Statistica Sinica Preprint No: SS-2019-0414

Title	Maximum Lq-likelihood Estimation in Functional Measurement Error Models
Manuscript ID	SS-2019-0414
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0414
Complete List of Authors	Patricia Giménez, Lucas Guarracino and Manuel Galea
Corresponding Author	Patricia Giménez
E-mail	pcgimene@mdp.edu.ar

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

Statistica Sinica

MAXIMUM LQ-LIKELIHOOD ESTIMATION IN FUNCTIONAL MEASUREMENT ERROR MODELS

Patricia Giménez[1], Lucas Guarracino[1] and Manuel Galea[2]

[1] *Universidad Nacional de Mar del Plata, Argentina and*

[2] *Pontificia Universidad Católica de Chile, Chile*

Abstract: We consider a robust parametric procedure for estimating the structural parameters in functional measurement error models. The methodology extends the maximum Lq-likelihood approach to the more general problem of independent, but not identically distributed observations and the presence of incidental parameters. The proposal replaces the incidental parameters in the Lq-likelihood with their estimates, which depend on the structural parameter. The resulting estimator, called the maximum Lq-likelihood estimator (MLqE) adapts according to the discrepancy between the data and the postulated model by tuning a single parameter q, with $0 < q < 1$, that controls the trade-off between robustness and efficiency. The maximum likelihood estimator is obtained as a particular case when $q = 1$. We provide asymptotic properties of the MLqE under appropriate regularity conditions. Moreover, we describe the estimating algorithm based on a reweighting procedure, as well as a data-driven proposal for the choice of the tuning parameter q. The approach is illustrated and applied to the problem of estimating a bivariate linear normal relationship, including a small simulation

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

study and an analysis of a real data set.

Key words and phrases: Functional measurement error models, incidental parameters, maximum Lq-likelihood, robustness.

1. Introduction

This study deals with robust estimation in functional measurement error models based on an extension of the maximum Lq-likelihood (MLq) approach proposed by Ferrari and Yang (2010). In a typical measurement error model, a response vector variable $\boldsymbol{Y}$ is functionally related to a vector covariate $\boldsymbol{\xi}$ that is not observed exactly. Instead, it is observed with an error, a case often encountered in practice. Disregarding these measurement errors when estimating the regression parameters results in asymptotically biased (i.e., inconsistent) estimators. Numerous methods have been proposed to correct for measurement errors; see Fuller (1987), Cheng and Van Ness (1999), Carroll et al. (2006), and Buonaccorsi (2010), and the references cited therein.

The classical measurement error model considers that we observe the surrogate $\boldsymbol{X} = \boldsymbol{\xi} + \boldsymbol{u}$, independent of $\boldsymbol{Y}$, where the measurement error $\boldsymbol{u}$ is a random variable. Inference is based on a sample of n independent observations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$, where $\boldsymbol{Z}_j = (\boldsymbol{X}_j^T, \boldsymbol{Y}_j^T)^T$, for $j = 1, \ldots, n$. If the unobserved covariates $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ are unknown constants, then the model is

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

referred to as a functional model, and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ are nuisance parameters, the number of which increases with the sample size, called incidental parameters (Neyman and Scott (1948)). If $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ are considered a random sample from some distribution, then the model is referred to as a structural model.

In this study, we consider functional models. We model the density function of $\boldsymbol{Z}_j$, for $j = 1, \ldots n$, by

$$f_j(\boldsymbol{z}_j; \boldsymbol{\theta}, \boldsymbol{\xi}_j) = f_Y(\boldsymbol{y}_j; \boldsymbol{\theta}_1, \boldsymbol{\xi}_j) f_X(\boldsymbol{x}_j; \boldsymbol{\theta}_2, \boldsymbol{\xi}_j), \tag{1.1}$$

where f_Y and f_X are the models describing the relationships with the true unobserved covariate of the response and the observed covariate, respectively. Furthermore, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ and $\boldsymbol{\xi}_j$, for $j = 1, \ldots, n$, are vectors of unknown parameters. Here, $\boldsymbol{\theta}$ is the same for each j and is called a structural parameter; $\boldsymbol{\xi}_j$, which appears only once (in f_j), is called an incidental parameter. Our main interest lies in estimating the structural parameter $\boldsymbol{\theta}$.

It is not generally true that a maximum likelihood estimation produces consistent estimators of $\boldsymbol{\theta}$ (Stefanski (1985)). The problem is due to the large number of nuisance parameters (Neyman and Scott (1948); Andersen (1970); Lancaster (2000)). Mak (1982) studied a method for estimating $\boldsymbol{\theta}$ that also covers the maximum likelihood procedure.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

On the other hand, the unwieldy functional likelihood and its failure to produce consistent estimators have motivated the search for alternative methods of estimation, for example, the conditional and corrected score (Stefanski and Carroll (1987); Nakamura (1990); Giménez and Bolfarine (2000); Carroll et al. (2006)).

