

and $\boldsymbol{\beta}_\psi(q)$ is the vector of regression coefficients that depend on the quantile level $q$. The model errors are $e_{\psi,dj}(q) = y_{dj} - \boldsymbol{x}'_{dj}\boldsymbol{\beta}_\psi(q)$ and assumed to be independent. Conditioned to $\boldsymbol{x}_{dj}$, they satisfy that $Q_q(e_{\psi,dj}(q); \sigma_q, \psi|\boldsymbol{x}_{dj}) = 0$ and the homoscedasticity assumption $\sigma_q = \mathrm{var}^{1/2}(e_{\psi,dj}(q))$ is required.

The iterative re-weighted least squares (IRLS) algorithm is used to fit the two-level MQ models (Bianchi and Salvati, 2015) and guarantees convergence to a unique solution for a continuous monotone influence function $\psi$. At the output, we obtain not only an estimate of $\boldsymbol{\beta}_\psi(q)$ and $\sigma_q$, but also a diagonal matrix with the final weights $W_\psi(q) = \underset{1\leq d\leq D}{\mathrm{diag}}\big(\underset{1\leq j\leq n_d}{\mathrm{diag}}\ (w_{\psi dj}(q))\big)$, where:

$$\hat{\boldsymbol{\beta}}_\psi(q) = \big(X'W_\psi(q)X\big)^{-1}X'W_\psi(q)\boldsymbol{y}, \tag{3}$$

$\boldsymbol{y} = \underset{1\leq d\leq D}{\mathrm{col}}\big(\underset{1\leq j\leq n_d}{\mathrm{col}}\ (y_{dj})\big)$, $X = \underset{1\leq d\leq D}{\mathrm{col}}\big(\underset{1\leq j\leq n_d}{\mathrm{col}}\ (\boldsymbol{x}'_{dj})\big)$ and $\hat{\sigma}_q = \widehat{\mathrm{var}}^{1/2}(\hat{e}_{\psi,dj}(q)) = \mathrm{mad}_{\psi,n}(\hat{e}_{\psi,dj}(q))/0.6745$ is the median absolute deviation (MAD). For this study, we consider the Huber function

$$\psi(u) = u\,\mathbb{I}_{(-c_\psi,\, c_\psi)}(u) + c_\psi\,\mathrm{sgn}(u)\,\mathbb{I}_{\{|u|\geq c_\psi\}}, \quad u \in \mathbb{R}, \quad c_\psi > 0. \tag{4}$$

A commonly adopted value for the tuning constant is $c_\psi = 1.345$, which ensures approximately 95% asymptotic efficiency under the assumption of normally distributed model errors.

For the applications to SAE, the unit-level MQ coefficient of unit $j$ of area $d$ (Chambers and Tzavidis, 2006; Dawber and Chambers, 2019) is

$$q_{dj} = \underset{0<q<1}{\mathrm{solution}}\Big\{Q_q(y_{dj}; \sigma_q, \psi|\boldsymbol{x}_{dj}) = y_{dj}\Big\}, \quad d = 1, \ldots, D,\ j = 1, \ldots, N_d. \tag{5}$$

The unit-level MQ coefficient $q_{dj}$ is the "most likely" quantile probability of unit $j$ of area $d$, so $y_{dj} = \boldsymbol{x}'_{dj}\boldsymbol{\beta}_\psi(q_{dj})$, and it can be predicted in the sampled units $j \in s_d$:

$$\hat{q}_{dj} = \underset{0<q<1}{\mathrm{solution}}\Big\{\hat{Q}_q(y_{dj}; \hat{\sigma}_q, \psi|\boldsymbol{x}_{dj}) = y_{dj}\Big\}, \quad \hat{Q}_q(y_{dj}; \hat{\sigma}_q, \psi|\boldsymbol{x}_{dj}) = \boldsymbol{x}'_{dj}\hat{\boldsymbol{\beta}}_\psi(q). \tag{6}$$

The unit-level MQ coefficients are determined at the population level. Therefore, if a hierarchical structure explains part of the variability of the population, units within areas defined by that hierarchy are expected to have similar unit-level MQ coefficients. The population means and the predicted small area sample means of the unit-level MQ coefficients –commonly referred to as area-specific MQ coefficients– are

$$\theta_d = \frac{1}{N_d}\sum_{j=1}^{N_d} q_{dj} \quad \text{and} \quad \hat{\theta}_d = \frac{1}{n_d}\sum_{j=1}^{n_d} \hat{q}_{dj}. \tag{7}$$

The two-level MQ model with $q = \hat{\theta}_d$ is expected to provide the most accurate predictions in area $d$. They can be used to predict various area-specific quantities, with population means $\overline{Y}_d = \frac{1}{N_d}\sum_{j=1}^{N_d} y_{dj}$ being a primary example. Based on a Taylor series expansion, a plug-in predictor of $\overline{Y}_d$ is

$$\hat{\overline{Y}}_d^{mq} = \frac{1}{N_d}\Big\{\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d} \boldsymbol{x}'_{dj}\hat{\boldsymbol{\beta}}_\psi\big(\hat{\theta}_d\big)\Big\}. \tag{8}$$

