

Estimation and Testing in M-quantile Regression with Applications to Small Area Estimation

Annamaria Bianchi¹, Enrico Fabrizi², Nicola Salvati³ 
and Nikos Tzavidis⁴

¹DSAEMQ, Università di Bergamo, Bergamo, Italy

²DISES, Università Cattolica del S. Cuore, Milan, Italy

³DEM, Università di Pisa, Pisa, Italy

⁴University of Southampton, Southampton, UK

E-mail: nicola.salvati@unipi.it

Summary

In recent years, M-quantile regression has been applied to small area estimation to obtain reliable and outlier robust estimators without recourse to strong parametric assumptions. In this paper, after a review of M-quantile regression and its application to small area estimation, we cover several topics related to model specification and selection for M-quantile regression that received little attention so far. Specifically, a pseudo- R^2 goodness-of-fit measure is proposed, along with likelihood ratio and Wald type tests for model specification. A test to assess the presence of actual area heterogeneity in the data is also proposed. Finally, we introduce a new estimator of the scale of the regression residuals, motivated by a representation of the M-quantile regression estimation as a regression model with Generalised Asymmetric Least Informative distributed error terms. The Generalised Asymmetric Least Informative distribution, introduced in this paper, generalises the asymmetric Laplace distribution often associated to quantile regression. As the testing procedures discussed in the paper are motivated asymptotically, their finite sample properties are empirically assessed in Monte Carlo simulations. Although the proposed methods apply generally to M-quantile regression, in this paper, their use is illustrated by means of an application to Small Area Estimation using a well known real dataset.

Key words: Generalised Asymmetric Least Informative distribution; goodness-of-fit; likelihood ratio type test; loss function; robust regression.

1 Introduction

In sample surveys, estimates of population descriptive quantities for a target variable y are usually needed both for the population as a whole and for subpopulations, known as domains or areas. Provided that large enough domain-specific sample sizes are available, statistical agencies can perform domain estimation by using the same design-based methods used for the estimation of population level quantities (direct estimation). In the case of small domain-specific sample

sizes, direct estimation may lead to estimates with large sampling variability. When direct estimation is not reliable in all or most of the domains, there is need to use small area estimation (SAE) techniques.

Area-level and unit-level linear mixed models have been studied in the literature to obtain empirical best linear unbiased predictors of small area means (Rao & Molina, 2015, chapter 5). Empirical best estimation is useful for estimating the small area means efficiently when normality holds, otherwise, its properties can be deteriorated especially by the presence of outliers in the data. Consequently, it is of interest to see how robust estimation can be adapted to SAE.

In recent years, Chambers & Tzavidis (2006) and Sinha & Rao (2009) addressed the issue of outlier robustness in SAE proposing techniques that can be used to down-weight any outliers when fitting the underlying model. Sinha & Rao (2009) addressed this issue from the perspective of linear mixed models. Chambers & Tzavidis (2006) proposed to apply the M-quantile (hereafter, MQ) regression models to SAE with the aim of obtaining reliable and outlier robust estimators without recourse to parametric assumptions for the residuals distribution using M-estimation theory. A comparison of these two alternative approaches can be found in Chambers *et al.* (2014a). The distinguishing features of the approach by Chambers & Tzavidis (2006) include the protection that a careful choice of a quantile-specific loss function $\rho_\tau(\cdot)$, $0 < \tau < 1$, offers against the effect of outliers and the characterisation of domain heterogeneity in terms of domain-specific MQs. They can be viewed as an alternative to random effects for measuring area-specific unobserved heterogeneity. Whenever there is insufficient evidence of this heterogeneity, a prediction based on a simpler median linear regression model would be more efficient.

A number of papers on MQ regression that focus on theoretical developments (Tzavidis *et al.*, 2010; Fabrizi *et al.*, 2012; Salvati *et al.*, 2012; Bianchi & Salvati, 2015; Chambers *et al.*, 2014a; Fabrizi *et al.*, 2014a; Tzavidis *et al.*, 2016; Alfò *et al.*, 2017), extensions to non-linear models (Pratesi *et al.*, 2009; Chambers *et al.*, 2014b; Dreassi *et al.*, 2014; Tzavidis *et al.*, 2015; Chambers *et al.*, 2016) and various small area applications (Tzavidis *et al.*, 2008; Pratesi *et al.*, 2008; Salvati *et al.*, 2011; Tzavidis *et al.*, 2012; Fabrizi *et al.*, 2014b) have been published in recent years. In view of this growing number of studies, in this paper, we review MQ linear regression with special focus on its application to SAE.

The aim of SAE is to complement and extend the published official statistics. As models play a key role in MQ-based SAE, it is important that they are carefully selected and checked. While in the SAE literature based on linear mixed models, diagnostics and model selection are widely discussed (Rao & Molina, 2015, section 5.4), these topics have received, comparatively, little attention in MQ applications to SAE. For this reason, we complement the review of MQ regression by proposing (a) a measure of goodness-of-fit of the model parallel to the ordinary R^2 ; (b) model specification tests based on likelihood ratio and Wald type tests; and (c) a test on the presence of actual area heterogeneity.

As a preliminary step, we introduce the parametric distribution associated with a general loss function $\rho_\tau(\cdot)$, that we will call Generalised Asymmetric Least Informative (GALI) distribution. This distribution relates to MQ regression in the same way as the normal distribution is linked to ordinary least squares and the asymmetric Laplace distribution to quantile regression (Yu & Moyeed, 2001). The likelihood under the GALI model is a working likelihood, that is, it is only used to facilitate maximum likelihood estimation. We use this term following Yang *et al.* (2015) who uses the same term for maximum likelihood estimation of quantile regression parameters under the asymmetric Laplace distribution. We propose a new estimator for the scale parameter based on the GALI distribution. Furthermore, if the loss function ρ_τ involves tuning constants that regulate the trade-off between robustness and efficiency (as is the case with the Huber

loss function), we propose to use the distribution associated to the loss function to estimate a data-driven value for these tuning constants.

The goodness-of-fit measure (analogous to the usual R^2) that we propose is similar to that introduced by Koenker & Machado (1999) for ordinary quantile regression. The likelihood ratio and the Wald type tests that we discuss are motivated asymptotically. We explore their finite sample properties by simulation exercises.

The paper is organised as follows. In Section 2, we review MQ regression, while its application to SAE is reviewed in Section 3. In Section 4, we introduce the GALI distribution and new estimators for the scale of the residuals in MQ regression and the tuning constant. In Section 5, we introduce the pseudo- R^2 goodness-of-fit measure and likelihood ratio and Wald type tests for linear hypotheses on the MQ regression parameters. Section 6 presents the test for assessing the presence of area-specific effects. In Section 7, we present simulation studies aimed at assessing the finite sample properties of the proposed tests and estimators. In Section 8, we discuss an application of the methods to real data. Finally, Section 9 concludes the paper with some final remarks.

2 A Review of M-quantile Regression

Quantile regression (Koenker & Bassett, 1978; Koenker, 2005) represents a useful generalisation of median regression whenever the interest is not limited to the estimation of a location parameter at the centre of the conditional distribution of the target variable, but it focuses also on location parameters (quantiles) at different parts of this conditional distribution. Similarly, expectile regression (Newey & Powell, 1987) generalises least squares regression to estimation of location parameters at other parts of the target conditional distribution, namely, expectiles. Breckling & Chambers (1988) introduce MQ regression that extends the ideas of M-estimation (Huber, 1964; Huber & Ronchetti, 2009) to a different set of location parameters of the target conditional distribution that lie between quantiles and expectiles. MQs aim at combining the robustness properties of quantiles with the efficiency properties of expectiles.

Given an (a.e.) continuously differentiable convex loss function $\rho(u)$, $u \in \mathcal{R}$, we define the tilted version of the loss function as

$$\rho_\tau(u) = |\tau - I(u < 0)|\rho(u), \quad (1)$$

where $\tau \in (0, 1)$ represents the MQ. In line with most of the applications cited in the introduction, in this paper, special attention is devoted to the tilted version of the popular Huber loss function,

$$\rho_\tau(u) = 2 \begin{cases} (c|u| - c^2/2)|\tau - I(u \leq 0)| & |u| > c \\ u^2/2|\tau - I(u \leq 0)| & |u| \leq c, \end{cases} \quad (2)$$

where $I(\cdot)$ is an indicator function, and c is a cutoff constant. By setting $\tau = 0.5$, a well-known distribution, the so-called least informative distribution, is associated to this function (Huber, 1981, section 4.5).

Given a random variable y with cdf $F(y)$, the τ -th MQ of y , denoted θ_τ , is obtained as the minimiser of

$$\int \rho_\tau(y - \theta_\tau)F(dy). \quad (3)$$

Depending on the choice of the loss function, MQs may reduce to ordinary quantiles ($\rho(u) = |u|$) or to expectiles ($\rho(u) = u^2$), while other choices are also possible (Dodge & Jureckova,

2000). However, as it is well known, quantiles and expectiles should be treated separately due to different properties of the corresponding influence functions (Wooldridge, 2010, p. 407).

In case of regression, the argument inside the loss function is replaced by standardised residuals. More precisely, let \mathbf{x} a p -dimensional random vector with first component $x_1 = 1$. The observed data $\{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$ are assumed to be a random sample of size n drawn from an infinite population; thus, (\mathbf{x}_i, y_i) are independent and identically distributed random variables. Assuming a linear model, for any $\tau \in (0, 1)$, the MQ of order τ of y_i given \mathbf{x}_i is defined by

$$MQ_\tau(y_i|\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_\tau, \quad (4)$$

where $\boldsymbol{\beta}_\tau \in \Theta \subset \mathcal{R}^p$ is the solution to

$$\min_{\boldsymbol{\beta}_\tau \in \Theta} E \left[\rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau}{\sigma_\tau} \right) \right], \quad (5)$$

and σ_τ is a scale parameter that characterises the distribution of $\varepsilon_{\tau i} = y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau$. The linear specification in (4) can be alternatively written as

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_\tau + \varepsilon_{\tau i},$$

where $\{\varepsilon_{\tau i}\}$ is a sequence of independent and identically distributed errors with unknown distribution function F_τ satisfying, by definition, $MQ_\tau(\varepsilon_{\tau i}|\mathbf{x}_i) = 0$. The estimator of the MQ regression coefficients (Breckling & Chambers, 1988) is defined as

$$\hat{\boldsymbol{\beta}}_\tau = \arg \min \sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau}{\hat{\sigma}_\tau} \right), \quad (6)$$

where $\hat{\sigma}_\tau$ is a consistent estimator of σ_τ . Because ρ is (a.e.) continuously differentiable and convex, the vector $\hat{\boldsymbol{\beta}}_\tau$ can equivalently be obtained as the solution of the following system of equations

$$\sum_{i=1}^n \psi_\tau \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau}{\hat{\sigma}_\tau} \right) \mathbf{x}_i = \mathbf{0}, \quad (7)$$

where $\psi_\tau(u) = d\rho_\tau(u)/du = |\tau - I(u < 0)|\psi(u)$, with $\psi(u) = d\rho(u)/du$. An iterative method is needed here to obtain a solution, like an iteratively re-weighted least squares algorithm or the Newton–Raphson algorithm.

Regarding the scale parameter σ_τ , it may generally be defined by an implicit relation of the form

$$E \left[\chi \left(\frac{\varepsilon_{\tau i}}{\sigma_\tau} \right) \right] = 0, \quad (8)$$

where the expectation is taken with respect to the distribution of $\varepsilon_{\tau i}$. In MQ regression, a typical choice for χ is $\chi(u) = \text{sgn}(|u - \text{Med}(u)| - 1)$, which leads to the scaled population median absolute deviation $\sigma_\tau = \text{Med}\{|\boldsymbol{\varepsilon}_\tau - \xi_{1/2,\tau}|\}/q$, where $\xi_{1/2,\tau} = \text{Med}(F_\tau(\boldsymbol{\varepsilon}_\tau))$,

$q = \Phi^{-1}(3/4) = 0.6745$, with Φ denoting the distribution function of the standard normal distribution. The constant q is generally added at the denominator of the mean absolute deviation (MAD), so that it corresponds to the standard deviation in case of normality (e.g., Falk, 1997). The corresponding estimator is the scaled sample MAD

$$\hat{\sigma}_\tau = \frac{\text{Med}\{|\hat{\mathbf{e}}_\tau - \text{Med}(\hat{\mathbf{e}}_\tau)|\}}{q}, \quad (9)$$

where $\hat{\mathbf{e}}_\tau = (\hat{e}_{\tau 1}, \dots, \hat{e}_{\tau n})$, $\hat{e}_{\tau i} = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\tau$. The consistency of the MAD estimator can be proved by standard theory of M-estimators (Wooldridge, 2010), assuming that (8) has a unique solution. Uniqueness of the solution in (8) is related to the absolute continuity of the distribution of the errors. We refer the reader to Hall & Welsh (1985) for detailed conditions and a proof of the strong consistency of the MAD estimator.

The asymptotic theory for MQ regression with independent and identically distributed (iid) errors and fixed regressors can be derived from the results in Huber (1973), as pointed out in Breckling & Chambers (1988). In case of stochastic regressors and in the presence of heteroskedasticity, Bianchi & Salvati (2015) show the consistency and the asymptotic normality of $\hat{\boldsymbol{\beta}}_\tau$ and the consistency of its asymptotic variance estimator,

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_\tau) = (n - p)^{-1} n \hat{\mathbf{W}}_\tau^{-1} \hat{\mathbf{G}}_\tau \hat{\mathbf{W}}_\tau^{-1} \quad (10)$$

where

$$\begin{aligned} \hat{\mathbf{W}}_\tau &= (n \hat{\sigma}_\tau)^{-1} \sum_{i=1}^n \hat{\psi}'_{\tau i} \mathbf{x}_i \mathbf{x}_i^T, \\ \hat{\mathbf{G}}_\tau &= n^{-1} \sum_{i=1}^n \hat{\psi}_{\tau i}^2 \mathbf{x}_i \mathbf{x}_i^T, \end{aligned}$$

with $\hat{\psi}'_{\tau i} := \psi'_\tau(\hat{e}_{i\tau}/\hat{\sigma}_\tau)$, $\hat{\psi}_{\tau i} = \psi_\tau(\hat{e}_{i\tau}/\hat{\sigma}_\tau)$.