Measurement error model regression procedures are known to be non-robust, because they are highly sensitive to outlying observations and/or mild deviations from the assumed model, and are less robust than standard regression procedures (Ammann and Van Ness (1988, 1989)). This has motivated the search for more robust methods of estimation. In the literature, studies on robust estimation in measurement error models focused mainly on the structural model. In this case, unlike the functional model, the results are also sensitive to an incorrect specification of the parametric distribution of the true covariate. In many ways, in structural measurement error models, the methods for obtaining robust estimates are analogous to those developed for the ordinary regression; see Cheng and Van Ness (1999, Chap. 7) and the references cited therein, as well as Fekri and Ruiz-Gazen (2004, 2006) and Croux, Fekri, and Ruiz-Gazen (2010). The number of robust options in the literature for estimating the structural parameter in a functional measurement error model is quite limited, owing to the presence

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

of the incidental parameters. Carroll and Gallo (1982) discuss the classical independent error model with replication in the predictors. Zamar (1985) investigates robust orthogonal regression M-estimators. Abdullah (1989) presents a computational scheme based on an iteratively reweighted regression method. Luong and Mak (1991) propose M-estimators. Vilca-Labra, Bolfarine, and Arellano-Valle (1998) obtain robust estimators by deriving maximum likelihood estimators in a functional linear model under the assumption of elliptical distributions of the errors. In the same way, Galea and de Castro (2017) study a functional model with replication using Student's t distribution.

We adopt a different approach, and propose a new robust fully parametric estimation procedure for functional measurement error models based on the MLq approach introduced by Ferrari and Yang (2010) in the context of small-tail inference. The method has an information-theoretical perspective because it is based on minimizing an empirical version of the Tsallis–Havdra–Charvat entropy, or q-entropy, employed in the context of statistical mechanics (Tsallis (1988)). Ferrari and La Vecchia (2012) examine its infinitesimal robustness properties, and show that the procedure is related to minimizing the power divergence, or q-divergence (Cressie and Read (1984)), between the assumed model and the true model density un-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

derlying the data, when the parameter is properly rescaled.

We obtain an estimator of the structural parameter of the model using the MLq approach when the incidental parameters are first replaced by their estimates, which depend on the structural parameter. The resulting estimator, called the maximum Lq-likelihood estimator (MLqE), adapts according to the discrepancy between the data and the postulated model by tuning a single parameter q $(0 < q \leq 1)$, which controls the trade-off between robustness and efficiency. When $q < 1$, data points with high likelihoods are assigned large weights. Outliers are usually assigned small weights because of their low likelihoods. When the data are consistent with the model and $q \to 1$, the maximum likelihood estimator (MLE) is obtained as a particular case.

The remainder of the paper is organized as follows. In Section 2, we review the MLq estimation approach in the ordinary independent and identically distributed (i.i.d.) case. Then, we extend it to estimation in functional measurement error models in the case of independent, but not identically distributed observations and the presence of incidental parameters. In Section 3, we provide asymptotic properties of the estimators under appropriate regularity conditions. Section 4 describes the estimation algorithm and its convergence properties. In Section 5, we briefly present a data-driven

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

proposal for the choice of the tuning parameter q. Section 6 applies the proposed approach to a bivariate linear normal relationship and illustrates the performance of the method using a small simulation study and an analysis of a real data set. Some concluding remarks are provided in Section 7. The proofs are included in the online Supplementary Material.

2. MLq estimation

2.1 MLq estimation for i.i.d. observations

Let G represent the true data-generating distribution having density g with respect to the Lebesgue measure. The true unknown density function g is modeled by the parametric family of densities $\mathcal{F} = \{f(.;\boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$. It is assumed that $f(.;\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}$ and g have common support $\mathcal{X} \subset \mathbb{R}^k$, and that the family $\mathcal{F}$ is identifiable. One way to estimate the parameters is to minimize a data-based estimate of some appropriate divergence between the assumed model and the true density underlying the data. Minimum divergence estimators can afford considerable robustness at minimal expense of efficiency (Beran (1977)). The MLq estimation approach introduced by Ferrari and Yang (2010) in the context of small-tail inference provides a fully parametric estimation method that minimizes the power divergence, or q-divergence, between the true density g generating the data and the

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

postulated model density $f_{\boldsymbol{\theta}}$, defined by

$$\mathcal{D}_q(g, f_{\boldsymbol{\theta}}) = -\frac{1}{q}\mathrm{E}_G\, L_q\left\{\frac{f(\boldsymbol{Z};\boldsymbol{\theta})}{g(\boldsymbol{z})}\right\} = -\frac{1}{q}\int_{\mathcal{X}} L_q\left\{\frac{f(\boldsymbol{z};\boldsymbol{\theta})}{g(\boldsymbol{z})}\right\} g(\boldsymbol{z})\, d\boldsymbol{z}, \quad (2.1)$$

where $L_q(u) = (u^{1-q} - 1)/1 - q$ for $q \neq 1$, and $L_q(u) = \log u$ for $q = 1$, recovering the Kullback–Leibler divergence. When $q \to 1$, $L_q(u) \to \log(u)$. (2.1) is a divergence in the sense that $\mathcal{D}_q(g, f_{\boldsymbol{\theta}}) \geq 0$ and $\mathcal{D}_q(g, f_{\boldsymbol{\theta}}) = 0$ if and only if the densities g and $f_{\boldsymbol{\theta}}$ are equal. Such a quantity was first considered by Cressie and Read (1984) in the context of goodness-of-fit testing.