The bias of $\hat{\overline{Y}}_d^{mq}$ is $B\big(\hat{\overline{Y}}_d^{mq}\big) = E\big[\hat{\overline{Y}}_d^{mq} - \overline{Y}_d\big]$ and a robust estimator is (Chambers et al., 2014)

$$\hat{B}_\phi\big(\hat{\overline{Y}}_d^{mq}\big) = -\Big(1 - \frac{n_d}{N_d}\Big)\frac{\hat{\sigma}_{\hat{\theta}_d}}{n_d}\sum_{j\in s_d}\phi\big(\hat{u}_{\psi,dj}\big), \tag{9}$$

where $\phi$ is an influence function with area-specific robustness parameter $c_{\phi,d} \geq 0$. Here, the residuals and standardized residuals for $q = \hat{\theta}_d$ are defined as $\hat{e}_{\psi,dj} \triangleq \hat{e}_{\psi,dj}(\hat{\theta}_d) = y_{dj} - \boldsymbol{x}'_{dj}\hat{\boldsymbol{\beta}}_\psi\big(\hat{\theta}_d\big)$ and $\hat{u}_{\psi,dj} = \hat{\sigma}_{\hat{\theta}_d}^{-1}\hat{e}_{\psi,dj}$, respectively. In practice, as the influence function $\phi$ for bias correction we use the Huber function given in (4). Accordingly, a robust bias-corrected MQ (BMQ) predictor of $\overline{Y}_d$ is

$$\hat{\overline{Y}}_d^{bmq} \;=\; \hat{\overline{Y}}_d^{mq} - \hat{B}_\phi\big(\hat{\overline{Y}}_d^{mq}\big) = \hat{\overline{Y}}_d^{mq} + \frac{1}{n_d}\Big(1 - \frac{n_d}{N_d}\Big)\sum_{j\in s_d}\hat{\sigma}_{\hat{\theta}_d}\phi\big(\hat{u}_{\psi,dj}\big). \tag{10}$$

The final term on the right-hand side of (10) serves to mitigate the potential bias of the MQ predictor, as discussed by Chambers et al. (2014). By appropriately selecting the tuning parameter $c_{\phi,d}$, one can regulate the bias-variance trade-off inherent in the BMQ predictors and their corresponding mean squared error (MSE). Indeed, the robustness parameters $c_{\phi,d}$ play a crucial role in enhancing the BMQ predictor over the MQ predictor; however, the selection of their optimal values had remained an open issue until recently. A common, but subjective, choice is $\hat{c}_{\phi,d} = 3,\ d = 1,\ldots,D$.

To address this optimally, Bugallo et al. (2025) propose a data-driven procedure for selecting predictor-specific values of $c_{\phi,d}$ that minimize the estimated MSE of $\hat{\overline{Y}}_d^{bmq}$:

$$\hat{c}_{\phi,d} \triangleq \hat{c}_{\phi,d}(\hat{\theta}_d) = \underset{c_{\phi,d}\geq 0}{\operatorname{argmin}}\ \operatorname{mse}_d^{bmq}(c_{\phi,d}) = \underset{c_{\phi,d}\geq 0}{\operatorname{argmin}}\ A_d(c_{\phi,d}), \quad d = 1,\ldots,D, \tag{11}$$

where $\operatorname{mse}_d^{bmq}(c_\phi)$ denotes an estimate of $\operatorname{MSE}(\hat{\overline{Y}}_d^{bmq})$, such as those proposed by (Chambers et al., 2011, 2014), and $A_d(c_{\phi,d})$ its $c_{\phi,d}$ - dependent part. Using any of these methods, it holds that

$$A_d(c_{\phi,d}) = \Big(1 - \frac{n_d}{N_d}\Big)^2\Big(\frac{\hat{\sigma}_{\hat{\theta}_d}}{n_d}\Big)^2\Big(\sum_{j\in s_d}\phi^2\big(\hat{u}_{\psi,dj}\big) + \Big(\sum_{j\in s_d}\big(\phi\big(\hat{u}_{\psi,dj}\big) - \hat{u}_{\psi,dj}\big)\Big)^2\Big), \quad d = 1,\ldots,D. \tag{12}$$

Solutions to the minimization problem in (11) are referred to as optimal robustness parameters for bias correction in MQ linear prediction and are denoted by $\hat{c}_{\phi,d},\ d = 1,\ldots,D$. Their existence and uniqueness for each area were established by Bugallo et al. (2025) in the context of temporal MQ models, and this result naturally extends to classical MQ models.

## 3. CONSISTENCY OF AREA-SPECIFIC M-QUANTILE COEFFICIENTS

The area-specific MQ coefficients $\hat{\theta}_d$, for $d = 1,\ldots,D$, are treated as random variables and interpreted as pseudo-random effects that characterize area-level patterns. The asymptotic theory is developed under the following assumptions, which hold as $n \to \infty$ for $d = 1,\ldots,D$:

$$\text{(N1)}\ \exists 0 < a_d < 1 : \sum_{d=1}^{D} a_d = 1 \text{ and } \frac{n_d}{n} \to a_d; \qquad \text{(N2)}\ \exists 0 < f_d < 1 : f_d \to 1 \text{ and } \frac{n_d}{N_d} \to f_d. \tag{13}$$

**Theorem**. Let $d = 1,\ldots,D$. Under assumptions (A1)–(A8) in Bianchi and Salvati (2015) and (N1)–(N2) in (13), the consistency of $\hat{\theta}_d$ holds; that is, $|\hat{\theta}_d - \theta_d| = o_p(1)$ as $n \to \infty$.

The empirical consistency of $\hat{\theta}_d$ was assessed via model-based simulations in Bugallo et al. (2025).

## 4. RESIDUAL ANALYSIS AND BOOTSTRAP INFERENCE

We analyze the residual behavior in two-level MQ models to better understand their role in small area linear prediction. From (3), it holds that $\hat{e}_{\psi,dj} = \Big[(\mathbf{I}_{n\times n} - \mathbf{H}(\hat{\theta}_d))\boldsymbol{y}\Big]_{dj}$, where

$$\mathbf{H}(\hat{\theta}_d) = \mathbf{X}\left(\mathbf{X}'\mathbf{W}_\psi(\hat{\theta}_d)\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}_\psi(\hat{\theta}_d) \tag{14}$$

is the homologue of the projection matrix in mixed models and exhibits common properties. Although $\mathbf{H}(\hat{\theta}_d)$ is not symmetric, it is idempotent and its diagonal elements are interpreted as the leverage of the

$j$-th unit in area $d$ in the fitting process of the two-level MQ model for $q = \hat{\theta}_d$. The diagonal elements also allow for the detection of influential observations.