3 A Review of M-quantile Models in Small Area Estimation

Let us now consider a finite population and suppose that it is divided into D non-overlapping small areas of size N_j , $j = 1, \dots, D$, so that $\sum_{j=1}^D N_j = N$. Suppose that a sample of size $n_j > 0$ is drawn from each small area. For simplicity of exposition, we do not consider the case $n_j = 0$, although the theory can be easily extended to it. For convenience, we introduce a second subscript in the notation to indicate the hierarchical nature of the data, $\{(\mathbf{x}_{ij}, y_{ij}), i = 1, \dots, n_j; j = 1, \dots, D\}$. Here, values y_{ij} represent the variable of interest, and values \mathbf{x}_{ij} of a $p \times 1$ vector are the individual level covariates. For the non-sampled population units, we assume that the values of \mathbf{x}_{ij} are known. We also assume that sampling is non-informative for the small area distribution of y_i given \mathbf{x}_{ij} , allowing us to use population level models with the sample data.

The papers by Chambers & Tzavidis (2006) and Aragon *et al.* (2005) were the first to introduce the idea of measuring heterogeneity in the data via MQs. In particular, Chambers & Tzavidis (2006) characterise the variability across the population of interest by introducing the idea of MQ coefficients. At the population level, the MQ coefficient for a unit within a small area is defined as the value τ_{ij} such that $MQ_{\tau_{ij}}(y_{ij}|\mathbf{x}_{ij}) = y_{ij}$. If a hierarchical structure does explain part of the variability, after accounting for the effect of covariates, units within small

area are expected to have similar MQ coefficients. Chambers & Tzavidis (2006) propose to characterise each small area j by the average of the MQ coefficients of the units that belong to that small area. The small area-specific MQ coefficient, denoted by τ_j , identifies the most characteristic MQ regression line for that small area. We can think of this in the context of linear mixed models as the group-specific regression line that is distinguished from population-average line by the random effect. The aim is to use the small area-specific MQ coefficient τ_j to predict various area specific quantities, including (but not only) the area mean of y . In what follows, we focus on the estimation of the small area j mean of y , denoted m_j . When (4) holds, and β_τ is a sufficiently smooth function of τ , Chambers & Tzavidis (2006) suggest a predictor of m_j of the form:

$$\hat{m}_j^{MQ} = N_j^{-1} \left\{ \sum_{i \in s_j} y_{ij} + \sum_{i \in r_j} \mathbf{x}_{ij}^T \hat{\beta}_{\hat{\tau}_j} \right\}, \quad (11)$$

where we use indices s and r to denote sample and non-sample quantities, respectively. Thus, the set s_j contains the n_j indices of the units drawn from the population, and the set r_j contains the $N_j - n_j$ indices of the non-sampled units in small area j . Here, $\hat{\tau}_j$ is an estimate of the average value of the MQ coefficients of the units in area j . The case of $n_j = 0$, mentioned earlier, can be easily dealt with by using a synthetic MQ predictor, which is obtained by setting $\hat{\tau}_j = 0.5 (\hat{m}_j^{MQ/SYN})$. According to Chambers *et al.* (2014a), such method is defined as *robust-projective* as it projects sample non-outlier (i.e. working model) behaviour onto the non-sampled part of the survey population. In contrast, Chambers *et al.* (2014a) propose methods to address the presence of representative outliers (Chambers, 1986), that is, sample outliers that are potentially drawn from a group of population outliers and hence cannot be unit weighted in estimation. This method allows for contributions from representative sample outliers, and it is defined as a *robust-predictive* method because it attempts to predict the contribution of the population outliers to the population quantity of interest. In the robust-predictive context, a bias-corrected version of estimator (11) is given by

$$\hat{m}_j^{MQ-BC} = N_j^{-1} \left\{ \sum_{i \in s_j} y_{ij} + \sum_{i \in r_j} \mathbf{x}_{ij}^T \hat{\beta}_{\hat{\tau}_j} + \frac{N_j - n_j}{n_j} \sum_{i \in s_j} \omega_{ij}^{MQ} \phi \left\{ \frac{y_{ij} - \mathbf{x}_{ij}^T \hat{\beta}_{\hat{\tau}_j}}{\omega_{ij}^{MQ}} \right\} \right\}, \quad (12)$$

where ω_{ij}^{MQ} is a robust estimator of the scale of the residual $y_{ij} - \mathbf{x}_{ij}^T \hat{\beta}_{\hat{\tau}_j}$ in area j . The robust influence function ψ , used to define $\hat{\beta}_{\hat{\tau}_j}$, is replaced in the third addend of (12) by ϕ ; such function is still bounded but more accommodating with respect to sample outliers, that is, such that $|\psi| \leq |\phi|$. Its purpose is to define an adjustment for the bias caused by the fact that the first two terms on the right hand side of (12) treat sample outliers as not representative (for details, see Chambers *et al.* 2014a). If ϕ is the identity function, predictor (12) corresponds to the Chambers and Dunstan estimator (Tzavidis *et al.*, 2010).

Two different analytic methods of mean squared error (MSE) estimation for MQ-based robust predictors of small area means under the robust-projective and robust-predictive approaches have been proposed in the literature. Both are developed under the assumption that the working model for inference is conditioned on the realised values of the area effects. As a consequence, the proposed MSE estimators are conditional estimators. The first method is introduced by Chambers *et al.* (2011) as a pseudo-linearisation estimator for the conditional MSE of predictor (11), and it is labelled as CCT (Chambers, Chandra, Tzavidis) estimator. The second method uses first-order approximations to the variances of solutions of estimating equations to develop conditional MSE estimators for predictors (11) and (12), and it is labelled as CST (Chambers,

Chandra, Salvati, Tzavidis) (Chambers *et al.*, 2014a). The MSE estimator for predictor (12) is based on the approximation

$$MSE(\hat{m}_j^{MQ-BC}) = \left(1 - \frac{n_j}{N_j}\right)^2 \left[\{\bar{\mathbf{x}}_{rj} - \bar{\mathbf{x}}_{sj}\}^T \hat{V}(\hat{\boldsymbol{\beta}}_{\hat{\tau}_j}) \{\bar{\mathbf{x}}_{rj} - \bar{\mathbf{x}}_{sj}\} + \hat{V}(\bar{e}_{rj}) + \frac{1}{n_j^2} \sum_{i \in s_j} \left\{ \omega_{ij}^{MQ} \phi \left\{ \frac{y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{\hat{\tau}_j}}{\omega_{ij}^{MQ}} \right\} \right\}^2 \right], \quad (13)$$

where $\hat{V}(\hat{\boldsymbol{\beta}}_{\hat{\tau}_j})$ is the estimated variance of the fitted MQ regression coefficients at $\tau = \hat{\tau}_j$, $\hat{V}(\bar{e}_{rj}) = (N_j - n_j)^{-1}(n - 1)^{-1} \sum_k \sum_{i \in s_k} (y_{ik} - \mathbf{x}_{ik}^T \hat{\boldsymbol{\beta}}_{\hat{\tau}_k})^2$ and $\bar{\mathbf{x}}_{rj}$, $\bar{\mathbf{x}}_{sj}$ denote the vectors of average values of \mathbf{x}_{ij} for the $N_j - n_j$ non-sampled units and the n_j of sampled units, respectively, in area j . The results of the simulation experiments in Chambers *et al.* (2014a) show that the CST estimator has lower bias than the CCT and that it is also more stable for both predictors (11) and (12).

Several methodological developments on MQ regression in SAE have been made in recent years and they are briefly reviewed in the following. Fabrizi *et al.* (2012) consider two problems relevant to practical small area applications. First, they propose a solution to guarantee the benchmarking property of small area estimators. The proposed procedure is consistent with the MQ regression framework, thus, it is theoretically more interesting than a simple ratio adjustment. Second, they consider the problem of the correction of the under/over-shrinkage of small area estimators. The authors note that the MQ small area estimators may under-shrink (under normality) or over-shrink (when the distribution of actual small area parameters is skewed). In line with most of the literature, notions of under-shrinkage and over-shrinkage are defined in terms of variance calculated over the ensemble of small area parameters. This may not be robust to the presence of outlying areas, but the method proposed by Fabrizi *et al.* (2012) can be readily extended to other descriptions of the variability of the ensemble of area parameters.

Fabrizi *et al.* (2014a) adopt a model-assisted approach for developing design-consistent (weighted) MQ small area estimators. The authors assume a working linear MQ model and consider only properties with respect to the randomisation distribution induced by the sample design. Fabrizi *et al.* (2014a) note that for the estimation of small area means and totals, the weighted MQ based estimators may be expressed in generalized regression (GREG) form and can therefore be easily interpreted. Moreover, this estimator remains consistent even under informative sampling design provided that the sampling weights incorporate all the relevant information about the selection process.

Salvati *et al.* (2012) incorporate the spatial information in small area predictors based on MQ models via geographically weighted regression. In particular, the authors specify an MQ-geographically weighted regression model that is a local model for the MQs of the conditional distribution of the outcome variable given the covariates. This model is then used to define a bias-robust predictor of the small area characteristic of interest that also accounts for spatial association in the data. Another approach to take into account spatial information in small area MQ predictors is by using a semiparametric MQ regression model as proposed by Pratesi *et al.* (2008). In this case, the response variable depends on the geographical position of the observations through an unknown smooth bivariate function estimated by low-rank thin plate splines. From the simulation results, the semiparametric MQ models in SAE appear to be a useful tool when the functional form of the relationship between the variable of interest and the covariates is left unspecified, and the data are characterised by complex patterns of spatial dependence (Salvati *et al.*, 2011).

The MQ approach to small area prediction has been extended to discrete responses. In particular, Tzavidis *et al.* (2015) propose a small area predictor based on a new semiparametric MQ model for counts that extends the ideas of Cantoni & Ronchetti (2001) and Chambers & Tzavidis (2006). This predictor can be viewed as an outlier robust alternative to the more commonly used conditional expectation predictor for counts that is based on a Poisson generalised linear mixed models with Gaussian random effects. Chambers *et al.* (2014b) introduce a semi-parametric approach to ecological regression for disease mapping, based on modelling the regression MQs of a negative binomial variable. The method is robust to outliers in the model covariates, including those due to measurement error, and can account for both spatial heterogeneity and spatial clustering. Chambers *et al.* (2016) extend the MQ approach to SAE for counts (Tzavidis *et al.*, 2015; Chambers *et al.*, 2014b) to the case where the response is binary. Modelling the MQs of a binary outcome presents more challenges than modelling the MQs of a count outcome. A detailed account of these challenges is provided in the paper. With the proposed approach, random effects are avoided, and between-area variation in the response is characterised by variation in area-specific values of MQ indices. Furthermore, outlier robust inference is achieved in the presence of both misclassification and measurement error.

A common criticism of the application of MQ regression in SAE is the limited availability of model selection and diagnostics tools that are well researched for linear mixed models. Focusing on linear MQ models, we partially fill this gap in Sections 5 and 6 discussing new tools for model selection and diagnostics. Some of these diagnostics can be applied to general MQ regression models, and some are specific to SAE problems.

4 A Working Likelihood for M-quantiles: The Generalised Asymmetric Least Informative distribution

Yu & Moyeed (2001) show the relationship between the loss function for quantile regression and the maximisation of a likelihood function for independently distributed asymmetric Laplace random variables. In this section, we show a similar relationship for MQ regression models.

Given the loss function ρ_τ , in an infinite population context, we define the GALI random variable as the random variable having density function

$$f_\tau(y; \mu_\tau, \sigma_\tau) = \frac{1}{\sigma_\tau B_\tau} \exp \left\{ -\rho_\tau \left(\frac{y - \mu_\tau}{\sigma_\tau} \right) \right\}, \quad -\infty < y < +\infty. \quad (14)$$

where $B_\tau = \int_{-\infty}^{+\infty} \frac{1}{\sigma_\tau} \exp \left\{ -\rho_\tau \left(\frac{y - \mu_\tau}{\sigma_\tau} \right) \right\} dy < +\infty$, and μ_τ and σ_τ are location and scale parameters, respectively. We note that μ_τ coincides with the τ^{th} MQ of the distribution; in fact, μ_τ can be obtained as the solution of

$$\int_{-\infty}^{+\infty} \psi_\tau \left(\frac{y - \mu_\tau}{\sigma_\tau} \right) f_\tau(y; \mu_\tau, \sigma_\tau) dy = 0,$$

that defines the MQ of the distribution.

Because the theory developed in this section can be applied more generally and not only to SAE, we drop subscript j from our notation. For linear MQ regression, that is, when $\mu_\tau = \mu_{\tau i} = \mathbf{x}_i^T \boldsymbol{\beta}_\tau$, the estimators of the unknown regression parameters $\boldsymbol{\beta}_\tau$ and the scale σ_τ may be obtained by maximising the log-likelihood function obtained from densities (14):

$$l_\tau(y) = -n \log \sigma_\tau - n \log B_\tau - \sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau}{\sigma_\tau} \right). \quad (15)$$

Given σ_τ , the estimating equations for the regression coefficients β_τ are the same as those of Equation (7). The estimating equation for σ_τ is

$$-\frac{n}{\sigma_\tau} + \frac{1}{\sigma_\tau^2} \sum_{i=1}^n \psi_\tau \left(\frac{y_i - \mathbf{x}_i^T \beta_\tau}{\sigma_\tau} \right) (y_i - \mathbf{x}_i^T \beta_\tau) = 0, \quad (16)$$

and its solution defines a new estimator for σ_τ alternative to (9). With respect to (8), in this case, $\chi(u) = -u\psi_\tau(u) - 1$, and the parameter is defined as the solution of

$$E \left[-\varepsilon_{\tau i} \psi_\tau \left(\frac{\varepsilon_{\tau i}}{\sigma_\tau} \right) \right] = \sigma_\tau.$$

This choice is in line with what Koenker & Machado (1999) and Yu & Zhang (2005) propose for quantile regression, considering the maximum likelihood estimator under the asymmetric Laplace distribution.