A direct minimization of (2.1) requires a nonparametric density estimation, which can be troublesome in multidimensional problems. A nonparametric density estimation can be avoided by approaching the minimization of (2.1) indirectly by minimizing a generalized information measure called q-entropy, or non-extensive entropy (Tsallis (1988)), given by

$$\mathcal{H}_q(g, f_{\boldsymbol{\theta}}) = -\int L_q\{f(\boldsymbol{z};\boldsymbol{\theta})\}g(\boldsymbol{z})\, d\boldsymbol{z} = -\mathrm{E}_G\{L_q\{f(\boldsymbol{Z};\boldsymbol{\theta})\}\}. \quad (2.2)$$

Ferrari and La Vecchia (2012) show that minimizing $\mathcal{H}_q(g, f_{\boldsymbol{\theta}})$ is equivalent to minimizing $\mathcal{D}_q(g^{(1/q)}, f_{\boldsymbol{\theta}})$, where $g^{(1/q)}$ is the density proportional to $g^{1/q}$. Therefore, a transformation on the estimates is required in order to obtain consistent estimates for the true density g. The key advantage of working with (2.2) instead (2.1) is that the former can be estimated easily from data averages.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

Let $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^*$ be the minimizers of $\mathcal{D}_q(g, f_{\boldsymbol{\theta}})$ and $\mathcal{H}_q(g, f_{\boldsymbol{\theta}})$, respectively; whereas $\boldsymbol{\theta}$ is a generic element of Θ, $\boldsymbol{\theta}^*$ is called the surrogate parameter. The existence and uniqueness of $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^*$ are assumed in the interior Θ° of Θ.

For $0 < \alpha < \infty$, the power transformation of a density g is defined by

$$g^{(\alpha)}(\boldsymbol{z}) = \frac{g(\boldsymbol{z})^\alpha}{\int g(\boldsymbol{z})^\alpha \, d\boldsymbol{z}}, \tag{2.3}$$

provided that the integral in the denominator converges and it is assumed that $\mathcal{F}$ is closed under (2.3), for all $0 < \alpha < 1$. A continuous mapping $\tau_\alpha \colon \Theta \to \Theta$ is defined satisfying $f(\boldsymbol{z}; \tau_\alpha(\boldsymbol{\theta})) = f^{(\alpha)}(\boldsymbol{z}; \boldsymbol{\theta}), \quad \forall\, \boldsymbol{z} \in \mathcal{X}$. That is, $\tau_\alpha(\boldsymbol{\theta})$ is simply the parameter of the density proportional to $f_{\boldsymbol{\theta}}^\alpha$, which can be computed analytically for common families of distributions, such as the exponential distribution.

The considerations above motivate the following estimation strategy. Given $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$, an i.i.d. sample from G, with g its corresponding density, a consistent estimator of the surrogate parameter $\boldsymbol{\theta}^*$ can be obtained by minimizing the empirical version of the q-entropy (2.2) or, equivalently, by maximizing the L_q-likelihood function. That is, the MLqE is defined by $\hat{\boldsymbol{\theta}}_n^* = \arg\max_{\boldsymbol{\theta} \in \Theta} \sum_{j=1}^n L_q\{f(\boldsymbol{Z}_j; \boldsymbol{\theta})\}$.

Let $\boldsymbol{U}^*(\boldsymbol{Z}; \boldsymbol{\theta}) = \nabla L_q\{f(\boldsymbol{Z}; \boldsymbol{\theta})\} = \boldsymbol{U}(\boldsymbol{Z}; \boldsymbol{\theta}) f(\boldsymbol{Z}; \boldsymbol{\theta})^{1-q}$ be the q-score function, where $\boldsymbol{U}(\boldsymbol{Z}; \boldsymbol{\theta}) = \nabla \log\{f(\boldsymbol{Z}; \boldsymbol{\theta})\}$ is the usual maximum likelihood

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

score function, and ∇ is the gradient with respect to $\boldsymbol{\theta}$. Then, $\hat{\boldsymbol{\theta}}_n^*$ is a solution of the estimating equations

$$\sum_{j=1}^{n} \boldsymbol{U}^*(\boldsymbol{Z}_j; \boldsymbol{\theta}) = \sum_{j=1}^{n} \boldsymbol{U}(\boldsymbol{Z}_j; \boldsymbol{\theta}) f(\boldsymbol{Z}_j; \boldsymbol{\theta})^{1-q} = \mathbf{0}. \tag{2.4}$$

When $q \neq 1$, (2.4) can be viewed as a weighted version of the efficient maximum likelihood score equation, with weights proportional to the $(1-q)$th power of the assumed density. Throughout this article, we assume $0 < q < 1$, such that observations that disagree with the model receive low weight, providing remarkably robust estimators with negligible efficiency losses compared with those of the maximum likelihood. If $q = 1$, all observations get weights equal to one and the MLqE coincides with the MLE.

Note that $\hat{\boldsymbol{\theta}}_n^*$ is weakly consistent for $\boldsymbol{\theta}^*$. Assuming that $\tau_q(\boldsymbol{\theta})$ is defined for all $\boldsymbol{\theta} \in \Theta^\circ$, Ferrari and La Vecchia (2012) show that $\hat{\boldsymbol{\theta}}_n = \tau_q(\hat{\boldsymbol{\theta}}_n^*)$ is weakly consistent for $\boldsymbol{\theta}^0$. This result is based on the fact that if $f_{\boldsymbol{\theta}^0}$ is the true density generating the data, then $\mathcal{D}_q(f_{\boldsymbol{\theta}^0}^{(1/q)}, f_{\boldsymbol{\theta}^*}) = \min_{\boldsymbol{\theta} \in \Theta} \mathcal{D}_q(f_{\boldsymbol{\theta}^0}^{(1/q)}, f_{\boldsymbol{\theta}})$, which is zero if and only if $f_{\boldsymbol{\theta}^*} = f_{\boldsymbol{\theta}^0}^{(1/q)}$, that is, $\boldsymbol{\theta}^* = \tau_{1/q}(\boldsymbol{\theta}^0)$.