The distribution of the target variables in MQ models has recently been studied by Bianchi et al. (2018). Following these authors, we approximate the distribution of the residuals as $\hat{\mathbf{e}}_\psi(\hat{\theta}_d) \sim \left(\mathbf{I}_{n\times n} - \mathbf{H}(\hat{\theta}_d)\right)' \boldsymbol{\xi}$, where $\boldsymbol{\xi} = \underset{1\le d\le D}{\text{col}}\left(\underset{1\le j\le n_d}{\text{col}}(\xi_{dj})\right)$ and $\xi_{gj} \sim \text{GALI}(\boldsymbol{x}'_{gj}\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d), \hat{\sigma}_{\hat{\theta}_d}, \hat{\theta}_d)$. GALI denotes the Generalized Asymmetric Least Informative distribution with location $\mu_{\hat{\theta}_d} = x'_{gj}\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d)$, scale $\hat{\sigma}_{\hat{\theta}_d}$ and probability $q = \hat{\theta}_d$. The demand for a working likelihood in MQ models for inference purposes motivated the definition of it, which generalizes the asymmetric Laplace distribution, often linked to quantile regression.

The distribution of the optimal robustness parameters $\hat{c}_{\phi,d}$ is estimated via bootstrap using three approaches: a non-parametric (NP) algorithm, a parametric (AP) method based on an approximation to the residuals' distribution and a Naïve (NA) alternative relying on the distribution of the model errors. Let $b = 1, \ldots, B$ denote the bootstrap replicates.

(a1) **Algorithm NP.** Generate $n_d$ values $\hat{u}^{*(b)}_{\psi,dj}$ by simple random sampling with replacement from the set of standardized residuals $\{\hat{u}_{\psi,d1}, \ldots, \hat{u}_{\psi,dn_d}\}$.

(a2) **Algorithm AP.** Generate $n_d$ values $\hat{e}^{*(b)}_{\psi,dj} = \left[\left(\mathbf{I}_{n\times n} - \mathbf{H}(\hat{\theta}_d)\right)' \boldsymbol{\xi}^{*(b)}\right]_{dj}$, where it holds that $\xi^{*(b)}_{gk} \sim \text{GALI}(x'_{gk}\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d), \hat{\sigma}_{\hat{\theta}_d}, \hat{\theta}_d)$. Define $\hat{u}^{*(b)}_{\psi,dj} = \left(\hat{\sigma}^{*(b)}_{\hat{\theta}_d}\right)^{-1} \hat{e}^{*(b)}_{\psi,dj}$, $\hat{\sigma}^{*(b)}_{\hat{\theta}_d} = \text{mad}_{\psi,n}(\hat{e}^{*(b)}_{\psi,dj})/0.6745$.

(a3) **Algorithm NA.** Generate $n_d$ i.i.d. values $\hat{u}^{*(b)}_{\psi,dj} \sim \text{GALI}(0, 1, \hat{\theta}_d)$.

Algorithm NP is based on the empirical distribution of the standardized residuals. It should therefore be preferable to Algorithm NA, but in SAE the sample sizes of the areas are expected to be quite small. Hence, the empirical distribution of $\hat{u}_{\psi,dj}$ would be an inaccurate approximation of the true distribution of $\hat{u}_{\psi,dj}$ because it would be approximated with too little data.

Algorithm AP is based on an approximation of the distribution of the residuals. It is expected to be the preferable method with small and moderate sample sizes and is considered an intermediate option between Algorithm NP and NA. Algorithm NA is the most conservative approach. For each area, the bootstrap distribution of $\hat{c}_{\phi,d}$, according to Algorithm NA, does not depend on $\hat{\sigma}_{\hat{\theta}_d}$ or $N_d$. In fact, it does not even depend on the distribution of the target variables, but only on $\hat{\theta}_d$ and $n_d$. In practice, Algorithm NA ignores the randomness derived from the estimation of $\boldsymbol{\beta}_\psi(\theta_d)$ and the prediction of $\theta_d$. However, it can be used to develop statistical methods conditioned on $\hat{\theta}_d$ and to present empirical results in Section 6 without specifying the model that generates the data. If one were to use the Naïve approach in mixed models, it would consist of approximating the distribution of the residuals by that of the model errors.

In this research, GALI random variables were generated in `R` using custom code.

Let $d = 1, \ldots, D$. The distribution of $\hat{c}_{\phi,d}$ can be approximated via bootstrap as follows:

1. Repeat $B$ times ($b = 1, \ldots, B$):
   (a) Generate standardized residuals using the NP, AP or NA algorithms.
   (b) Based on the bootstrap sample $\{\hat{u}^{*(b)}_{\psi,dj} : \ j = 1, \ldots, n_d\}$, define
   $$A^{*(b)}_d(c_{\phi,d}) = \left(1 - \frac{n_d}{N_d}\right)^2 \left(\frac{\hat{\sigma}_{\hat{\theta}_d}}{n_d}\right)^2 \left(\sum_{j\in s_d} \phi^2\left(\hat{u}^{*(b)}_{\psi,dj}\right) + \left(\sum_{j\in s_d} \left(\phi\left(\hat{u}^{*(b)}_{\psi,dj}\right) - \hat{u}^{*(b)}_{\psi,dj}\right)\right)^2\right). \quad (15)$$
   (c) Solve the minimization problem in the bootstrap world: $\hat{c}^{*(b)}_{\phi,d} = \underset{c_{\phi,d}\ge 0}{\text{argmin}}\ A^{*(b)}_d(c_{\phi,d})$.
2. For $b = 1, \ldots, B$, sort the values $\hat{c}^{*(b)}_{\phi,d}$ from smallest to largest to estimate the distribution of $\hat{c}_{\phi,d}$ via parametric bootstrap. They are $\hat{c}^*_{\phi,d\ (1)} \le \ldots \le \hat{c}^*_{\phi,d\ (B)}$.