Solving Equations (7) and (16) requires an iterative algorithm. The steps of this algorithm are as follows:

1. For specified τ , define initial estimates $\hat{\beta}_\tau^{(0)}$ and $\hat{\sigma}_\tau^{(0)}$.
2. At each iteration t , calculate $w_{\tau i}^{(t-1)} = \psi_\tau(u_{\tau i}^{(t-1)})/u_{\tau i}^{(t-1)}$ with $u_{\tau i}^{(t-1)} = (y_i - \mathbf{x}_i^T \hat{\beta}_\tau^{(t-1)})/\hat{\sigma}_\tau^{(t-1)}$.
3. Compute the new weighted least squares estimates from

$$\hat{\beta}_\tau^{(t)} = \left\{ \sum_{i=1}^n (w_{\tau i}^{(t-1)} \mathbf{x}_i \mathbf{x}_i^T) \right\}^{-1} \left\{ \sum_{i=1}^n (y_i w_{\tau i}^{(t-1)} \mathbf{x}_i) \right\}. \quad (17)$$

4. Compute the new estimate of $\hat{\sigma}_\tau$ by

$$\hat{\sigma}_\tau^{(t)} = \left\{ n^{-1} \sum_{i=1}^n w_{\tau i}^{(t-1)} (y_i - \mathbf{x}_i^T \hat{\beta}_\tau^{(t-1)})^2 \right\}^{1/2}. \quad (18)$$

5. Repeat Steps 2–4 until convergence. Convergence is achieved when the difference between the estimated model parameters obtained from two successive iterations is less than a small pre-specified value. This algorithm is similar to that proposed by Street *et al.* (1988). In simulation experiments, the proposed algorithm usually converges, on average, after 10 iterations.

If $\rho_\tau(\cdot)$ is the Huber loss function defined in (2), we call (14) the asymmetric least informative (ALI) distribution. In this case, the normalizing constant B_τ is given by

$$B_\tau = \sqrt{\frac{\pi}{\tau}} \left[\Phi(c\sqrt{2\tau}) - 1/2 \right] + \sqrt{\frac{\pi}{1-\tau}} \left[\Phi(c\sqrt{2(1-\tau)}) - 1/2 \right] \\ + \frac{1}{2c\tau} \exp\{-c^2\tau\} + \frac{1}{2c(1-\tau)} \exp\{-c^2(1-\tau)\}. \quad (19)$$

This distribution is essentially a modified normal distribution with heavier tails (when $|y| > c$). For $\tau = 0.5$, this distribution was derived by Huber (1981, section 4.5) as the one minimizing the Fisher information in the ε -contaminated neighbourhood of the normal distribution. Formulae for the cumulative distribution function and moments of the ALI distribution ($\tau \in (0, 1)$) are provided in Appendix A.

The ALI distribution depends on the tuning constant c . Conventionally, in M-regression, the tuning constant is set by the data analyst in a way that provides a trade-off between robustness and efficiency. Huber (1981, p.18) suggests that ‘good choices are in the range between 1 and 2, say, 1.5’. The default value for c in the R package MASS (`rlm` function) is 1.345 which corresponds to 90% of efficiency of the estimates under normality. When the errors are normally distributed, the best choice is to set the tuning constant equal to a large value, that is, ∞ . Using a smaller value, say 1.345, in this case will offer unnecessary robustness at the cost of reduced efficiency of the estimates. To overcome this *ad hoc* approach to selecting the tuning constant, Wang *et al.* (2007) propose a data-driven method such that the tuning constant is numerically chosen in a way that maximises the asymptotic efficiency of the estimated parameters.

We can treat c as a parameter of the density f_τ and estimate this (together with β_τ and σ_τ) by maximising the log-likelihood function (15). In fact, in the ALI distribution, c can be interpreted as an additional parameter that determines how heavy the tails of the distribution are. For estimating the tuning constant, there is no closed form expression. In this case, the compass search algorithm or the Nelder–Mead algorithm (Griva *et al.*, 2008) can be used. The final estimating procedure works by adding to the proposed iterative algorithm the new step 4’:

- 4’ Given $\hat{\beta}_\tau^{(t)}$ and $\hat{\sigma}_\tau^{(t)}$ maximise the log-likelihood function (15) with respect to c using the compass search algorithm (Bottai *et al.*, 2015) or the Nelder–Mead algorithm.

An R function that implements an iterative algorithm for estimating the parameters is available from the authors.

The idea of estimating the tuning constant using likelihood equations can be applied to other loss functions as well whenever they include an additional parameter or tuning constant.

5 Goodness-of-Fit and Likelihood Ratio Type Tests in M-quantile Regression

In this section, we present, in an infinite population context, a pseudo- R^2 goodness-of-fit measure for MQ regression and likelihood ratio and Wald type tests for linear hypotheses on the regression parameters. The asymptotic theory is developed according to standard M-estimation theory, as in Gourieroux & Monfort (1989) and Wooldridge (2010).

5.1 A Goodness-of-Fit Measure

For a given MQ, the introduction of the pseudo- R^2 is motivated by the need for a measure analogous to the ordinary R^2 used in least squares regression, where the goodness-of-fit is expressed with respect to the null model under which all the coefficients except for the intercept are set to zero. Consider the general MQ model (4), where the first component of the

p -dimensional vector \mathbf{x} is the intercept, that is, $x_1 = 1$ and the MQ estimator $\hat{\beta}_\tau$ of β_τ . Now consider the null model, given by

$$MQ_\tau(y_i|\mathbf{x}_i) = \beta_{1\tau}, \quad (20)$$

and denote by $\tilde{\beta}_{1\tau}$ its MQ estimator. A relative goodness-of-fit measure comparing the full to the null MQ regression model is defined as

$$R_\rho^2(\tau) = 1 - \frac{\sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \hat{\beta}_\tau}{\hat{\sigma}_\tau} \right)}{\sum_{i=1}^n \rho_\tau \left(\frac{y_i - \tilde{\beta}_{1\tau}}{\hat{\sigma}_\tau} \right)}. \quad (21)$$

Because $\sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \hat{\beta}_\tau}{\hat{\sigma}_\tau} \right) \leq \sum_{i=1}^n \rho_\tau \left(\frac{y_i - \tilde{\beta}_{1\tau}}{\hat{\sigma}_\tau} \right)$, this measure is always between 0 and

1. $R_\rho^2(\tau)$ is a local relative measure of goodness-of-fit of the MQ regression model with respect to the null model at a specific τ . Because this goodness-of-fit statistic depends on τ , it is useful to study its variation across MQs. As in Koenker & Machado (1999), we explore the behaviour of the index $R_\rho^2(\tau)$ using a range of simulated data. We consider a simple bivariate regression setting under three different scenarios with $n = 100$:

- Gaussian noise: the data are generated with y_i iid standard normal distribution and independent of x . Values x_i are generated as iid $N(5, 1)$.
- Gaussian location shift: the data are generated according to the model

$$y_i = x_i + \epsilon_i$$

with ϵ_i iid $N(0, 1)$ and x_i iid $N(5, 1)$.

- Gaussian scale shift: a heteroskedastic version of the regression model is given by

$$y_i = \left(x_i + \frac{1}{4} x_i^2 \right) \epsilon_i$$

with ϵ_i iid $N(0, 1/100)$ and x_i iid $N(3, 1)$.

First row in Figure 1 illustrates the MQ model lines fitted at $\tau = (0.01, 0.10, 0.25, 0.50, 0.75, 0.90, \text{ and } 0.99)$. The bottom row of the figure shows the values of $R_\rho^2(\tau)$ as a function of τ . As we expected, under Gaussian noise, the values of $R_\rho^2(\tau)$ are nearly 0 over the entire range $\tau \in (0, 1)$. Under the Gaussian location shift scenario, the values of $R_\rho^2(\tau)$ show a flat relationship between y and x for each τ . This indicates that all the conditional MQs are equally successful in reducing variability (Koenker & Machado, 1999). In the case of heteroskedasticity (scenario 3), the conditional median and the unconditional median are both 0, so in line with what we expected under this data generation scenario, we have $R_\rho^2(0.5) \cong 0$. For the other values of τ , there is a clear benefit from the MQ form of the conditional MQ specification.

5.2 Hypothesis Testing

We start by partitioning MQ regression as follows:

$$MQ_\tau(y_i|\mathbf{x}_i) = \mathbf{x}_{i1}^T \beta_{1\tau} + \mathbf{x}_{i2}^T \beta_{2\tau}, \quad (22)$$

where $\beta_\tau = (\beta_{1\tau}^T, \beta_{2\tau}^T)^T$, $\beta_{1\tau}$ is a $(p - k) \times 1$ vector and $\beta_{2\tau}$ is a $k \times 1$ ($0 < k < p$) vector and \mathbf{x}_{i1} , \mathbf{x}_{i2} are defined accordingly. We are interested in testing the null hypothesis:

$$H_0 : \beta_{2\tau} = \mathbf{0}. \quad (23)$$

Let $\hat{\beta}_\tau$ denote the MQ estimator of the full model and let $\tilde{\beta}_\tau = (\tilde{\beta}_{1\tau}^T, \mathbf{0}^T)^T$ denote the MQ estimator under the null hypothesis (23). For testing the null hypothesis (23), we propose a likelihood ratio type test that is valid when the residuals follow a general distribution. Let

$$\hat{V}(\tau) = \sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \hat{\beta}_\tau}{\sigma_\tau} \right), \quad \tilde{V}(\tau) = \sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \tilde{\beta}_\tau}{\sigma_\tau} \right).$$

Consider the following regularity conditions:

- (C1) Θ compact set in \mathcal{R}^p ;
- (C2) ρ is (a.e.) twice continuously differentiable;
- (C3) $|\sup_{\beta_\tau \in \Theta} \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \beta_\tau}{\sigma_\tau} \right)| < h(\mathbf{x}_i, y_i)$ and $|\sup_{\beta_\tau \in \Theta} \psi'_\tau \left(\frac{y_i - \mathbf{x}_i^T \beta_\tau}{\sigma_\tau} \right) \mathbf{x}_i \mathbf{x}_i^T| < g(\mathbf{x}_i, y_i)$, with h and g are P -integrable functions;
- (C4) $E[\mathbf{x}_i \mathbf{x}_i^T \psi'((y_i - \mathbf{x}_i^T \beta_\tau)/\sigma_\tau)]$ is uniformly nonsingular for $\beta_\tau \in \Theta$.
- (C5) The errors $\varepsilon_{\tau i}$ are independent of \mathbf{x}_i .

Assumption (C3) guarantees the applicability of the Uniform Law of Large Numbers. In case of the Huber loss function, (C3) is satisfied provided $E|\mathbf{x}_i|^2 < +\infty$ and $E|y_i| < +\infty$. Assumption (C5) is required for the validity of the generalised information equality. This would hold also if the \mathbf{x}_i 's are fixed regressors. The information equality is needed for the validity of the likelihood ratio type test. It can be relaxed for the Wald type test. The following theorem presents the distribution of the likelihood ratio statistic.

Theorem 1. *Provided conditions (C1)–(C5) are satisfied, under the null hypothesis H_0*

$$-2 \frac{E \psi'_{\tau i}}{E \psi_{\tau i}^2} (\hat{V}(\tau) - \tilde{V}(\tau)) \xrightarrow{d} \chi_k^2, \quad (24)$$

where $\psi'_{\tau i} = \psi'_\tau(\varepsilon_{\tau i}/\sigma_\tau)$, $\psi_{\tau i} = \psi_\tau(\varepsilon_{\tau i}/\sigma_\tau)$.

The proof of Theorem 1 is reported in Appendix B. A hypothesis test for H_0 is obtained by substituting the unknown quantities in (24) with consistent estimators leading to,

$$-2 \frac{(n-p)^{-1} \sum_{i=1}^n \hat{\psi}'_{\tau i}}{n^{-1} \sum_{i=1}^n \hat{\psi}_{\tau i}^2} \left[\sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \hat{\beta}_\tau}{\hat{\sigma}_\tau} \right) - \sum_{i=1}^n \rho_\tau \left(\frac{y_i - \mathbf{x}_i^T \tilde{\beta}_\tau}{\hat{\sigma}_\tau} \right) \right], \quad (25)$$

where $\hat{\psi}'_{\tau i}$ and $\hat{\psi}_{\tau i}$ have been previously defined, and the nuisance parameter σ_τ is estimated under the full model. This is to ensure that the test statistic is nonnegative. Even though the asymptotic distribution of (25) is not exactly χ_k^2 , simulations show that χ_k^2 is still a good approximation for it (Section 7.2). This is due to the fact that the contribution of the estimation of σ_τ to the asymptotic variance is negligible, as it was noticed in Bianchi & Salvati (2015). The same approach was adopted in Schrader & Hettmansperger (1980). This test is more commonly known as likelihood ratio (LR)-type test because the density of the $\varepsilon_{\tau i}$ does not have to correspond to the loss function. Note also that the proposed test can be easily extended to test more

general linear hypotheses, for example, $H_0 : \mathbf{R}\boldsymbol{\beta}_\tau = \mathbf{r}$, where \mathbf{R} is a $k \times p$ full rank matrix, and \mathbf{r} is a $k \times 1$ vector. Similar results for M-regression estimators are provided by Schrader & Hettmansperger (1980) in the case of fixed regressors and for quantile regression with fixed regressors by Koenker & Machado (1999).