2.2 MLq estimation in functional measurement error models

Here, we extend the MLq approach to estimate the structural parameter in functional measurement error models. Adapting the approach to the case of independent, but not identically distributed observations and the presence

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

of incidental parameters requires a substantial and nontrivial extension to the approach followed in the case of i.i.d. observations.

Let us assume that our data $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent, where $\boldsymbol{Z}_j = (\boldsymbol{X}_j^T, \boldsymbol{Y}_j^T)^T$, with $\boldsymbol{Y}_j$ the vector response variable and $\boldsymbol{X}_j$ the observed covariate. We assume that each $\boldsymbol{Z}_j$ has distribution function G_j and density g_j with respect to the Lebesgue measure. We want to model g_j by the family $\mathcal{F}_j = \{f_j(.;\boldsymbol{\theta}, \boldsymbol{\xi}_j),\ \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p,\ \boldsymbol{\xi}_j \in \Xi \subset \mathbb{R}^r\}$, for all $j = 1, \ldots n$, where $f_j(.;\boldsymbol{\theta}, \boldsymbol{\xi}_j) = f_{\boldsymbol{\theta}, \boldsymbol{\xi}_j}$ is the assumed model density of $\boldsymbol{Z}_j$ given in (1.1). Here, the observations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent, but not identically distributed. The structural parameter $\boldsymbol{\theta}$ is the same for each j, but f_j depends also on the incidental parameter $\boldsymbol{\xi}_j$.

We assume that for $j = 1, 2, \ldots$, $\mathcal{F}_j$ is closed under the power transformation (2.3). We define a continuous mapping $\tau_\alpha \colon \Theta \times \Xi \to \Theta \times \Xi$, where $\tau_\alpha(\boldsymbol{\theta}, \boldsymbol{\xi}_j) = (\tau_\alpha^1(\boldsymbol{\theta}), \tau_\alpha^2(\boldsymbol{\xi}_j))$, satisfying $f_j(\boldsymbol{z}; \tau_\alpha^1(\boldsymbol{\theta}), \tau_\alpha^2(\boldsymbol{\xi}_j)) = f_j^{(\alpha)}(\boldsymbol{z}; \boldsymbol{\theta}, \boldsymbol{\xi}_j)$ $\forall\, \boldsymbol{z} \in \mathcal{X}$, for $j = 1, 2, \ldots$ The closure of $\mathcal{F}_j$ under (2.3) in the functional model seems to be a stronger condition than in the structural model. However, a closed form for τ_α can be obtained in some relevant cases. Consider, for example, the regression setting where the response variable $\boldsymbol{Y}_j$ follows the exponential family of densities

$$f_Y(\boldsymbol{y}_j; \boldsymbol{\beta}, \boldsymbol{\xi}_j) = \exp\{\boldsymbol{\eta}(\boldsymbol{\beta}^T \boldsymbol{\xi}_j)^T \boldsymbol{a}(\boldsymbol{y}_j) - b(\boldsymbol{\beta}^T \boldsymbol{\xi}_j)\},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

with $\boldsymbol{a}(.)$ and $b(.)$ known functions. Here, the explanatory vector $\boldsymbol{\xi}_j$ is measured with an independent normal error such that $\boldsymbol{X}_j$ follows a multivariate normal density with mean $\boldsymbol{\xi}_j$ and covariance matrix Σ. When $\boldsymbol{\eta}(\boldsymbol{\beta}^T\boldsymbol{\xi}_j) = \boldsymbol{\beta}^T\boldsymbol{\xi}_j$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \text{vech}^T\Sigma)^T$, we obtain $\tau_\alpha^1(\boldsymbol{\theta}) = (\alpha\boldsymbol{\beta}^T, \alpha^{-1}\text{vech}^T\Sigma)$ and $\tau_\alpha^2(\boldsymbol{\xi}_j) = \boldsymbol{\xi}_j$. Another example is given in Section 6.

We also assume the existence and uniqueness of $(\boldsymbol{\theta}^0, \boldsymbol{\xi}_1^0, \ldots, \boldsymbol{\xi}_n^0)$ and $(\boldsymbol{\theta}^*, \boldsymbol{\xi}_1^*, \ldots, \boldsymbol{\xi}_n^*)$ as the minimizers of the averaged q-divergence $\frac{1}{n}\sum_{j=1}^n \mathcal{D}_q(g_j, f_{\boldsymbol{\theta},\boldsymbol{\xi}_j})$ and the averaged q-entropy $\frac{1}{n}\sum_{j=1}^n \mathcal{H}_q(g_j, f_{\boldsymbol{\theta},\boldsymbol{\xi}_j})$, respectively, for all large n. We call $\boldsymbol{\theta}^*$ the surrogate structural parameter. Finally, we assume that τ_q is defined for all $(\boldsymbol{\theta}, \boldsymbol{\xi})$ in the interior of $\Theta \times \Xi$.