## 5. A NEW TEST TO DETECT ATYPICAL AREAS

The main contribution of this research is the detection of atypical areas based on the optimal robustness parameters. For two-level MQ models, we propose to check whether the area-specific MQ coefficients $\theta_d$ are equal to 0.5 in order to identify the atypical condition of an area $d$. Our idea comes from the deviation of the mean of the standardized residuals from the origin, which becomes more pronounced as $\theta_d$ moves

away from 0.5 and translates into a higher bias correction. However, this cannot be done directly because the distribution of $q_{dj}$, $d = 1, \ldots, D$, $j = 1, \ldots, N_d$, is unknown, and so is that of the random variables $\theta_d$.

To overcome this challenge, we will perform inference conditional on $\theta_d$, so that it can be treated as a parameter $0 < \theta_d < 1$. We formulate the test

$$\mathrm{H}_0 : \ \theta_d = 0.5 \ \ \text{vs} \ \ \mathrm{H}_1 : \ \theta_d \neq 0.5, \quad d = 1, \ldots, D. \tag{16}$$

The critical point is $\hat{c}^*_{\phi,d(\lfloor(1-\alpha)B\rfloor)}(0.5)$ and the rejection region is $\left[\hat{c}^*_{\phi,d(\lfloor(1-\alpha)B\rfloor)}(0.5), \infty\right)$, for $0 < \alpha < 1$. To calculate $\hat{c}^*_{\phi,d(\lfloor(1-\alpha)B\rfloor)}(0.5)$, one may use the NP, AP or NA algorithms. Areas with unusually large bias corrections in the BMQ predictor are deemed atypical.

At this regard, the results of test (16) should be interpreted solely in terms of the contrast reflecting the atypical nature of the area $d$ to which they apply, as this is a procedure of conditional inference on $\theta_d$. More specifically, the identical distribution of any subcollection of the set of random variables $\{\theta_d : \ d = 1, \ldots, D\}$ is not intended to be tested.

## 6. SOME OUTSTANDING EMPIRICAL RESULTS

First, the model that generates the data is not specified, only $n_d$ and $\theta_d$, so the outputs are necessarily relative to Algorithm NA in Section 4, with $B = 2000$ bootstrap replicates. Figure 1 plots the kernel density estimation (KDE) of the bootstrap distribution of $\hat{c}_{\phi,d}$ for $\theta_d \in \{0.5, 0.75, 0.95\}$ and $n_d = 10$, the bootstrap median (dashed blue line) and the quantile $q_{0.95}$ (red dots).

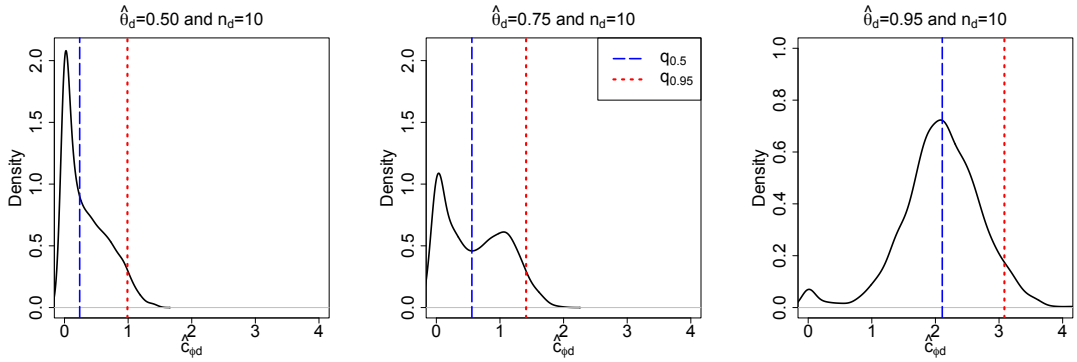



Figure 1: KDE of the bootstrap distribution of $\hat{c}_{\phi,d}$ for $\theta_d \in \{0.5, 0.75, 0.95\}$, $n_d = 10$ and $B = 2000$.

The density of $\hat{c}_{\phi,d}$ for $\theta_d = 0.5$ is unimodal and concentrates around the origin, with a second mode appearing and overtaking the first mode as $\theta_d$ moves away from 0.5. In fact, as $\theta_d$ increases, the distribution shifts to the right even though it maintains an increasingly less appreciable mode around the origin.

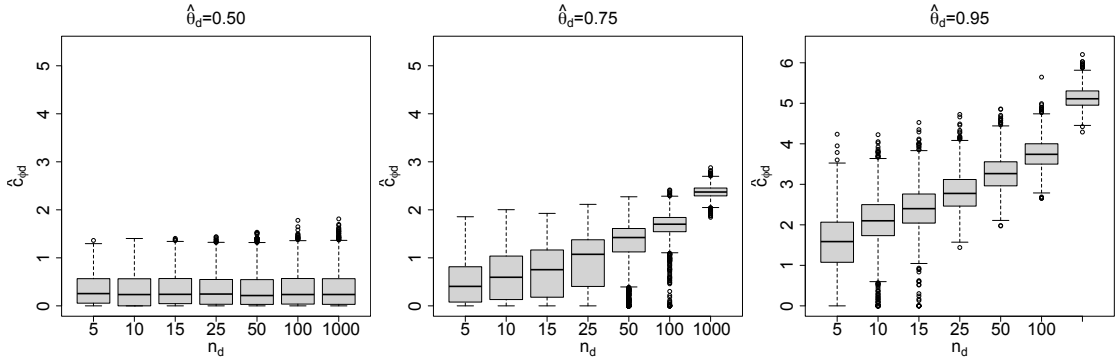



Figure 2: Boxplots of $\hat{c}_{\phi,d}$ for $\theta_d \in \{0.5, 0.75, 0.95\}$ and $n_d \in \{5, 10, 15, 25, 100, 1000\}$ calculated by parametric bootstrap with $B = 2000$.