An alternative to the LR-type test is to use a Wald type test. The test statistic is derived from Theorem 2.1 in Bianchi & Salvati (2015). Let $\mathbf{R} = [\mathbf{0} : \mathbf{I}_k]$. It follows that under H_0

$$n(\mathbf{R}\hat{\boldsymbol{\beta}}_\tau)^T [\mathbf{R}\boldsymbol{\Sigma}_\tau \mathbf{R}]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}}_\tau) \xrightarrow{d} \chi_k^2,$$

where $\boldsymbol{\Sigma}_\tau$ is defined in (B4). Replacing $\boldsymbol{\Sigma}_\tau$ with its consistent estimator

$$\hat{\boldsymbol{\Sigma}}_\tau = \hat{\sigma}_\tau^2 \frac{(n-p)^{-1} \sum_{i=1}^n \hat{\psi}_{\tau i}^2}{n^{-1} \sum_{i=1}^n \hat{\psi}'_{\tau i}} \left[\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right]^{-1},$$

the statistic

$$W \equiv n(\mathbf{R}\hat{\boldsymbol{\beta}}_\tau)^T [\mathbf{R}\hat{\boldsymbol{\Sigma}}_\tau \mathbf{R}]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}}_\tau),$$

follows asymptotically a χ_k^2 distribution. A major difference between the LR-type test and the Wald type test is that the latter can be made robust to the presence of heteroskedasticity by using a robust estimator of the covariance matrix in place of $\hat{\boldsymbol{\Sigma}}_\tau$.

6 A Test to Assess the Presence of Area-Specific Effects

In this section, we introduce an LR-type test for the presence of unobserved heterogeneity (area-specific effects). The proposed test has a similar aim to that for the strict positiveness of variance components in the case of a linear mixed (random) effects model. Testing for the presence of significant cluster effects is a well-known problem in the literature (Greven *et al.*, 2008; Crainiceanu & Ruppert, 2004; Datta *et al.*, 2011). Clustering can exist either because of the design used to collect the data (i.e. use of a multi-stage cluster design) or because of the natural structures that exist in the population (e.g. pupils nested within schools or individuals nested within households). The discussion in this section will pay special attention to the existence of area-effects in SAE.

Our aim is to test for the presence of significant area/cluster effects by proposing a testing procedure based on the cluster-specific MQ coefficients τ_j . The development considers an infinite population.

We define the MQ coefficients $\boldsymbol{\tau} = (\tau_1, \dots, \tau_d)^T$ by adopting an approach that is explicitly based on the loss function. Within group j , τ_j is defined to be the one that uniquely solves

$$\min_{\tau} E \left[\rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_\tau}{\sigma} \right) | j \right].$$

Intuitively, τ_j is defined as the MQ for which the regression plane identified by $\boldsymbol{\beta}_{\tau_j}$ is the closest to observations from group j , according to the metrics of $\rho(\cdot)$. Note that $\rho(\cdot)$ is the untilted loss function, that is, $\rho_{0.5}(\cdot)$, so that the scale σ coincides with $\sigma_{0.5}$. The use of the untilted loss function is motivated by the search of the regression plane that best fits the units in a specific subgroup of the population. Testing for the presence of clustering is equivalent to testing whether the group-specific MQ coefficients are all equal, that is,

$$\begin{aligned} H_0 : \tau_j &= 0.5 \quad \forall j = 1, \dots, d \\ H_A : \tau_j &\neq 0.5 \text{ for at least one } j. \end{aligned}$$

Of course, $\tau = 0.5$ represents the global minimiser when considering all groups $j = 1, \dots, d$. A natural estimator $\hat{\tau}_j$ for τ_j is obtained by solving

$$\min_{\tau} \sum_{i=1}^{n_j} \rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{\tau}}{\hat{\sigma}} \right), \quad (26)$$

where $\hat{\sigma}$ is an estimator of σ such as the one obtained solving (16) for $\tau = 0.5$. Because ρ is a positive function, the problem may be rewritten as follows. The vector of estimated MQ coefficients $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \dots, \hat{\tau}_d)^T$ is obtained as the solution of

$$\min_{(\tau_1, \dots, \tau_d)} \sum_{j=1}^d \sum_{i=1}^{n_j} \rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{\tau_j}}{\hat{\sigma}} \right). \quad (27)$$

Note that Chambers & Tzavidis (2006) define τ_j in a different way: they define it as the average value of the MQ coefficients within the area, suggesting that alternative definitions (such as the median value) and consequently alternative estimators can be used. In this paper, we define τ_j and $\hat{\tau}_j$ as the solution to the minimisation problem defined in (26). Although different from the original definition used in Chambers & Tzavidis (2006), we expect that the area-specific coefficients under the two definitions will lead to similar solutions. The advantage of the definition we use in this paper is that it provides an easier approach to study the distribution of the test statistic we are interested in.

Assuming that conditions (C1)–(C5) are satisfied and that $\boldsymbol{\beta}_{\tau}$ is differentiable in τ with $\partial^2 \boldsymbol{\beta}_{\tau} / \partial \tau^2 = 0$ (i.e. $\boldsymbol{\beta}_{\tau}$ linear in τ), it may be shown that under H_0

$$-2 \frac{E \psi'_{ij}}{E \psi^2_{ij}} \left[\sum_{j=1}^d \sum_{i=1}^{n_j} \rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{\hat{\tau}_j}}{\sigma} \right) - \sum_{j=1}^d \sum_{i=1}^{n_j} \rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{0.5}}{\sigma} \right) \right] \xrightarrow{d} \chi^2_{d-1}, \quad (28)$$

where $\psi'_{ij} = \psi'(\varepsilon_{0.5ij}/\sigma)$, $\psi_{ij} = \psi(\varepsilon_{0.5ij}/\sigma)$ and $\varepsilon_{0.5ij} = (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{0.5})$ (for a sketch of the proof of (28), see Appendix C). By simulation (Section 7.3), we show that by substituting the unknown parameters in (28) by consistent estimators, the asymptotic distribution is still well approximated by a χ^2_{d-1} distribution. Hence, a hypothesis test may be based on

$$-2 \frac{(n-p)^{-1} \sum_{ij} \hat{\psi}'_{ij}}{n^{-1} \sum_{ij} \hat{\psi}^2_{ij}} \left[\sum_{j=1}^d \sum_{i=1}^{n_j} \rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{\hat{\tau}_j}}{\hat{\sigma}} \right) - \sum_{j=1}^d \sum_{i=1}^{n_j} \rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{0.5}}{\hat{\sigma}} \right) \right],$$

where $\hat{\psi}'_{ij} = \psi'(\hat{\varepsilon}_{0.5ij}/\hat{\sigma})$, $\hat{\psi}_{ij} = \psi(\hat{\varepsilon}_{0.5ij}/\hat{\sigma})$, $\hat{\varepsilon}_{0.5ij} = (y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{0.5})$ and $\boldsymbol{\beta}_{\hat{\tau}_j}$ and $\boldsymbol{\beta}_{0.5}$ are replaced by the corresponding consistent estimators.

The proposed test can assist the decision to include or not cluster effects in the model. We note that the asymptotic result holds if $n_j \rightarrow +\infty$ for each $j = 1, \dots, d$. Even though the test is asymptotically valid when the sample size within each group tends to infinity, we empirically show in Section 7 that it provides reasonable results in the SAE context as well.

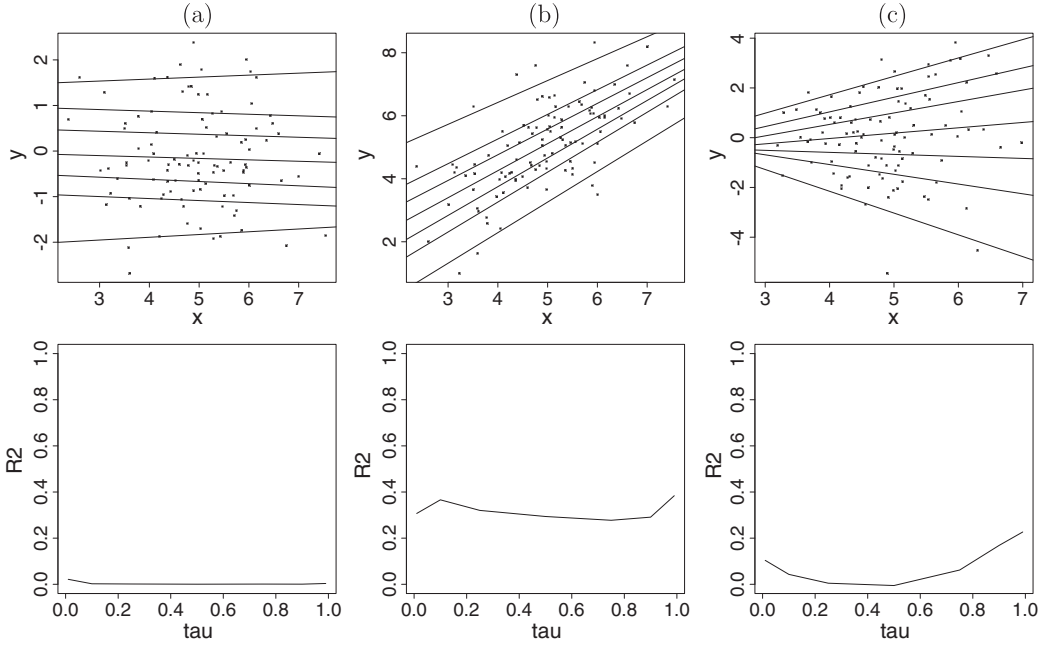


Figure 1. The figure shows three different scenarios and their associated $R^2_\rho(\tau)$: (a) Gaussian noise; (b) Gaussian location shift; (c) Gaussian scale shift. The top row presents the data and in solid font the M-quantile model lines fitted at $\tau = (0.01, 0.10, 0.25, 0.50, 0.75, 0.90, 0.99)$. The second row depicts the values of $R^2_\rho(\tau)$ at different M-quantiles.

The test we propose has a different aim to that of specification tests such as that recently proposed by Parente & Santos Silva (2013) as we are not testing the assumptions needed for the estimation of β_τ but whether units belonging to the same cluster are characterised by similar quantile coefficients, which is useful in prediction.

7 Simulation Studies

In this section, we present results from three simulation studies used to investigate the method for selecting the tuning constant c proposed in Section 4, the finite sample properties of the tests proposed in Section 5 and of the test statistic used for assessing the presence of area-specific effects proposed in Section 6. All these tools are related to model selection and checking. This is very important in SAE as its aim is to produce model-based estimates that practitioners are confident to use.

Although estimation of the tuning constant and LR and Wald type tests are useful tools for general MQ regression, we consider a typical small area setting with observations clustered within area for all simulations, in line with our focus on this type of application. Thus, in the simulation, we generate data under linear mixed (random) effects models that incorporate area specific variation.

7.1 Choosing the Tuning Constant

We consider data generated under the following mixed (random) effects model,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \varepsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, d, \quad (29)$$

where j indexes areas (clusters), and i indexes units within areas, $\beta_0 = 1$, $\beta_1 = 2$, x follows a uniform distribution $(0, 5)$, $d = 100$, $n_j = 5$ ($n = 500$). The error terms of the mixed model, u_i and ε_{ij} , are generated by using different parametric assumptions. The random effects u_i are generated from a Normal distribution with mean 0, and $\sigma_u^2 = 1$ and ε are drawn from different error distributions.

1. Gaussian with mean 0, variance 1: ‘regularly’ noisy data;
2. t -Student with 3 degrees of freedom (t_3): more noisy data with heavy tails;
3. Contaminated Normal with $\varepsilon \sim (1 - \gamma)N(0, 1) + \gamma N(0, 25)$ where γ is an independently generated Bernoulli random variable with $Pr(\gamma = 1) = 0.1$; and
4. Cauchy with location 0 and scale 1: noisy data with the likely presence of extreme and outlying observations.

These assumptions on the errors replicate the design in Sinha & Rao (2009), Giusti *et al.* (2014) and Salvati *et al.* (2010). We have chosen these settings, ranging from a situation of ‘regularly’ noisy data to situations of more noisy data with huge extreme values, for evaluating the estimation of the tuning constant c under the ALI distribution as proposed in Section 4. The residuals are rescaled so their variance is equal to 1, and the value of intracluster correlation under different scenarios is always approximately equal to 0.3. For each Monte Carlo sample, we estimate the tuning constant c under the ALI distribution as proposed in Section 4. Figure 2 shows the distribution, obtained with 10 000 Monte Carlo samples of the estimated tuning constants for the four scenarios at $\tau = 0.25, 0.5, 0.75$. The horizontal dashed line represents the usual choice of $c = 1.345$. Under the Gaussian setting, the values of the tuning constants are clearly larger than the value 1.345 at each τ . The estimated value of the tuning constant suggests that using a robust estimator in this case is not justified as one would expect under the assumptions of normality. In contrast, the values of the estimated tuning constant are smaller than 1.345 in the contaminated and Cauchy scenarios. For instance, in the case of the contaminated scenario, the median value of the estimated tuning constant at $\tau = 0.5$ is 0.794. In the case of the Cauchy scenario, the median value of the estimated tuning constant, at each quantile, degenerates to 0 because the Cauchy distribution has very heavy tails. For the t -student scenario, the median value of the estimated tuning constant is 1.27 at $\tau = 0.5$, and it becomes higher than 1.345 (about 2.0) at $\tau = 0.25, 0.75$.

In applications, a unique c should be chosen; it can be the optimal one at 0.5 or chosen taking into consideration also optimal values at other quantiles.