Given the sample $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$, the ML$q$E $\hat{\boldsymbol{\theta}}_n^*, \hat{\boldsymbol{\xi}}_1^*, \ldots, \hat{\boldsymbol{\xi}}_n^*$ is defined by $(\hat{\boldsymbol{\theta}}_n^*, \hat{\boldsymbol{\xi}}_1^*, \ldots, \hat{\boldsymbol{\xi}}_n^*) = \arg\max_{\boldsymbol{\theta},\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n} \sum_{j=1}^n L_q\{f_j(\boldsymbol{Z}_j; \boldsymbol{\theta}, \boldsymbol{\xi}_j)\}$. The main interest lies in estimating the structural parameter. The approach pursued here replaces the incidental parameters $\boldsymbol{\xi}_j$ with the estimators $\hat{\boldsymbol{\xi}}_j = \hat{\boldsymbol{\xi}}_j(\boldsymbol{Z}_j; \boldsymbol{\theta})$, for $j = 1, 2, \ldots,$ given by $\hat{\boldsymbol{\xi}}_j = \arg\max_{\boldsymbol{\xi}_j \in \Xi} L_q\{f_j(\boldsymbol{Z}_j; \boldsymbol{\theta}, \boldsymbol{\xi}_j)\}$, for $j = 1, 2, \ldots$ Let

$$H_n(\boldsymbol{\theta}) = \sum_{j=1}^n L_q\{f_j(\boldsymbol{Z}_j; \boldsymbol{\theta}, \hat{\boldsymbol{\xi}}_j(\boldsymbol{Z}_j; \boldsymbol{\theta}))\} = \sum_{j=1}^n h_j(\boldsymbol{Z}_j; \boldsymbol{\theta}) \tag{2.5}$$

be the objective function. Then, we can characterize $\hat{\boldsymbol{\theta}}_n^*$, if it exists, as $\hat{\boldsymbol{\theta}}_n^* = \arg\max_{\boldsymbol{\theta}\in\Theta} H_n(\boldsymbol{\theta})$.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

We assume that the following derivatives exist a.e. for all j :

$$\nabla_k h_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = U^{\dagger}_{jk}(\boldsymbol{Z}_j;\boldsymbol{\theta}) \text{ and } \nabla_{kl} h_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = I^{\dagger}_{jkl}(X_j;\boldsymbol{\theta}), \quad k,l=1,\ldots,p,$$

where ∇_k and ∇_{kl} represent the partial derivatives with respect to the indicated components of $\boldsymbol{\theta}$. Let $\boldsymbol{U}^{\dagger}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = (U^{\dagger}_{j1}(\boldsymbol{Z}_j;\boldsymbol{\theta}),\ldots,U^{\dagger}_{jp}(\boldsymbol{Z}_j;\boldsymbol{\theta}))^T$ and $\boldsymbol{I}^{\dagger}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})$ be the symmetric matrix with (k,l)th element equal to $I^{\dagger}_{jkl}(\boldsymbol{Z}_j;\boldsymbol{\theta})$.

Differentiating (2.5) with respect to $\boldsymbol{\theta}$, we have the following estimating equation:

$$\sum_{j=1}^{n} \boldsymbol{U}^{\dagger}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = \sum_{j=1}^{n} \tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q} = \boldsymbol{0}, \tag{2.6}$$

where

$$\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = f_j(\boldsymbol{Z}_j;\boldsymbol{\theta},\hat{\boldsymbol{\xi}}_j) \quad \text{and} \quad \tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = \boldsymbol{U}_j(\boldsymbol{Z}_j;\boldsymbol{\theta},\hat{\boldsymbol{\xi}}_j), \tag{2.7}$$

with $\boldsymbol{U}_j(\boldsymbol{Z}_j;\boldsymbol{\theta},\boldsymbol{\xi}_j) = \nabla \log f_j(\boldsymbol{Z}_j;\boldsymbol{\theta},\boldsymbol{\xi}_j)$.

The MLqE $\hat{\boldsymbol{\theta}}^*_n$ is obtained as a solution of the estimating equation (2.6). This equation is satisfied by the maximizer of $H_n(\boldsymbol{\theta})$ in (2.5), whenever this maximum exists.

We also define the matrices

$$\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\left[\boldsymbol{I}^{\dagger}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\right] \text{ and } \overline{\boldsymbol{\Gamma}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\left[\boldsymbol{U}^{\dagger}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\boldsymbol{U}^{\dagger}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^T\right].$$

If the true densities belong to the model family, then these matrices will depend on the incidental parameters $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n$.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

3. Asymptotic properties

In this section, we study the asymptotic properties the MLqE of the structural parameter. The consistency and asymptotic normality of the MLqE $\hat{\boldsymbol{\theta}}_n^*$ can be derived under appropriate regularity conditions. Because in a functional model, the basic assumption that the observations are i.i.d. is not met, these regularity conditions are quite different from the usual i.i.d. case. We assume the following:

C0. $\frac{1}{n}\sum_{j=1}^{n} E_{G_j}[h_j(\boldsymbol{Z}_j; .)]$ converges uniformly to a function $\bar{h}(.)$ in a neighborhood of $\boldsymbol{\theta}^\dagger$, where $\boldsymbol{\theta}^\dagger \in \Theta^\circ$ is a local maximum of $\bar{h}$.