It follows that the bias correction is more necessary for larger (and smaller) values of $\theta_d$. Actually, the relationship between $\theta_d$ and $\hat{c}_{\phi,d}$ is symmetric around $\theta_d = 0.5$, i.e. smaller values of $\theta_d$ also correspond to

atypical areas. Looking at the vertical scale of the three plots in Figure 1, the data are highly concentrated around the origin for $\theta_d = 0.5$ and become more dispersed as $\theta_d$ increases. The evidence against choosing $\hat{c}_{\phi,d} = 3$ by default (Chambers et al., 2014, 2011) is another notable finding. Figure 1 reveals that the selection of $\hat{c}_{\phi,d} = 3$ is not advisable for non-atypical areas, where $\theta_d$ is close to 0.5, and may even result in an exaggerated bias correction for moderately high (low) values of $\theta_d$, such as $\theta_d = 0.75$ (and $\theta_d = 0.25$)

We would also like to highlight the relationship between $\hat{c}_{\phi,d}$ and $n_d$. The boxplots in Figure 2 show how the distribution of $\hat{c}_{\phi,d}$ moves to the right as the sample size increases. This shift is minimal for $\theta_d = 0.5$ and very noticeable for more distant area-specific MQ coefficients, such as $\theta_d = 0.95$. Again, the selection of $\hat{c}_{\phi,d} = 3$ by default is ruled out as a valid option due to its unquestionable dependence on $n_d$.

Moving on to other topic, the performance in model-based simulations of the new test to detect atypical areas in Section 5 is studied. The simulation design is based on Chambers et al. (2014). Random area and unit-level effects are generated under three scenarios:

$[0, 0]$ – No outliers: $u_d \sim \mathcal{N}(0, 3)$ and $e_{dj} \sim \mathcal{N}(0, 6)$.

$[e, 0]$ – Unit-level outliers: $e_{dj} \sim \delta\,\mathcal{N}(0, 6) + (1 - \delta)\,\mathcal{N}(20, 150)$, where $\delta$ is a Bernoulli variable with $\mathbb{P}(\delta = 1) = 0.97$.

$[e, u]$ – Outliers in both area and unit-level effects: the unit-level errors follow the same distribution as in the previous case and $u_d \sim \mathcal{N}(9, \sqrt{20})$ for areas $37 \leq d \leq 40$.

Each simulation scenario was run with $S = 500$ iterations. The target variables are defined as

$$y_{dj}^{(s)} = 100 + 5x_{dj} + u_d^{(s)} + e_{dj}^{(s)},\ s = 1, \ldots, S,\ j \in U_d,\ d = 1, \ldots, D,$$

where $u_d^{(s)}$ and $e_{dj}^{(s)}$ are generated depending on the specific scenario and $x_{dj} \backsim \text{LogN}(1, 0.5)$.

The simulation results of the area-level outlier detection test using Algorithm AP (the best performer) are presented in Table 1. The proposed method reliably identifies areas as anomalous when they are truly outliers (see column [e, u]), and effectively detects the majority of atypical cases.

| **Scenario** | $[0, 0]$ | $[e, 0]$ | $[e, u]$ |
|---|---|---|---|
| $1 \leq d \leq 40$ | 0.008 | 0.047 | |
| $1 \leq d \leq 36$ | | | 0.048 |
| $37 \leq d \leq 40$ | | | **0.845** |

Table 1: Proportion of outliers detected by Algorithm AP in the atypical area detection test at 5% in model-based simulations.

## 7. APPLICATION TO REAL DATA

In this section, we apply the methodology to analyze income trends in the $D = 8$ provinces of Andalucía (southern Spain) using data from the 2022 Spanish Living Conditions Survey (SLCS) and auxiliary variables from the 2021 Census provided by the Spanish National Institute of Statistics. We have chosen only Andalusian provinces to enhance the graphical interpretability of the results and for space considerations. Sampling fractions are all below 0.11%, underscoring the small-area context. The response variable is the equivalized disposable income per person and unit of consumption, in thousands of euros. Its correlation with the elevation factors is -0.090, suggesting a non-informative sampling design. Since the analysis is based on unit-level data, the range of available auxiliary variables is limited to only two: *sex* and *age4*, a categorical variable divided into four groups: 0–25, 26–45, 46–64 and 65+.

Figure 3 plots the KDE of the bootstrap distribution of the optimal robustness parameters $\hat{c}_{\phi,d}$ by province, sorted by $|\hat{\theta}_d - 0.5|$. Bootstrap estimation of the distribution of $\hat{c}_{\phi,d}$ has been carried out using Algorithm AP of Section 4 with $B = 2000$ replicates. All estimated densities in Figure 3 are unimodal, except for minor perturbations, with the mode shifting to the right as $|\hat{\theta}_d - 0.5|$ increases. Consequently, higher optimal robustness parameters are expected. Higher values of $\hat{c}_{\phi,d}$ indicate greater atypicality in the areas, as shown in Figure 4, where the probability of atypicality is plotted against the confidence level $1 - \alpha$, and calculated as $\frac{1}{B}\sum_{b=1}^{B} I\big(\hat{c}_{\phi,d}^{*(b)} > \hat{c}^*_{\phi,d(\lfloor(1-\alpha)B\rfloor)}(0.5)\big)$, where $\hat{c}_{\phi,d}^{*(b)}$ and $\hat{c}^*_{\phi,d(\lfloor(1-\alpha)B\rfloor)}(0.5)$ have been calculated by bootstrap with $B = 2000$.