7.2 Likelihood Ratio Type Test

For evaluating the LR and Wald type tests for linear hypotheses on the MQ regression parameters, data are generated under the following mixed (random) effects model

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + u_i + \varepsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, d. \quad (30)$$

The regression coefficients are set as follows: $\beta_0 = 0$, $\beta_1 = 0.5$ and β_2, β_3 vary pairwise from 0 to 1, that is, $(\beta_2, \beta_3) = (0, 0)$, $(\beta_2, \beta_3) = (0.25, 0.25)$, $(\beta_2, \beta_3) = (0.5, 0.5)$ and $(\beta_2, \beta_3) = (1, 1)$. With respect to the choice of the values of the regression coefficients, we consider departures from the null hypothesis up to a level where the power of the test is expected to reach level 1. Equality of β_2 and β_3 is motivated only by simplicity, and it is in line with the simulations proposed by Koenker & Machado (1999). The values of x_1 , x_2 and x_3 are drawn from a Normal distribution with mean 5, 3 and 2, respectively, and variance equals to 1. The number of small areas is set equal to $d = 8, 20, 100$ and sample size in each small area $n_j = 5$, so we consider three different overall sample sizes: $n = 40, 100, 500$. We consider $n = 40$ for

evaluating the properties of the LR and the Wald type tests under a small sample size. Note that in the application (Section 8), the sample size is 37. The error terms of the mixed model, u_i and ε_{ij} , are generated by using different parametric assumptions. Three settings for generating ε_i are considered.

1. Gaussian with mean 0, variance 1;
2. t -Student distribution with 3 degrees of freedom (t_3); and
3. Chi-squared errors with 2 degrees of freedom (χ_2^2).

t -Student and chi-squared random variables are re-scaled so to have variance equal to 1; in the case of chi-squared, we subtract the mean to generate zero-mean residuals. The random effects are generated from a Normal distribution with mean 0 and $\sigma_u^2 = 0.43$. This entails

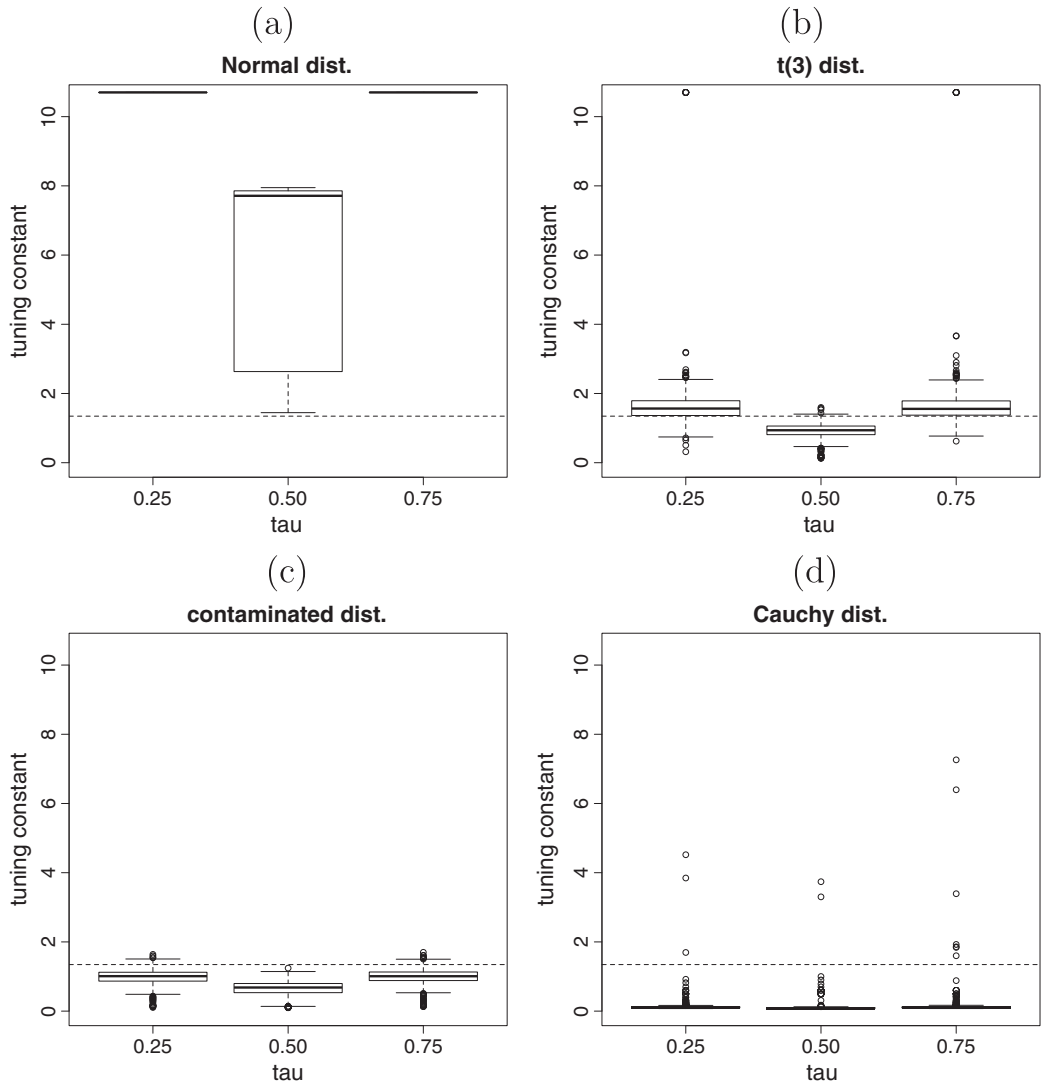


Figure 2. The distribution of the values of the tuning constant over Monte Carlo samples and different settings for the error distribution at $\tau = 0.25, 0.50, 0.75$ and $d = 100$: (a) Gaussian distribution; (b) t -student distribution; (c) Contaminated Normal distribution; (d) Cauchy distribution. The horizontal dashed line represents the choice of $c = 1.345$.

that for all the scenarios the value of the intraclass correlation is approximately equal to 0.3. These choices define a $4 \times 3 \times 3$ design of simulations. As in the previous section, we consider different settings ranging from a situation of regularly noisy data to situations of skewed distributed data. Each scenario is independently simulated $T = 10\,000$ times. MQ regression is fitted at $\tau = 0.5, 0.75, 0.90$ by using the Huber influence function with $c = 1.345$ for t -Student and chi-squared errors, $c = 100$ for Gaussian errors and the maximum likelihood estimator (18) based on the ALI distribution as the estimator of σ_τ . Setting c equal to 1.345 gives reasonably high efficiency under normality and protects against outliers when the Gaussian assumption is violated (Huber, 1981). For the Gaussian scenario, the resistance against outliers is not necessary, and a large value for the tuning constant is preferred.

The results for the LR-type test for the null hypothesis

$$H_0 : \beta_{2\tau} = \beta_{3\tau} = 0$$

at the significance level $\alpha = 0.10, 0.05, 0.01$ are presented in Table 1. In all cases, when $\beta_2 = \beta_3 = 0$ and the null hypothesis is true, the Type I error is very close to the nominal α , with deviations in the case of $\tau = 0.9$ in the t_3 and χ^2_2 scenarios with $d = 8$ and 20 ($n = 40, 100$) where the test turns out to be conservative. For the Gaussian scenario, the power of the test tends to 1 as soon as the values of β_2 and β_3 increase, that is, the null hypothesis is rejected for all sample sizes. In case of departures from normality, for example, under the t_3 scenario, the power of the test tends to 1 at $\tau = 0.5$ and 0.75 once $\beta_2, \beta_3 = 0.25$ especially for $d = 100$ ($n = 500$). At $\tau = 0.9$, the LR type test performs well as regression coefficients increase (as soon as $\beta_2, \beta_3 = 0.5$). Under the chi-squared setting, the test at $\tau = 0.75, 0.90$ appears to have lower power in rejecting the null hypothesis especially for the scenario with $d = 8$ and $d = 20$. Results for this scenario improve as the number of groups, d , and the values of the regression parameters (β_2, β_3) increase. In the interest of brevity, results for the Wald type test are not reported. They are available from the authors upon request. However, here, we provide a summary of the comparison between the Wald and LR-type tests. Under the Gaussian scenario, results from the two tests are very similar. Under the t_3 and χ^2_2 scenarios, convergence for the Wald type test is slower than convergence for the LR-type test under the null hypothesis, especially for extreme MQs. Of course, the two tests are equivalent when the sample size is large. In general, we have a slight preference towards the use of the LR-type test.

7.3 Testing for the Presence of Area-Specific Heterogeneity

In this section, we present an empirical evaluation of the properties of the test for the presence of area heterogeneity, and we show how this test can be useful in the SAE context. For these simulations, data are generated under model (29). We consider three scenarios for the number of groups d , $d = 8, d = 20$ and $d = 100$ and three scenarios for the within group samples size, $n_j = 5, n_j = 20$ and $n_j = 50$ with the overall sample size that varies from 40 to 500. The error terms of the mixed model, u_i and ε_{ij} , are generated by using different parametric assumptions. In particular, the random effects are generated from a Normal distribution with mean 0 and different scenarios for the level 2 variance components $\sigma_u^2 = 0, 1, 2.5, 7.5$. For $\sigma_u^2 = 0$, data are generated under the null hypothesis of no clustering. For the values of σ_u^2 other than 0, clustering is introduced in the simulated data. Individual effects are generated according to Normal distribution with mean 0 and variance 5. The intraclass correlation varies between 0 and 0.60. In general, in SAE applications, the observed intraclass correlation is about 0.30. When $\sigma_u^2 = 0$, that is, under the null hypothesis, we empirically study the Type I error by using the proposed test. For all other scenarios of $\sigma_u^2 \neq 0$, we study the power of the proposed test. Each scenario is independently simulated $T = 10\,000$ times.

Table 1. Type I error and power of the proposed likelihood ratio type test under Gaussian, t_3 and χ^2_2 distributions at $\tau = 0.50, 0.75, 0.90$ with β_2, β_3 varying pairwise from 0 to 1, $\alpha = 0.10, 0.05, 0.01$ and $d = 8, 20, 100$ with $n_j = 5$.

d	α	Gaussian, $c = 100$			$t_3, c = 1.345$			$\chi^2_2, c = 1.345$		
		$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
		$(\beta_2, \beta_3) = (0, 0)$								
8	0.10	0.117	0.132	0.181	0.119	0.143	0.322	0.120	0.159	0.355
	0.05	0.066	0.079	0.117	0.067	0.087	0.232	0.069	0.096	0.268
	0.01	0.018	0.021	0.044	0.017	0.028	0.128	0.017	0.029	0.150
	0.10	0.110	0.114	0.133	0.103	0.114	0.147	0.109	0.120	0.181
20	0.05	0.059	0.062	0.075	0.050	0.063	0.089	0.057	0.064	0.112
	0.01	0.012	0.015	0.021	0.012	0.016	0.030	0.012	0.016	0.049
	0.10	0.101	0.105	0.109	0.102	0.108	0.122	0.103	0.106	0.126
100	0.05	0.052	0.058	0.058	0.052	0.053	0.063	0.050	0.055	0.069
	0.01	0.010	0.011	0.012	0.013	0.012	0.017	0.010	0.011	0.018
	$(\beta_2, \beta_3) = (0.25, 0.25)$									
8	0.10	0.392	0.394	0.400	0.547	0.491	0.490	0.353	0.248	0.188
	0.05	0.283	0.282	0.301	0.430	0.380	0.401	0.251	0.166	0.102
	0.01	0.130	0.135	0.159	0.229	0.205	0.156	0.104	0.068	0.073
20	0.10	0.574	0.547	0.481	0.681	0.605	0.457	0.497	0.313	0.273
	0.05	0.453	0.430	0.371	0.566	0.488	0.357	0.375	0.215	0.191
	0.01	0.245	0.225	0.192	0.337	0.267	0.191	0.184	0.088	0.082
100	0.10	1.000	0.999	0.964	1.000	0.996	0.909	0.984	0.823	0.395
	0.05	0.991	0.998	0.934	0.998	0.991	0.846	0.967	0.728	0.282
	0.01	0.962	0.989	0.914	0.991	0.968	0.671	0.903	0.498	0.128
$(\beta_2, \beta_3) = (0.50, 0.50)$										
8	0.10	0.849	0.827	0.774	0.952	0.900	0.761	0.779	0.498	0.437
	0.05	0.776	0.746	0.694	0.919	0.846	0.692	0.689	0.391	0.397
	0.01	0.580	0.554	0.516	0.807	0.702	0.546	0.485	0.212	0.200
	0.10	0.978	0.962	0.920	0.993	0.982	0.852	0.944	0.729	0.449
20	0.05	0.960	0.941	0.873	0.987	0.961	0.784	0.905	0.619	0.352
	0.01	0.883	0.841	0.729	0.953	0.890	0.619	0.774	0.400	0.196
	0.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.872
100	0.05	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.795
	0.01	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.590
	$(\beta_2, \beta_3) = (1, 1)$									
8	0.10	1.000	1.000	0.995	1.000	0.998	0.973	0.994	0.906	0.731
	0.05	1.000	1.000	0.991	1.000	0.997	0.958	0.990	0.854	0.657
	0.01	0.995	0.991	0.968	0.998	0.991	0.916	0.966	0.708	0.501
	0.10	1.000	1.000	1.000	1.000	1.000	0.998	1.000	0.996	0.841
20	0.05	1.000	1.000	1.000	1.000	1.000	0.996	1.000	0.990	0.767
	0.01	1.000	1.000	1.000	1.000	1.000	0.985	1.000	0.965	0.604
	0.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	0.05	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.01	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

In this Monte Carlo simulation, MQ regression is fitted by using the Huber influence function with $c = 100$ and the maximum likelihood estimator for the scale (18) under the ALI distribution. The use of a large value for the tuning constant is justified by the normality of the simulated data. Table 2 reports the results of the simulation experiment. The table shows the values of the intracluster correlation, $r = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2)$, the Type I error and the power of the proposed test statistic for $\alpha = 0.01, 0.05, 0.10$. To start with, we note that under the null hypothesis, the Type I error is very close to the nominal value of α . As the value of σ_u^2 increases, the power of the test increases too. The power increases more sharply for larger within cluster sample sizes. The number of clusters also seems to impact on the power of the test. The power

Table 2. Type I error and power of the proposed test statistic for clustering under Gaussian distribution with r varying between 0 and 0.6, $\alpha = 0.10, 0.05, 0.01$, $d = 8, 20, 100$ and $n_j = 5, 20, 50$.