Some additional regularity conditions (C1 to C6) are provided in the Supplementary Material to establish the asymptotic properties of the MLqE. Moreover, the asymptotic results of this section are derived under the assumption that at $\boldsymbol{\theta}^\dagger$, we have

$$E_{G_j}[\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j; \boldsymbol{\theta}^\dagger)] = \mathbf{0}, \quad j = 1, 2, \ldots \tag{3.1}$$

Theorem 1. *Let $\boldsymbol{\theta}^\dagger$ in the interior of Θ satisfy the regularity conditions C0 and C1 to C6 in the Supplementary Material, as well as assumption (3.1),*

(i) With probability tending to one, (2.6) has a root $\hat{\boldsymbol{\theta}}_n^*$, which converges in probability to $\boldsymbol{\theta}^\dagger$. If $\tilde{\boldsymbol{\theta}}_n$ is any other consistent root of (2.6), then $\hat{\boldsymbol{\theta}}_n^* = \tilde{\boldsymbol{\theta}}_n$ with probability tending to one.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

(ii) $\hat{\boldsymbol{\theta}}_n^*$ *is asymptotically normal with mean* $\boldsymbol{\theta}^\dagger$ *and covariance matrix* $n^{-1}\boldsymbol{\Omega}_n(\boldsymbol{\theta}^\dagger)$, *where* $\boldsymbol{\Omega}_n(\boldsymbol{\theta}) = \overline{\boldsymbol{\Lambda}}_n^{-1}(\boldsymbol{\theta})\overline{\boldsymbol{\Gamma}}_n(\boldsymbol{\theta})\overline{\boldsymbol{\Lambda}}_n^{-1}(\boldsymbol{\theta})^T$.

The proof is provided in the Supplementary Material.

Remark 1. When all the true densities g_j belong to the model family so that $g_j = f_j(.;\boldsymbol{\theta}^0, \boldsymbol{\xi}_j^0)$, for some common $\boldsymbol{\theta}^0$, we have that

$$\sum_{j=1}^n \mathcal{D}_q(f^{(1/q)}_{\boldsymbol{\theta}^0,\boldsymbol{\xi}_j^0}, f_{\boldsymbol{\theta}^*,\boldsymbol{\xi}_j^*}) = \min_{\boldsymbol{\theta},\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n} \sum_{j=1}^n \mathcal{D}_q(f^{(1/q)}_{\boldsymbol{\theta}^0,\boldsymbol{\xi}_j^0}, f_{\boldsymbol{\theta},\boldsymbol{\xi}_j}),$$

for all n, which is zero if and only if $f_{\boldsymbol{\theta}^*,\boldsymbol{\xi}_j^*} = f^{(1/q)}_{\boldsymbol{\theta}^0,\boldsymbol{\xi}_j^0}$, for $j = 1, 2, \ldots,$ that is $\boldsymbol{\theta}^* = \tau^1_{1/q}(\boldsymbol{\theta}^0)$ and $\boldsymbol{\xi}_j^* = \tau^2_{1/q}(\boldsymbol{\xi}_j^0)$, for $j = 1, 2, \ldots$. Owing to the noise in estimating $\hat{\boldsymbol{\xi}}_j$, in general, $\boldsymbol{\theta}^\dagger$ is not equal to $\boldsymbol{\theta}^*$, unlike what happens for an i.i.d. sample. In this case, $\boldsymbol{\theta}^\dagger$ depends in general on the surrogate structural parameter $\boldsymbol{\theta}^*$ and the incidental parameters $\boldsymbol{\xi}_1^*, \ldots, \boldsymbol{\xi}_n^*$. Assumption (3.1) is satisfied when it is possible to obtain estimators $\hat{\boldsymbol{\xi}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})$ so that $\boldsymbol{\theta}^\dagger$ is determined only by $\boldsymbol{\theta}^*$ and is independent of the incidental parameters. In this case, there exists a function $\rho(.)$ such that $\boldsymbol{\theta}^\dagger = \rho(\boldsymbol{\theta}^*)$, as shown in the model considered in Section 6.

Remark 2. When the true densities belong to the model family and there exists a one-to-one function $\rho(.)$ such that $\boldsymbol{\theta}^\dagger = \rho(\boldsymbol{\theta}^*)$, a consistent estimator of $\boldsymbol{\theta}^0$ is given by $\hat{\boldsymbol{\theta}}_n = \eta_q(\hat{\boldsymbol{\theta}}_n^*)$, where $\eta_q = \tau_q^1 \circ \rho^{-1}$. When this is the case, we call $\hat{\boldsymbol{\theta}}_n$ the corrected MLqE. We have that $\hat{\boldsymbol{\theta}}_n$ is asymptotically normal with

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

mean $\boldsymbol{\theta}^0$ and covariance matrix $n^{-1}\boldsymbol{D}\boldsymbol{\Omega}_n(\boldsymbol{\theta}^\dagger)\boldsymbol{D}^T$, where $\boldsymbol{D}$ is the matrix $\nabla\eta_q$ evaluated in $\boldsymbol{\theta}^\dagger$. The asymptotic properties of the MLqE depend on the asymptotic behavior assumed for the incidental parameter sequence. When $q = 1$, we can obtain, as a particular case, the asymptotic results derived by Mak (1982), which include the maximum likelihood estimation in functional measurement error models.