The analysis of provincial atypicality reveals deviations from the overall pattern of the target variable. In this case, Huelva, Almería and Jaén –identified as the most atypical provinces in Figure 4– stand out due

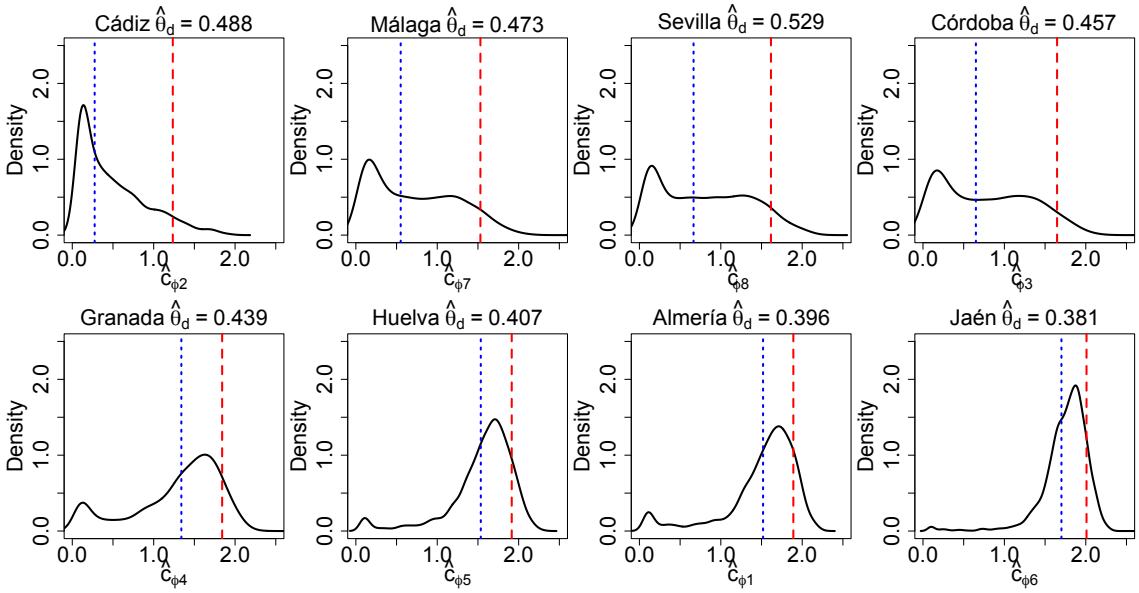



Figure 3: KDE of the bootstrap distribution of the optimal robustness parameters $\hat{c}_{\phi,d}$ by province. The median is a dashed blue line, while the 95% quantile is shown with red dots.

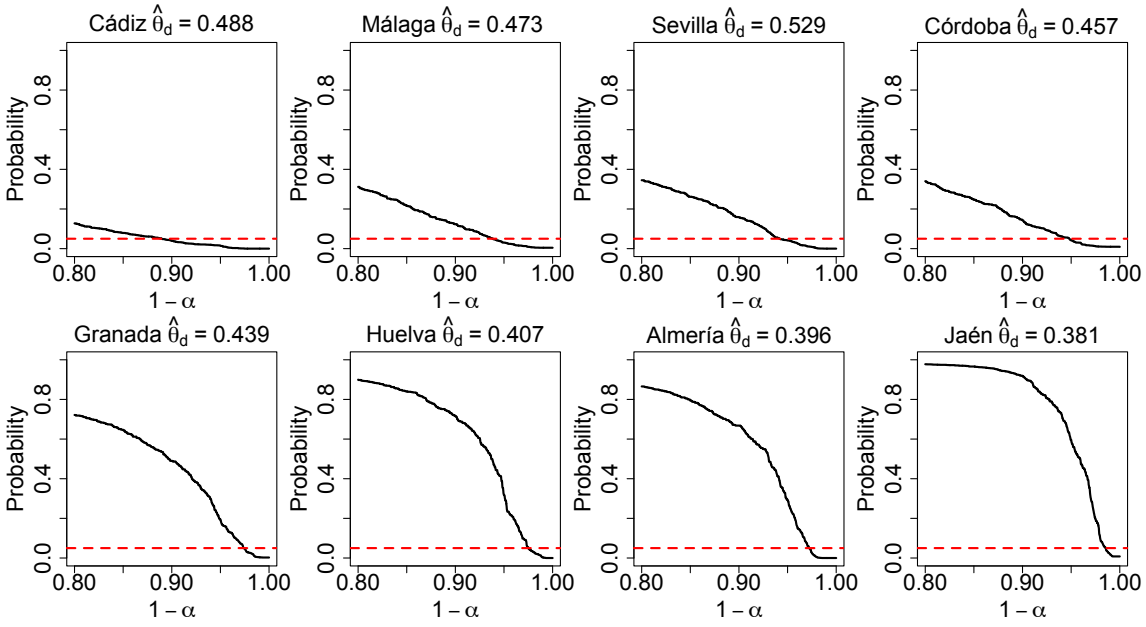
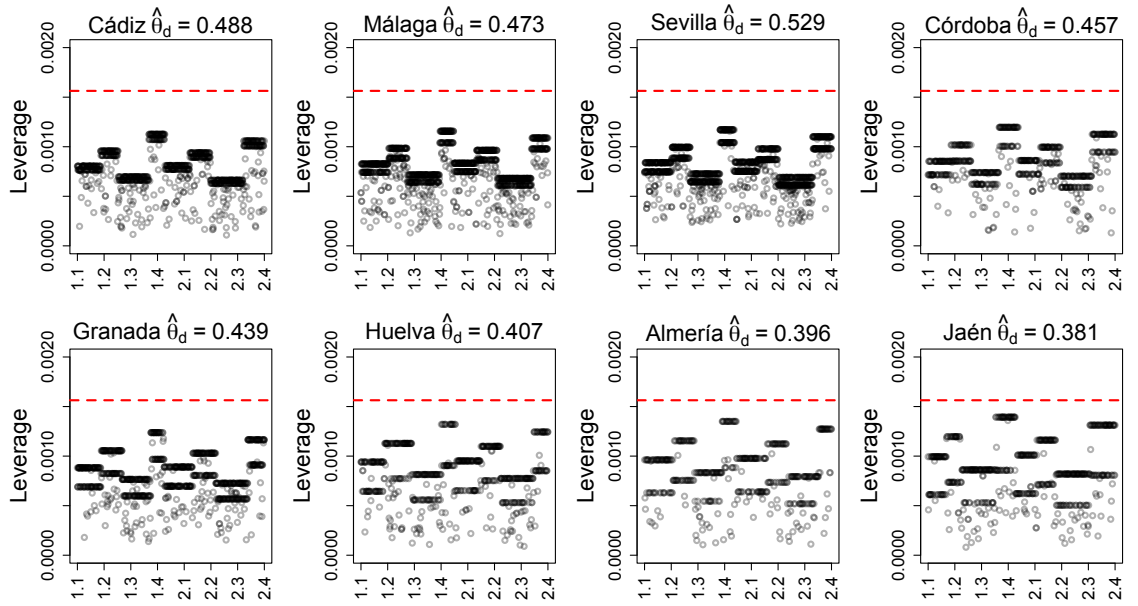