α	$d = 8$			$d = 20$			$d = 100$		
	$n_j = 5$	$n_j = 20$	$n_j = 50$	$n_j = 5$	$n_j = 20$	$n_j = 50$	$n_j = 5$	$n_j = 20$	$n_j = 50$
$r = 0$									
0.10	0.114	0.085	0.103	0.141	0.104	0.099	0.120	0.089	0.103
0.05	0.062	0.035	0.048	0.075	0.059	0.047	0.060	0.036	0.042
0.01	0.008	0.007	0.014	0.015	0.012	0.008	0.018	0.009	0.009
$r = 0.16$									
0.10	0.413	0.910	0.991	0.702	0.999	1.000	0.983	1.000	1.000
0.05	0.213	0.875	0.985	0.565	0.998	1.000	0.969	1.000	1.000
0.01	0.118	0.765	0.971	0.325	0.992	1.000	0.906	1.000	1.000
$r = 0.33$									
0.10	0.707	0.983	0.998	0.954	1.000	1.000	1.000	1.000	1.000
0.05	0.572	0.981	0.998	0.904	1.000	1.000	1.000	1.000	1.000
0.01	0.330	0.955	0.995	0.763	1.000	1.000	1.000	1.000	1.000
$r = 0.60$									
0.10	0.933	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
0.05	0.881	0.999	1.000	0.998	1.000	1.000	1.000	1.000	1.000
0.01	0.720	0.995	1.000	0.989	1.000	1.000	1.000	1.000	1.000

of the test increases fairly sharply when we have a larger number of clusters even if each cluster consists of a small number of units.

Under the null hypothesis, we have also computed the empirical expected value and variance of the test statistic. We expect that, under the χ^2_{d-1} asymptotic approximation, the expected value of the test statistic will be equal to $d - 1$ and the variance equal to $2 \times (d - 1)$. These expectations are confirmed by the simulation results.

Finally, we have run a simulation where the individual effects are generated according to t -student with 3 degrees of freedom and the MQ regression is fitted by using the Huber influence function with $c = 1.345$. Also, in this case, under the null hypothesis, the Type I error is very close to the nominal value of α , and the power of the test increases as the value of σ_u^2 increases. The detailed results are available to the interested reader from the authors.

The test can be used in the SAE framework to detect the presence of area effects. If the test rejects H_0 , it means that there is unobserved heterogeneity between areas and predictor (11) can be used to estimate the small area mean. Otherwise, if H_0 is not rejected, the synthetic estimator is preferred for predicting the small area quantity, because, in the case of absence of unobserved heterogeneity between areas, it guarantees less variability and bias than estimator (11). To evaluate the performance of the synthetic predictor and the MQ predictor (11), the absolute relative bias (ARB) and the relative root mean squared error (RRMSE) of estimates of the mean value in each small area are computed. Table 3 reports the average values over areas of these indices for $n_j = 5, 20, 50$ and $d = 100$. The results for $d = 8$ and $d = 20$ are not reported because these are very similar to those for $d = 100$ but are available from the authors upon request. Table 3 shows that the average ARB and RRMSE of the synthetic predictor increase as the intracluster correlation increases. The average values of ARB and RRMSE for estimator (11) remain constant at different values of r given the sample size. From the results in Table 3, it is apparent that when the assumption of significant between-area heterogeneity is not rejected, the synthetic estimator offers the best performance. On the other hand, as soon as the intracluster correlation increases, the predictor (11) performs best. Thus, the LR-type test for the presence of clustering can drive the choice of the MQ predictor in SAE. The increase in the RRMSE when unnecessarily incorporating the area effect into prediction

has been documented by other authors (Datta *et al.*, 2011). Our work extends these results to the case of SAE based on MQ regression.

8 Application

In this section, we use a dataset well-known in the SAE literature for illustrating how the proposed model fit, selection and diagnostic criteria work in a finite population context. Battese *et al.* (1988) analyse survey and satellite data for corn and soybean production for 12 counties in North Central Iowa. The dataset comes from the June 1978 Enumerative Survey, it consists of 37 observations and it includes information on the number of segments in each county, the number of hectares of corn and soybeans for each sample segment, the number of pixels classified by the LANDSAT satellite as corn and soybeans for each sample segment and the mean number of pixels per segment in each county classified as corn and soybeans. These data were used by Battese *et al.* (1988) to predict the hectares of corn and soybean by county. We use this dataset to compute the tuning constant c (Huber loss function is going to be adopted) and the R^2 goodness-of-fit measure and to perform the LR-type test for specifying the explanatory variables to be included in MQ regression. Note that county-specific random effects were introduced by Battese *et al.* (1988) to improve prediction, so we apply the LR-type test proposed in Section 6 to test whether there is significant between-county variation in the MQ coefficients, something that would justify the inclusion of county-specific MQs.

The response variables are y_1 , the number of hectares of corn, and y_2 , the number of hectares of soybeans. The models for the two variables are independent and include two fixed effects, x_1 and x_2 that represent the number of pixels classified by the LANDSAT satellite as corn and soybeans, respectively, for each sample segment. Battese *et al.* (1988) use the following two-level linear mixed model where i denotes the counties and j denotes the segments.

$$y_{hij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_i + e_{hij}, \quad h = 1, 2.$$

A random effect u_i is specified at the county level. This model will be used for benchmarking our results. Diagnostic for this model is reported in other papers (e.g. Sinha & Rao, 2009).

Table 3. Values of the average absolute relative bias (ARB) and average relative root mean squared error (RRMSE) over small areas for synthetic and (11) predictors under Gaussian distribution with r varying between 0 and 0.6, $d = 100$ and $n_j = 5, 20, 50$. Values are expressed as percentages.

Predictor	$n_j = 5$		$n_j = 20$		$n_j = 50$	
	ARB	RRMSE	ARB	RRMSE	ARB	RRMSE
$r = 0$						
\hat{m}_j^{MQ}	11.07	13.62	5.66	7.04	3.55	4.45
$\hat{m}_j^{MQ/SYN}$	1.39	1.74	0.99	1.24	0.86	1.08
$r = 0.16$						
\hat{m}_j^{MQ}	10.63	13.25	5.44	6.82	3.45	4.33
$\hat{m}_j^{MQ/SYN}$	11.41	14.29	11.20	14.02	10.84	13.58
$r = 0.33$						
\hat{m}_j^{MQ}	10.54	13.20	5.60	7.10	3.73	4.87
$\hat{m}_j^{MQ/SYN}$	17.96	22.50	17.67	22.13	17.12	21.44
$r = 0.60$						
\hat{m}_j^{MQ}	11.71	15.10	7.17	10.40	5.46	8.91
$\hat{m}_j^{MQ/SYN}$	31.07	38.92	30.59	38.31	29.65	37.13

They indicate that for the soybean model normality of u and e approximately holds. This is confirmed by a Shapiro–Wilk normality test, which does not reject the null hypothesis that the residuals follow a normal distribution (p -values: level 1 = 0.8583, level 2 = 0.2929). For the corn variable, on the other hand, there is an influential outlier in the Hardin county. Despite this, the null hypothesis of the Shapiro–Wilk test is not rejected (p -values: level 1 = 0.9987, level 2 = 0.1704).

We present results for MQ regression at $\tau = 0.05, 0.10, 0.25, 0.5, 0.75, 0.90, 0.95$. We further compare our results at $\tau = 0.5$ to model diagnostics from the linear mixed model used by Battese *et al.* (1988). For the analysis of the corn outcome, the estimate of the tuning constant c using the GALI pseudo-likelihood at $\tau = 0.5$ is equal to 1.94, a relatively low value, consistent with the presence of the outlier identified in diagnostic analysis. For the soybean variable, the tuning constant c estimate at $\tau = 0.5$ is 7.85. This value suggests that there are no issues with contamination. Using $c = 1.345$ or the value we chose for corn in this case would increase the robustness unnecessarily at the cost of lower efficiency. Similar conclusions hold for other values of τ .

Estimates of the scale parameter σ_τ obtained with the GALI-based method are shown in Figure 3. We note that these are sensitive to the MQ being considered and exhibit an inverted u-shape: for quantiles far from 0.5, the proportion of residuals for which $|u| > c$ is larger, and this proportion reduces their average size. For τ close to 0.5, estimates are close to those obtained using the MAD estimator (9). On the contrary, MAD estimates are larger for quantiles far from 0.5 compared with those obtained in the central part of the distribution. This can be due to the fact that the scaling constant q in (9) should be quantile-adjusted. Looking at the R^2 model fit criterion, the R^2 increases as τ increases for the corn outcome (Figure 3(b) solid line). For the soybean outcome, there appears to be an almost constant high value of R^2 at all values of τ (Figure 3 (b) dashed line). Overall, for both outcomes, there appears to be a moderate to strong linear relationship between the outcome and the explanatory variables at the different values of τ .

The LR-type tests results for the corn outcome are presented in Table 4 and for the soybean outcome in Table 5. The use of the LR-type test is justified by the simulation results obtained in Section 7.2, according to which this type of test should be preferred to the Wald type test in case of normality, when limited sample size is available and inference has to be made on extreme quantiles. When testing jointly the significance of x_1 and x_2 , the tests suggest that these covariates are significant for explaining the variability in both outcomes. For the corn outcome, the tests show that after controlling for the number of pixels classified by the LANDSAT satellite as corn (x_1), the number of pixels classified by the LANDSAT satellite as soybean (x_2) is not significant. Similarly, for the soybean outcome, after controlling for the number of pixels classified by the LANDSAT satellite as soybean (x_2), the number of pixels classified by the LANDSAT satellite as corn (x_1) is not significant. Hence, the model specification can be simplified by dropping non-significant terms. The same conclusions can be obtained by using the Wald type test. For validating these results at $\tau = 0.5$, we run the same analysis under the two-level linear mixed model used by Battese *et al.* (1988). For the corn outcome, after controlling for x_1 , the p -value for including x_2 is equal to 0.6315 indicating that x_2 can be dropped from the model. For the corn outcome, after controlling for x_2 , the p -value for including x_1 is equal to 0.6049 indicating that x_1 can be dropped from the model.

We turn our attention to testing the significance of the between-county variability. The two scatter plots in Figure 4 show the relationship between the predicted county random effects computed with the mixed model and the MQ county coefficients computed with the MQ model for the corn outcome (scatter plot (a)) and the soybean outcome (scatter plot (b)). For both outcomes, the two measures of county effects are well-correlated. For testing the significance

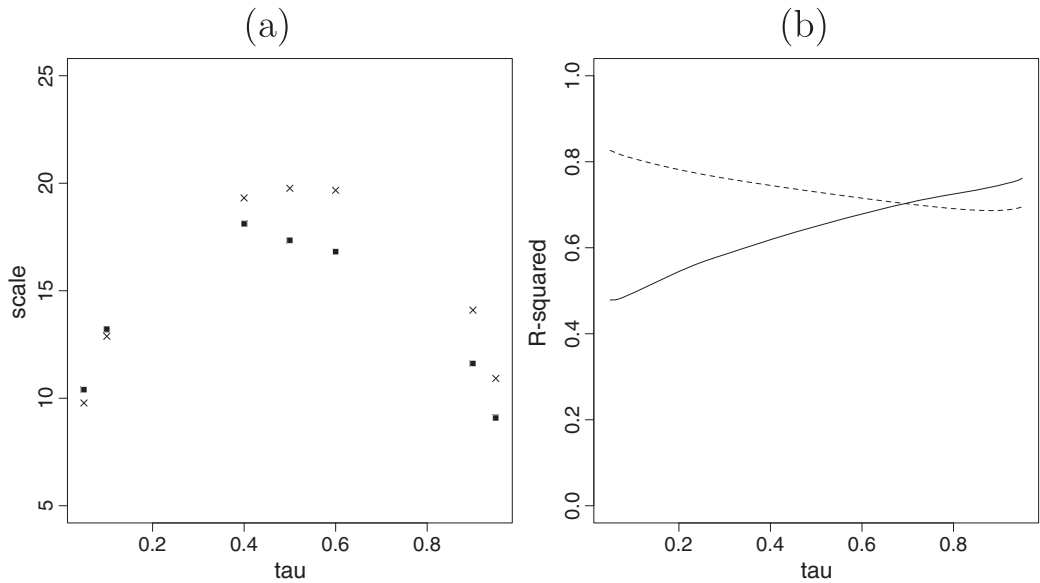


Figure 3. Plot (a) shows the values of the estimated scale at different value of τ for corn (■) and soybean (×). Plot (b) presents the R-squared, as defined in (21), at different value of τ for corn (solid line) and soybean (dashed line).

Table 4. Likelihood ratio (LR)-type test for the model specification of the corn outcome, $H_0 : (\beta_1, \beta_2) = 0$ and $H_0 : \beta_2 = 0$.

τ	$H_0 : (\beta_1, \beta_2) = 0$		$H_0 : \beta_2 = 0$	
	LR test	p-value	LR test	p-value
0.05	21.4	0.000	1.4	0.4935
0.10	23.8	0.000	0.3	0.8350
0.25	38.4	0.000	0.0	0.9996
0.50	68.3	0.000	0.4	0.7855
0.75	105.1	0.000	0.6	0.7376
0.90	97.1	0.000	0.1	0.9534
0.95	65.8	0.000	0.0	0.9959

Table 5. Likelihood ratio (LR)-type test for the model specification of the soybean outcome, $H_0 : (\beta_1, \beta_2) = 0$ and $H_0 : \beta_1 = 0$.