4. Estimation algorithm

The form of the estimating equation suggests exploring reweighting strategies to compute the estimates. If we define

$$\omega_j = \omega_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = \frac{\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q}}{\sum_{k=1}^n \tilde{f}_k(\boldsymbol{Z}_k;\boldsymbol{\theta})^{1-q}}, \tag{3.2}$$

the estimating equation can be written as $\sum_{j=1}^n \omega_j \tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = \mathbf{0}$, where ω_j, for $j = 1,\dots,n$, are weights that depend on the assumed model density, such that $\sum_{j=1}^n \omega_j = 1$. If $\boldsymbol{\theta}^{(s)}$ denotes the estimator in step s, then the estimator in step $s+1$ satisfies

$$\sum_{j=1}^n \omega_j^{(s)} \tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}^{(s+1)}) = \mathbf{0}, \tag{3.3}$$

where the weights $\omega_j^{(s)} = \omega_j(\boldsymbol{Z}_j;\boldsymbol{\theta}^{(s)})$, computed using (3.2), are updated at each step. If $0 < q < 1$, observations that disagree with the model receive a low weight. In the case $q = 1$, all observations receive the same weight and the estimators coincide with the MLE. The algorithm can be initialized by setting $\boldsymbol{\theta}^{(0)}$ as the MLE. The process continues until the convergence criterion $\|\boldsymbol{\theta}^{(s+1)} - \boldsymbol{\theta}^{(s)}\| < \delta$, for a certain stopping rule δ, is satisfied. The results in Arslan (2004) can be used to study the convergence properties of the algorithm.

It can be shown that the algorithm (3.3) produces at each step an ascent of the objective function H_n given in (2.5). If we define the function $\mathbb{Q}_n \colon \Theta \times \Theta \to \mathbb{R}$ by $\mathbb{Q}_n(\boldsymbol{u}, \boldsymbol{v}) = -\sum_{j=1}^{n} \tilde{f}_j(\boldsymbol{z}_j; \boldsymbol{v})^{1-q} \log \tilde{f}_j(\boldsymbol{z}_j; \boldsymbol{u})$, then the iterative procedure given by (3.3) implicitly defines a mapping $M_n \colon \Theta \to \Theta$ with $\boldsymbol{\theta}^{(s+1)} = M_n(\boldsymbol{\theta}^{(s)})$, for $s = 0, 1, 2, \dots$, where $M_n(\boldsymbol{v}) = \arg\min_{\boldsymbol{u} \in \Theta} \mathbb{Q}_n(\boldsymbol{u}, \boldsymbol{v})$. Because $\frac{\partial}{\partial \boldsymbol{u}} \mathbb{Q}_n(\boldsymbol{u}, \boldsymbol{v}) = -\sum_{j=1}^{n} \tilde{f}_j(\boldsymbol{z}_j; \boldsymbol{v})^{1-q} \tilde{\boldsymbol{U}}_j(\boldsymbol{z}_j; \boldsymbol{u})$, we can see that $\boldsymbol{v}$ is a stationary point of H_n if and only if it is a fixed point of M_n.

In the following proposition, we show that if $\boldsymbol{\theta}^{(s)}$, for $s = 0, 1, 2, \dots$, is not a fixed point of M_n, then the sequence $\{H_n(\boldsymbol{\theta}^{(s)})\}$, for $s = 0, 1, 2, \dots$, forms a monotone increasing sequence.

Proposition 1. *If $\boldsymbol{v}$ is not a fixed point of M_n, then $H_n(\boldsymbol{v}) < H_n(M_n(\boldsymbol{v}))$.*

The global convergence behavior of the sequences $\left\{\boldsymbol{\theta}^{(s)}\right\}_{s \geq 0}$ and $\left\{H_n(\boldsymbol{\theta}^{(s)})\right\}_{s \geq 0}$ is presented in the following proposition.

Proposition 2. *Let $\left\{\boldsymbol{\theta}^{(s)}\right\}_{s \geq 0}$ be a sequence generated by the equation $\boldsymbol{\theta}^{(s+1)} \in M_n(\boldsymbol{\theta}^{(s)})$, with an initial point $\boldsymbol{\theta}^{(0)} \in \Theta$. If all the points $\left\{\boldsymbol{\theta}^{(s)}\right\}_{s \geq 0}$*

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

are contained in a compact subset of Θ, *then all the limit points of* $\{\boldsymbol{\theta}^{(s)}\}_{s\geq 0}$ *are stationary points of* H_n, *and* $H_n(\boldsymbol{\theta}^{(s)})$ *converges increasingly to* $H_n(\boldsymbol{\theta}^*)$, *where* $\boldsymbol{\theta}^* \in \Theta$ *is a stationary point of* H_n.

Note that the notation $\boldsymbol{\theta}^{(0)}$ here refers to the starting point, and not to the true value of the parameter $\boldsymbol{\theta}^0$.

When the reweighting algorithm starts from a point very close to the maximum of the objective function H_n, then the sequence $\{\boldsymbol{\theta}^{(s)}\}_{s\geq 0}$ converges to it (Arslan (2004)). If the MLE cannot be computed efficiently, then it might be better to take a random starting point. Another alternative could be to take the naive MLE, which maximizes $\sum_{j=1}^{n} \log\{f_Y(\boldsymbol{y}_j; \boldsymbol{\theta}, \boldsymbol{x}_j)\}$, assuming that the covariates $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ are observed without error.

5. Choice of parameter q

Having a reasonable strategy for selecting q is crucial for applying the method to practical real-data scenarios, because there may be substantial variation in the performance of the estimators. The meaning of the parameters of each model could determine a more restrictive range of admissible values of q, as is the case for the application of Section 6.