Figure 4: Probability of atypicality for the provinces as a function of the confidence level $1 - \alpha$.

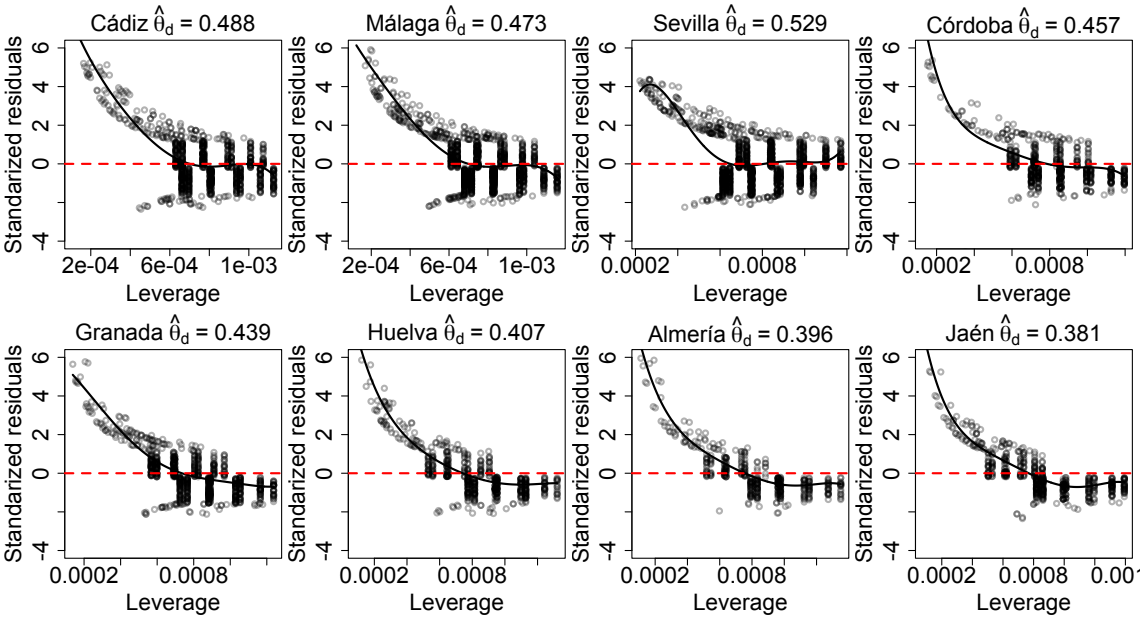
to their significantly lower average equivalized disposable incomes. Jaén, in particular, reports the lowest average not only within Andalucía but also across all the country (Spain), while Huelva and Almería also rank among the poorest. As outliers, these provinces face distinct economic challenges and often diverge from broader national trends.

We now turn to the analysis of the individuals. Section 4 lists properties of the projection matrix in MQ models (around equation (14)), analogous to those of mixed models. One of them is the interpretation of the diagonal elements as leverages in model fitting. An observation $j$ of area $d$ is said to be influential if and only if $h_{ddjj}(q) > 2\tilde{p}/n,\ d = 1,\dots,D,\ j = 1,\dots,n_d,\ 0 < q < 1$, where $\tilde{p}$ is the number of linearly independent covariables, so $\tilde{p} = 5$ in the practical case: $\beta_{\psi_1}$, $sex1 : age41$; $\beta_{\psi_2}$, $sex2$; $\beta_{\psi_3}$, $age42$; $\beta_{\psi_4}$, $age43$; and $\beta_{\psi_5}$, $age44$. According to this criterion, we do not report influential observations for the model fitting with the probabilities $\hat{\theta}_d$, $d = 1,\dots,D$. The maximum of the leverages has been 0.0014 for

$d = 5$ and $2\tilde{p}/n = 0.0016$. Figure 5 (a) displays the leverage values by age and sex; all are below the threshold. Figure 5 (b) shows a clear negative trend between leverages and standardized residuals.




(a) Leverage by age group and sex. Red line: threshold $2\tilde{p}/n = 0.0016$.




(b) Leverage vs. standardized residuals with local polynomial fit.

Figure 5: Results for the fitting of the $D = 8$ two-level MQ regression models.