τ	$H_0 : (\beta_1, \beta_2) = 0$		$H_0 : \beta_1 = 0$	
	LR test	p-value	LR Test	p-value
0.05	195.7	0.000	2.6	0.2696
0.10	146.6	0.000	1.2	0.5496
0.25	116.0	0.000	0.3	0.8557
0.50	91.8	0.000	0.0	0.9972
0.75	66.7	0.000	0.4	0.8129
0.90	61.9	0.000	1.2	0.5380
0.95	65.3	0.000	01.6	0.4532

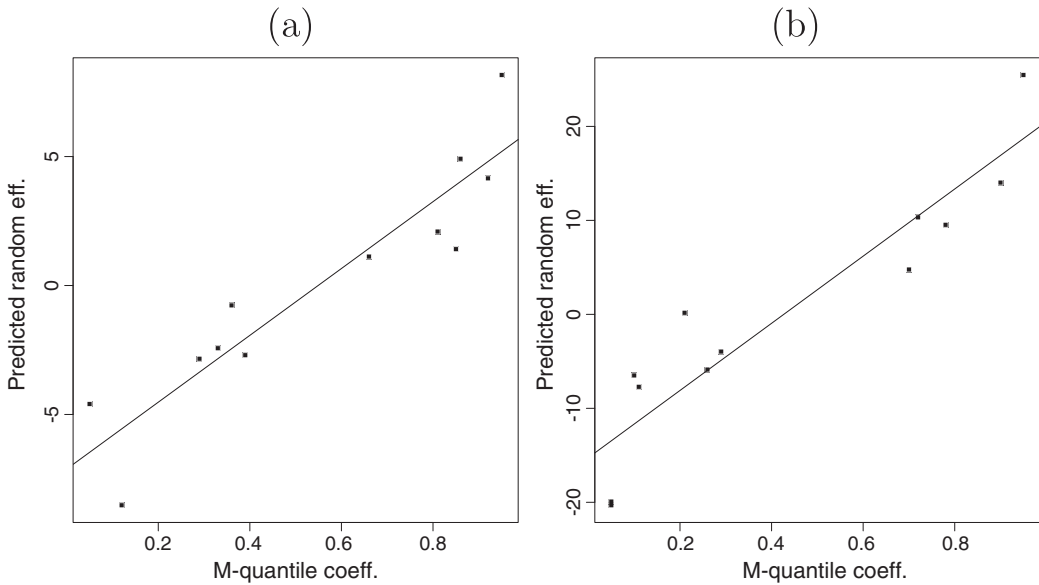


Figure 4. Scatter plots for the relationship between the predicted county random effects (computed with the mixed model) and the M-quantile county coefficients (computed with the M-quantile model) for the corn outcome (a) and for the soybean outcome (b).

of the county MQ coefficients, we use the proposed LR-type test. For the corn outcome, the value of the test statistic is 17.152, and the corresponding p -value = 0.103. For comparison purposes, we have also conducted the hypothesis test for the presence of significant between-county variation by using the linear mixed model. Under the null hypothesis $\sigma_u^2 = 0$, the test statistic has an asymptotic distribution which is an equal mixture of a point mass at zero and a χ^2 -distribution with 1 degree of freedom, denoted $1/2\chi_0^2 + 1/2\chi_1^2$ (Self & Liang, 1987). This type of test leads to a p -value equal to 0.5 thus suggesting that there is no presence of between-county variation in agreement with the result of the test based on MQ. Alternatively, for testing the null hypothesis of a zero between-county variation, we could compute the conditional-Akaike Information Criterion (cAIC) value (Vaida & Blanchard, 2005) and compare this to the AIC value for a linear regression model without random effects. The cAIC for the linear mixed model is 327.5109, and the AIC for the linear regression model is 327.4116. This indicates that the linear model without random effects fits almost and the more complex model that includes random effects. Hence, random effects may not be needed in the analysis of the corn outcome.

For the soybean outcome, the value of the LR-type test based on MQ coefficients for the presence of clustering is 26.791, and the corresponding p -value = 0.0049. As in the case of the corn outcome, we have also conducted the hypothesis test for the presence of significant between-county variation by using the linear mixed model. Under the null hypothesis $\sigma_u^2 = 0$ and using the Self & Liang (1987) testing procedure, the p -value of the LR-type test is 0.00127. The cAIC for the linear mixed model is 311.8459, and the AIC for the linear regression model is 333.8107. Overall, these results indicate that the linear model with county random effects fits better than the simpler model that ignores the random effects.

9 Final Remarks

In this paper, we review the MQ regression model and its application to SAE. We also extend the available toolkit for inference in MQ regression. For a given τ , we propose a pseudo- R^2

goodness-of-fit measure and LR and Wald type tests for testing linear hypotheses on the MQ regression parameters.

The cluster-specific MQ coefficients are used for proposing a new test for the presence of clustering in the data. The set of tests we present in the paper is useful in the SAE framework to validate the MQ models used for prediction. For a large class of continuously differentiable convex loss functions, we show the relationship between the loss function used in MQ regression and the maximisation of a likelihood function formed by combining independently distributed GALI densities. Using this parametrisation, we further propose an estimator of the scale parameter and a data-driven tuning constant to be used in the loss function. For each test, the asymptotic theory is developed, according to Gourieroux & Monfort (1989) and involving recent work on inference by Bianchi & Salvati (2015).

The simulation results for studying the finite sample properties of the model-fit criteria and the tests show that the Type I error of the LR-type test and the clustering test is very close to the nominal level α . For both tests, the results also indicate that the power tends to 1 as the values of the regression coefficients and the intraclass correlation coefficient increase. In the simulation experiments, we also investigate the behaviour of the method proposed for estimating the tuning constant in the Huber loss function. The tuning constant derived by using the likelihood method is able to reflect different levels of contamination in the data.

Despite the wide range of topics covered in this paper, some research problems remain open. For example, our proposed toolkit assumes the sampling design to be non-informative. Developing the theory for model selection and diagnostics under informative sampling remains an open research problem.

Moreover, the parametric representation of MQ regression developed in Section 4 could be used for proposing an AIC or a Bayesian Information Criterion for model selection. Finally, the pseudo- R^2 proposed in Section 5.1 could be an alternative method for estimating the optimal tuning constant: this goodness-of-fit measure can be computed with different values of the tuning constant to select the one that best fits the data. Both proposals are beyond the scope of this paper, however, and are left for further research.

Acknowledgements

The authors are indebted to the Editor, Associate Editor and referees for comments that significantly improved the paper. Tzavidis gratefully acknowledges support by grant ES/N011619/1 - Innovations in Small Area Estimation Methodologies from the UK Economic and Social Research Council. Salvati and Tzavidis gratefully acknowledge support by grant 730998 InGRID-2 Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy (Horizon 2020 research and innovation programme). Bianchi acknowledges financial support by the ex 60% University of Bergamo, Bianchi grant.

References

- Alfö, M., Salvati, N. & Ranalli, M. G. (2017). Finite mixtures of quantile and M-quantile regression models. *Stat. Comput.*, **27**, 547–570.
- Aragon, Y., Casanova, S., Chambers, R. & Leconte, E. (2005). Conditional ordering using nonparametric expectiles. *J. Off. Stat.*, **21**, 617–633.
- Battese, G., Harter, R. & Fuller, W. (1988). An error component model for prediction of county crop areas using survey and satellite data. *J. Am. Stat. Assoc.*, **83**, 28–36.
- Bianchi, A. & Salvati, S. (2015). Asymptotic properties and variance estimators of the M-quantile regression coefficients estimators, to appear in. *Commun. Stat. - Theory Methods*, **44**, 2016–2429.
- Bottai, M., Orsini, N. & Geraci, M. (2015). A gradient search maximization algorithm for the asymmetric Laplace likelihood. *Commun. Stat. - Comput. Simul.*, **85**, 1919–1925.

- Breckling, J. & Chambers, R. (1988). M -quantiles. *Biometrika*, **75**, 761–771.
- Cantoni, E. & Ronchetti, E. (2001). Robust inference for generalised linear models. *J. Am. Stat. Assoc.*, **96**, 1022–1030.
- Chambers, R. (1986). Outlier robust finite population estimation. *J. Am. Stat. Assoc.*, **81**, 1063–1069.
- Chambers, R. & Tzavidis, N. (2006). M -quantile models for small area estimation. *Biometrika*, **93**, 255–268.
- Chambers, R., Chandra, H. & Tzavidis, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Surv. Methodol.*, **37**, 153–170.
- Chambers, R., Chandra, H., Salvati, N. & Tzavidis, N. (2014a). Outlier robust small area estimation. *J. R. Stat. Soc., Ser. B*, **76**, 47–69.
- Chambers, R., Dreassi, E. & Salvati, N. (2014b). Disease mapping via negative binomial regression M -quantiles. *Stat. Med.*, **33**, 4805–4824.
- Chambers, R., Salvati, N. & Tzavidis, N. (2016). Semiparametric small area estimation for binary outcomes with application to unemployment estimation for local authorities in the UK. *J. R. Stat. Soc., Ser. A*, **179**, 453–479.
- Crainiceanu, C. M. & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance components. *J. R. Stat. Soc., Ser. B*, **66**, 165–185.
- Datta, G. S., Hall, P. & Mandal, A. (2011). Model selection by testing for the presence of small-area effects, and application to area level data. *J. Am. Stat. Assoc.*, **106**, 362–374.
- Dodge, Y. & Jureckova, J. (2000). *Adaptive Regression*. New York: Springer.
- Dreassi, E., Ranalli, M. G. & Salvati, N. (2014). Semiparametric M -quantile regression for count data. *Stat. Methods Med. Res.*, **23**, 591–610.
- Fabrizi, E., Salvati, N. & Pratesi, M. (2012). Constrained small area estimators based on M -quantile methods. *J. Off. Stat.*, **28**, 89–106.
- Fabrizi, E., Salvati, N., Pratesi, M. & Tzavidis, N. (2014a). Outlier robust model-assisted small area estimation. *Biom. J.*, **56**, 157–175.
- Fabrizi, E., Salvati, N., Giusti, C. & Tzavidis, N. (2014b). Mapping average equivalized income using robust small area methods. *Pap. Reg. Sci.*, **93**, 685–702.
- Falk, M. (1997). On mad and comedians. *Ann. Inst. Stat. Math.*, **49**, 615–644.
- Giusti, C., Tzavidis, N., Pratesi, M. & Salvati, N. (2014). Resistance to outliers of M -quantile and robust random effects small area models. *Commun. Stat.-Simul. Comput.*, **43**, 549–568.
- Gourieroux, C. & Monfort, A. (1989). *Statistics and Econometric Models* Cambridge, UK: Cambridge University Press.
- Greven, S., Crainiceanu, C. M., Kuechenhoff, H. & Peters, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *J. Comput. Graphical Stat.*, **17**, 870–891.
- Griva, I., Nash, S. G. & Sofer, A. (2008). *Linear and Nonlinear Optimization*, 2nd ed. Philadelphia: SIAM.
- Hall, P. & Welsh, A. H. (1985). Limit theorems for the median deviation. *Ann. Inst. Stat. Math.*, **37**, 27–36.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**, 73–101.
- Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Stat.*, **1**, 799–821.
- Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley & Sons.
- Huber, P. J. & Ronchetti, E. M. (2009). *Robust Statistics*. New York: John Wiley & Sons.
- Koenker, R. (2005). *Quantile Regression, Economic Society Monographs*. New York: Cambridge University press.
- Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker, R. & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *J. Am. Stat. Assoc.*, **94**, 1296–1310.
- Newey, W. K. & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.
- Parente, P. M. D. C. & Santos Silva, J. M. C. (2013). Quantile regression with clustered data. Paper number 13/05, Economics Department Discussion Papers Series.
- Pratesi, M., Ranalli, M. G. & Salvati, N. (2008). Semiparametric M -quantile regression for estimating the proportion of acidic lakes in 8-digit HUCs of the northeastern US. *Environmetrics*, **19**, 687–701.
- Pratesi, M., Ranalli, M. G. & Salvati, N. (2009). Nonparametric M -quantile regression using penalised splines. *J. Nonparametric Stat.*, **21**, 287–304.
- Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation*. New Jersey, US: Wiley & Sons.
- Salvati, N., Chandra, H., Ranalli, M. G. & Chambers, R. (2010). Small area estimation using a nonparametric model-based direct estimator. *Comput. Stat. Data Anal.*, **54**, 2159–2171.
- Salvati, N., Ranalli, M. G. & Pratesi, M. (2011). Small area estimation of the mean using non-parametric M -quantile regression: A comparison when a linear mixed model does not hold. *J. Stat. Comput. Simul.*, **81**, 945–964.
- Salvati, N., Tzavidis, N., Pratesi, M. & Chambers, R. (2012). Small area estimation via M -quantile geographically weighted regression. *TEST*, **21**, 1–28.
- Schrader, R. M. & Hettmansperger, T. P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika*, **67**, 93–101.