Finally, Figure 6 (left) maps the BMQ predictions for the average equivalized disposable income in 2022 for the $D = 8$ provinces of Andalucía. This allows for a graphical representation of the differences between provinces. Huelva, Almería and Jaén have lower average incomes, and the opposite is true for Sevilla. The relative root mean squared error (RRMSE) estimates have been computed using the MSE estimator of the BMQ predictor proposed in Chambers et al. (2014) as the numerator and the BMQ predictor itself as the denominator. They are mapped in Figure 6 (right). In terms of RRMSE, the results are really good for a SAE problem, with values below 6% in all provinces in 2022. This performance stands in sharp contrast to that of the direct estimators currently used by the Spanish National Institute of Statistics, which typically exhibit substantially higher RRMSE values.

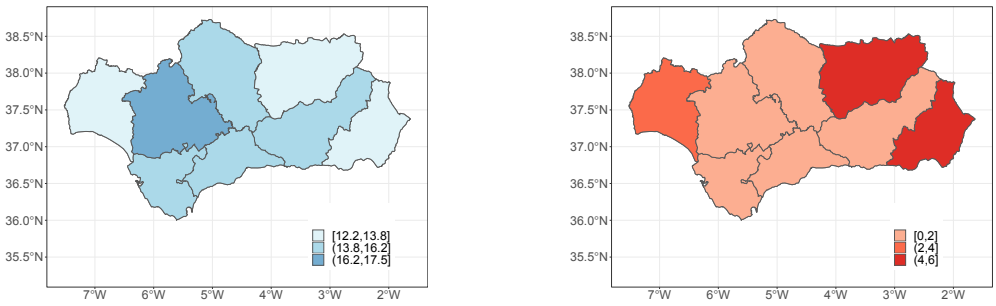



Figure 6: Estimates of average equivalized disposable income (left) and RRMSE estimates (right) in 2022 by province in Andalucía. Results for the BMQ predictor.

## 8. CONCLUSIONS

This research covers the analysis of the residuals in two-level MQ models and calculates their distribution, which are eventually approximated for practical use. Based on these distributions, which are unit-level dependent, both parametric and non-parametric bootstrap methods are proposed to estimate the distribution of the optimal robustness parameters. The idea of the latter is not only to understand their role in the bias correction, but also to alleviate the problem of subjective –but commonly used– selection of 3 (Chambers et al., 2014; Dawber and Chambers, 2019) and to analyse their relationship with the atypical condition of an area.

While robust methods are increasingly applied in SAE (Chambers et al., 2011; Chambers and Tzavidis, 2006), little attention has been paid to detecting atypical areas (Bugallo et al., 2025). We address this gap by proposing a bootstrap-based test to identify outliers, aiming to balance reliability and parsimony in $p$-value-based detection. This test is grounded in the behavior of the residuals and standardized residuals. Simulation studies support the validity of our approach: the test effectively detects clear outliers, though its sensitivity to moderately atypical areas is limited –highlighting the nuanced nature of robustness in SAE. The application to the SLCS2022 data demonstrates the practical value of the research. The bootstrap distribution of the optimal robustness parameters offers a natural way to rank atypicality and assess deviations from regional economic patterns.

Future studies should investigate the robustness properties of predictors derived from MQ models and examine how the selection of the optimal robustness parameters influences their performance. For instance, assessing robustness in the presence of outliers or heavily skewed distributions remains an open question. Additionally, the insights derived from the optimal selection of the robustness parameters offer valuable guidance improving future MQ applications.

## REFERENCES

Bianchi, A. and N. Salvati (2015). Asymptotic properties and variance estimators of the M-quantile regression coefficients estimators. *Communications in Statistics 44*, 813–827.

Bianchi, A., E. Fabrizi, N. Salvati, and N. Tzavidis (2018). Estimation and testing in M-quantile regression with applications to Small Area Estimation. *International Statistical Review 86*, 541–570.

Breckling, J. and R. Chambers (1988). M-quantiles. *Biometrika 75(4)*, 761–771.

Bugallo, M., Esteban, M. D., Hobza, T., Morales, D., and Pérez, A. (2024). Small Area Estimation of labour force indicators under unit-level multinomial mixed models. *Journal of the Royal Statistical Society: Series A 188(1)*, 241–270

Bugallo, M., D. Morales, N. Salvati, and F. Schirripa (2025). Temporal M-quantile models and robust bias-corrected small area predictors. *Journal of Survey Statistics and Methodology. DOI: 10.1093/jssam/smaf007.*

Bugallo, M. and Morales, D. (2025). Inference and diagnosis in M-quantile models with applications to Small Area Estimation. *Unpublished manuscript.*

Chambers, R. and N. Tzavidis (2006). M-quantile models for Small Area Estimation. *Biometrika 93(2)*, 255–268.

Chambers, R., H. Chandra, and N. Tzavidis (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methodology 37(2)*, 153–170.

Chambers, R., H. Chandra, N. Salvati, and N. Tzavidis (2014). Outlier robust Small Area Estimation. *Journal of the Royal Statistical Society: Series B 76 (1)*, 47–69.

Dawber, J. and R. Chambers (2019). Modelling group heterogeneity for Small Area Estimation using M-quantiles. *International Statistical Review 87 (1)*, 50–63.

Marchetti, S., Beresewicz, M., Salvati, N., Szymkowiak, M., and Wawrowski, L. (2018). The use of a three-level M-quantile model to map poverty at local administrative unit 1 in Poland. *Journal of the Royal Statistical Society: Series A 181(4)*, 1077–1104.

Morales, D., M. D. Esteban, A. Pérez, and T. Hobza (2021). *A Course on Small Area Estimation and Mixed Models: Methods, Theory and Applications in R*. Springer Nature.

Rao, J. and I. Molina (2015). *Small Area Estimation*. John Wiley & Sons.

Salvati, N., Tzavidis, N., Pratesi, M., and Chambers, R. (2012). Small Area Estimation via M-quantile Geographically Weighted Regression. *TEST 21*, 1–28.