- Self, S. G. & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.*, **82**, 605–610.
- Sinha, S. K. & Rao, J. N. K. (2009). Robust small area estimation. *Can. J. Stat.*, **37**, 381–399.
- Street, J., Carroll, R. & Ruppert, D. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *Am. Stat.*, **42**, 152–154.
- Tzavidis, N., Salvati, N., Pratesi, M. & Chambers, R. (2008). M-quantile models with application to poverty mapping. *Stat. Methods Appl.*, **17**, 393–411.
- Tzavidis, N., Marchetti, S. & Chambers, R. (2010). Robust estimation of small-area means and quantiles. *Aust. N. Z. J. Stat.*, **52**, 167–186.
- Tzavidis, N., Salvati, N., Chambers, R. & Chandra, H. (2012). Small area estimation in practice: An application to agricultural business survey data. *J. Indian Soc. Agric. Stat.*, **66**, 213–238.
- Tzavidis, N., Ranalli, M. G., Salvati, N., Dreassi, E. & Chambers, R. (2015). Robust small area prediction for counts. *Stat. Methods Med. Res.*, **24**, 373–395.
- Tzavidis, N., Salvati, N., Schmid, T., Flouri, E. & Midouhas, E. (2016). Longitudinal analysis of the strengths and difficulties questionnaire scores of the Millennium Cohort Study children in England using M-quantile random-effects regression. *J. R. Stat. Soc., Ser. A*, **179**, 427–452.
- Vaida, F. & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**, 351–370.
- Wang, Y., Lin, X., Zhu, M. & Bai, Z. (2007). Robust estimation using the Huber function with a data-dependent tuning constant. *J. Comput. Graphical Stat.*, **16**, 468–481.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge (Mass.): The MIT Press.
- Yang, Y., Wang, H. J. & He, X. (2015). Posterior inference in bayesian quantile regression with asymmetric laplace likelihood. *Int. Stat. Rev.*, **84**, 327–344.
- Yu, K. & Moyeed, R. A. (2001). Bayesian quantile regression. *Stat. Probab. Lett.*, **54**, 437–447.
- Yu, K. & Zhang, J. (2005). A three-parameters asymmetric Laplace distribution and its extension. *Commun. Stat.: Theory Methods*, **34**, 1867–1879.

[Received May 2017, accepted March 2018]

Appendix A: Properties of the Asymmetric Least Informative

In this appendix, we provide some more properties for special case of the GALI distribution when the $\rho(\cdot)$ is given by (2), that is the ALI we introduced in Section 4. Suppose that U is a random variable with the standard ALI density ($\mu_\tau = 0, \sigma_\tau = 1$), then its cumulative distribution function is written as

$$F(u) = \begin{cases} \frac{1}{2c(1-\tau)B_\tau} \exp\{[2cu + c^2](1-\tau)\} & u \leq -c \\ \frac{1}{B_\tau} \left\{ \frac{1}{2c(1-\tau)} e^{-c^2(1-\tau)} + \sqrt{\frac{\pi}{1-\tau}} \left[\Phi(u\sqrt{2(1-\tau)}) - \Phi(-c\sqrt{2(1-\tau)}) \right] \right\} & -c < u \leq 0 \\ \frac{1}{B_\tau} \left\{ \frac{1}{2c(1-\tau)} e^{-c^2(1-\tau)} + \sqrt{\frac{\pi}{1-\tau}} \left[\Phi(c\sqrt{2(1-\tau)}) - 1/2 \right] + \sqrt{\frac{\pi}{\tau}} \left[\Phi(u\sqrt{2\tau}) - 1/2 \right] \right\} & 0 < u \leq c \\ \frac{1}{B_\tau} \left\{ \frac{1}{2c\tau} e^{-c^2\tau} - \frac{1}{2c\tau} \exp\{-2\tau cu + c^2\tau\} \right\} & u > c \end{cases}.$$

For obtaining the expected value and the variance of U , the moment generating function is computed, and it can be written as

$$\begin{aligned} M_\tau(t) &= \frac{1}{B_\tau[2c(1-\tau) + t]} \exp\{-c^2(1-\tau) - ct\} \\ &+ \frac{\exp\{\frac{t^2}{4(1-\tau)}\}}{B_\tau} \sqrt{\frac{\pi}{1-\tau}} \left[\Phi\left(-\frac{t}{\sqrt{2(1-\tau)}}\right) - \Phi\left(\frac{-2c(1-\tau) - t}{\sqrt{2(1-\tau)}}\right) \right] \\ &+ \frac{\exp\{\frac{t^2}{4\tau}\}}{B_\tau} \sqrt{\frac{\pi}{\tau}} \left[\Phi\left(\frac{2c\tau - t}{\sqrt{2\tau}}\right) - \Phi\left(-\frac{t}{\sqrt{2\tau}}\right) \right] - \frac{1}{B_\tau(t - 2c\tau)} \exp\{-c^2\tau + ct\}, \end{aligned}$$

for $-2c(1-\tau) < t < 2c\tau$.

The first moment then is

$$E(U) = -\frac{1}{4B_\tau c^2(1-\tau)^2} \exp\{-c^2(1-\tau)\} + \frac{1}{4B_\tau c^2 \tau^2} \exp\{-c^2 \tau\} + \frac{1-2\tau}{2\tau(1-\tau)B_\tau},$$

and the variance is

$$\begin{aligned} Var(U) = \frac{1}{B_\tau} & \left[e^{-c^2(1-\tau)} \frac{1+2c^2(1-\tau)}{4c^3(1-\tau)^3} + e^{-c^2\tau} \frac{1+2c^2\tau}{4c^3\tau^3} + \frac{1}{2} \frac{\sqrt{\pi} [\Phi(c\sqrt{2\tau}) - 0.5]}{\tau^{3/2}} \right] \\ & + \frac{1}{2} \frac{\sqrt{\pi} [\Phi(c\sqrt{2(1-\tau)}) - 0.5]}{(1-\tau)^{3/2}} \Big]. \end{aligned}$$

These formulae may be easily generalised to the location and scale case. They can be used to obtain method of moments estimates of c and σ_τ to be used as initial values when minimizing (15) when $\rho_\tau(\cdot)$ is the Huber loss function, in line with Yu & Zhang (2005). The computations for obtaining the moment generating function, the expected value and the variance of U are not reported in the paper, but they are available from the authors upon request.

Appendix B: Proof of Theorem 1

Using a second order Taylor expansion,

$$2[\tilde{V}(\tau) - \hat{V}(\tau)] = \sqrt{n}(\tilde{\beta}_\tau - \hat{\beta}_\tau)^T (\Psi_\tau / \sigma_\tau) \sqrt{n}(\tilde{\beta}_\tau - \hat{\beta}_\tau) + o_p(1), \quad (B1)$$

where, by using (C4), $\Psi_\tau = \sigma_\tau^{-1} E(\psi'_{\tau i}) E(\mathbf{x}_i \mathbf{x}_i^T)$. Theorem 2.1 in Bianchi & Salvati (2015) ensures that

$$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) = \Psi_\tau^{-1} n^{-1/2} \sum_{i=1}^n \psi_{\tau i} \mathbf{x}_i + o_p(1). \quad (B2)$$

Similarly, a standard mean value expansion (under H_0) gives

$$n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i = n^{-1/2} \sum_{i=1}^n \psi_{\tau i} \mathbf{x}_i - \Psi_\tau \sqrt{n}(\tilde{\beta}_\tau - \beta_\tau) + o_p(1),$$

where $\tilde{\psi}_{\tau i} = \psi_\tau(\tilde{\varepsilon}_{\tau i} / \sigma_\tau)$, $\tilde{\varepsilon}_{\tau i} = y_i - \mathbf{x}_i^T \tilde{\beta}_\tau$. Hence,

$$\sqrt{n}(\tilde{\beta}_\tau - \beta_\tau) = \Psi_\tau^{-1} n^{-1/2} \left[-\sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i + \sum_{i=1}^n \psi_{\tau i} \mathbf{x}_i \right] + o_p(1). \quad (B3)$$

Substituting (B2) and (B3) into (B1), we obtain

$$2[\tilde{V}(\tau) - \hat{V}(\tau)] = \left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right)^T (E(\psi'_{\tau i}) E(\mathbf{x}_i \mathbf{x}_i^T))^{-1} \left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right) + o_p(1).$$

Following Wooldridge (2010), we introduce the $k \times p$ full rank matrix $\mathbf{R} = [\mathbf{0} : \mathbf{I}_k]$ and write H_0 as $\mathbf{R}\boldsymbol{\beta}_\tau = \mathbf{0}$. Because $\mathbf{R}\sqrt{n}(\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) = \mathbf{0}$, it can be proved (multiplying Equation (B3) by $\mathbf{R}\Psi_\tau^{-1}$) that

$$\mathbf{R}\Psi_\tau^{-1} n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \xrightarrow{d} N(\mathbf{0}, \mathbf{R}\Sigma_\tau \mathbf{R}^T),$$

where

$$\Sigma_\tau = \sigma_\tau^2 \frac{E\psi_{\tau i}^2}{E\psi'_{\tau i}} E[\mathbf{x}_i \mathbf{x}_i^T]^{-1}, \quad (\text{B4})$$

so that

$$\left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right)^T \Psi_\tau^{-1} \mathbf{R}^T (\mathbf{R}\Sigma_\tau \mathbf{R}^T)^{-1} \mathbf{R}\Psi_\tau^{-1} \left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right) \xrightarrow{d} \chi_k^2.$$

The previous expression can be simplified to

$$\left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right)^T (E(\psi_{\tau i}^2) E(\mathbf{x}_i \mathbf{x}_i^T))^{-1} \left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right) \xrightarrow{d} \chi_k^2,$$

and therefore, we have that

$$\begin{aligned} & 2 \frac{E\psi'_{\tau i}}{E\psi_{\tau i}^2} [\tilde{V}(\tau) - \hat{V}(\tau)] \\ &= \left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right)^T (E(\psi_{\tau i}^2) E(\mathbf{x}_i \mathbf{x}_i^T))^{-1} \left(n^{-1/2} \sum_{i=1}^n \tilde{\psi}_{\tau i} \mathbf{x}_i \right) + o_p(1) \xrightarrow{d} \chi_k^2. \end{aligned} \quad (\text{B5})$$

Appendix C: Sketch of the proof of Equation (28)

Under the assumptions of the theorem, convergence of $\hat{\tau}_j$ to τ_j is verified by using standard Taylor linearisation techniques. For the asymptotic distribution of the test statistic, let $Q(\boldsymbol{\tau}) = \sum_{j=1}^d \sum_{i=1}^{n_j} \rho \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{\tau_j}}{\sigma} \right)$, $\mathbf{s}(\boldsymbol{\tau}) = \left\{ \frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} \frac{\partial \rho}{\partial \tau_j} \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{\tau_j}}{\sigma} \right) \right\}_{j=1}^d$, $\mathbf{H}(\boldsymbol{\tau}) = \text{diag} \left\{ \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\partial^2 \rho}{\partial \tau_j^2} \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{\tau_j}}{\sigma} \right) \right\}$ and $\mathbf{n} = (n_1, \dots, n_d)^T$. Let $\mathbf{A}_0 = \text{diag}\{a_j\}$ and $\mathbf{B}_0 = \text{diag}\{b_j\}$ with

$$a_j = E \left[\frac{\partial^2 \rho}{\partial \tau_j^2} \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{\tau_j}}{\sigma} \right) \Big|_{0.5} \right] = \sigma^{-2} E \psi'_{ij} E \left(\mathbf{x}_{ij}^T \frac{\partial \boldsymbol{\beta}_{\tau_j}}{\partial \tau_j} \Big|_{0.5} \right)^2$$

$$b_j = E \left[\frac{\partial \rho}{\partial \tau_j} \left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_{\tau_j}}{\sigma} \right) \Big|_{0.5} \right]^2 = \sigma^{-2} E \psi_{ij}^2 E \left(\mathbf{x}_{ij}^T \frac{\partial \boldsymbol{\beta}_{\tau_j}}{\partial \tau_j} \Big|_{0.5} \right)^2.$$

Under H_0 , a mean value expansion yields

$$\mathbf{0} = \mathbf{s}(\hat{\boldsymbol{\tau}}) = \mathbf{s}(\mathbf{0.5}) + \sqrt{\mathbf{n}} \cdot (\hat{\boldsymbol{\tau}} - \mathbf{0.5}) + o_p(1),$$

implying $\sqrt{\mathbf{n}} \cdot (\hat{\boldsymbol{\tau}} - \mathbf{0.5}) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}_0^{-1} \mathbf{B}_0 \mathbf{A}_0^{-1})$, as $n_j \rightarrow +\infty$, $j = 1, \dots, d$, where \cdot denotes the Hadamard product.

Then,

$$Q(\mathbf{0.5}) - Q(\hat{\boldsymbol{\tau}}) = \frac{1}{2} (\hat{\boldsymbol{\tau}} - \mathbf{0.5})^T H(\dot{\boldsymbol{\tau}}) (\hat{\boldsymbol{\tau}} - \mathbf{0.5})$$

$$= \frac{1}{2} [\sqrt{\mathbf{n}} \cdot (\hat{\boldsymbol{\tau}} - \mathbf{0.5})]^T \mathbf{A}_0 [\sqrt{\mathbf{n}} \cdot (\hat{\boldsymbol{\tau}} - \mathbf{0.5})] + o_p(1),$$

where $\dot{\boldsymbol{\tau}}$ is a value between $\hat{\boldsymbol{\tau}}$ and $\mathbf{0.5}$. Hence,

$$2[Q(\mathbf{0.5}) - Q(\hat{\boldsymbol{\tau}})] \frac{E \psi'_{ij}}{E \psi_{ij}^2} = [\sqrt{\mathbf{n}} \cdot (\hat{\boldsymbol{\tau}} - \mathbf{0.5})]^T [\mathbf{A}_0^{-1} \mathbf{B}_0 \mathbf{A}_0^{-1}]^{-1} [\sqrt{\mathbf{n}} \cdot (\hat{\boldsymbol{\tau}} - \mathbf{0.5})] + o_p(1).$$

Intuitively, for reasons of symmetry, if $\tau_j = 0.5$ for $j = 1, \dots, d-1$, also $\tau_d = 0.5$ (as the global minimiser is $\tau = 0.5$). The same relationship needs to hold for the corresponding estimators $\hat{\tau}_j$'s. So the previous expression may be reparametrised leading to a χ_{d-1}^2 asymptotic distribution.