There can be no universal way of selecting an appropriate value of q in a given situation. Choosing the tuning parameter to minimize the estimated

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

summed mean squared error is a quite general approach for families of robust methods indexed by tuning parameters. Warwick and Jones (2005) and Ghosh and Basu (2015) minimize of the estimated asymptotic mean squared error. For the numerical example of Section 6, we consider a data-based choice of an appropriate value of q by minimizing a parametric bootstrap estimate of the mean squared error of the corrected MLqE, $\hat{\boldsymbol{\theta}}_n$. Given the observed sample $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$, for q fixed in a grid $Q = \{q_1, \ldots, q_m\}$, the corresponding $\hat{\boldsymbol{\theta}}_n^{(q)}$ is computed and $\hat{\xi}_j^{(q)} = \hat{\xi}_j(\boldsymbol{z}_j; \hat{\boldsymbol{\theta}}_n^{(q)})$, for $j = 1, \ldots, n$. Then, B bootstrap samples of size n, $\{\boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_n^*\}$, are taken from the densities $f_j(\,.\,; \hat{\boldsymbol{\theta}}_n^{(q)}, \hat{\xi}_j^{(q)})$, for $j = 1, \ldots, n$. For each of the B samples, the estimators $\hat{\boldsymbol{\theta}}_{n,(b)}^{(q)}$, for $b = 1, \ldots, B$, are calculated and the mean squared error (MSE) of $\hat{\boldsymbol{\theta}}_n^{(q)}$ is estimated using $\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\theta}}_n^{(q)}) = B^{-1} \sum_{b=1}^{B} \|\hat{\boldsymbol{\theta}}_{n,(b)}^{(q)} - \hat{\boldsymbol{\theta}}_n^{(q)}\|^2$. The procedure is repeated for all $q \in Q$, and the optimal q is chosen as $q_{\mathrm{opt}} = \arg\min_{q \in Q} \widehat{\mathrm{MSE}}(\hat{\boldsymbol{\theta}}_n^{(q)})$. A parametric bootstrap estimate of the MSE may be more accurate than its nonparametric version, because in the latter case, outliers in the original sample may appear multiple times in a resample.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

6. Application to the simple linear functional model with normal errors

Consider the simple linear regression model represented by the equations

$$Y_j = \alpha + \beta\xi_j + e_j, \quad \text{and} \quad X_j = \xi_j + u_j, \quad j = 1, \ldots, n, \qquad (3.4)$$

where e_j and u_j are independent and normally distributed with zero means and variances σ_e^2 and σ_u^2, respectively. Here, we consider the case where $\lambda = \sigma_e^2/\sigma_u^2$ is assumed to be known for the identifiability of the model. Without loss of generality, it is assumed that $\lambda = 1$, implying that $\sigma_e^2 = \sigma_u^2 = \phi$. The structural parameter is $\boldsymbol{\theta} = (\alpha, \beta, \phi)^T$. The unknown quantities $\xi_1, \ldots, \xi_n$ are incidental parameters because their number increases with the sample size. Denote the true unknown parameters by $\boldsymbol{\theta}^0 = (\alpha^0, \beta^0, \phi^0)^T$ and ξ_j^0, for $j = 1, 2, \ldots$. Model (3.4) can be written as $\boldsymbol{Z}_j = \boldsymbol{a} + \boldsymbol{b}\xi_j + \boldsymbol{\epsilon}_j$, for $j = 1, \ldots, n$, where $\boldsymbol{Z}_j = (X_j, Y_j)^T$ are observable, $\boldsymbol{\epsilon}_j = (u_j, e_j)^T$, for $j = 1, \ldots, n$ are i.i.d. with $\boldsymbol{\epsilon}_j \sim N_2(\mathbf{0}, \phi\boldsymbol{I}_2)$, $\boldsymbol{a} = (0, \alpha)^T$, and $\boldsymbol{b} = (1, \beta)^T$. Then, $\boldsymbol{Z}_j \sim N_2(\boldsymbol{\mu}_j, \phi\boldsymbol{I}_2)$, where $\boldsymbol{\mu}_j = \boldsymbol{a} + \boldsymbol{b}\xi_j$. We denote by $f_j(\boldsymbol{z}_j; \boldsymbol{\theta}, \xi_j)$ the assumed model density. Observations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent, but not identically distributed. Here, the structural parameter $\boldsymbol{\theta} = (\alpha, \beta, \phi)^T$ is the same for each j, but f_j depends also on the incidental parameter ξ_j. Note that the family density model is closed under transformation (2.3),

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

where $\tau_q(\boldsymbol{\theta}, \xi_j) = (\tau_q^1(\boldsymbol{\theta}), \tau_q^2(\xi_j))$, with $\tau_q^1(\boldsymbol{\theta}) = (\alpha, \beta, q^{-1}\phi)$ and $\tau_q^2(\xi_j) = \xi_j$.

The Lq-likelihood function is given by $\sum_{j=1}^n L_q\{f_j(\boldsymbol{Z}_j; \boldsymbol{\theta}, \xi_j)\}$. Solving $\partial L_q\{f_j(\boldsymbol{Z}_j; \boldsymbol{\theta}, \xi_j)\}/\partial \xi_j = 0$, we obtain for each j and a given $\boldsymbol{\theta}$, the conditional MLqE of ξ_j given $\boldsymbol{\theta}$, $\hat{\xi}_j = \hat{\xi}_j(\boldsymbol{Z}_j; \boldsymbol{\theta}) = c^{-1}[X_j + \beta(Y_j - \alpha)]$, where $c = \boldsymbol{b}^T\boldsymbol{b} = 1 + \beta^2$. Then, replacing ξ_j by $\hat{\xi}_j$ in the Lq-likelihood, we obtain the objective function

$$H_n = \sum_{j=1}^n h_j(\boldsymbol{Z}_j; \boldsymbol{\theta}) = \sum_{j=1}^n L_q\{\tilde{f}_j(\boldsymbol{Z}_j; \boldsymbol{\theta})\}, \tag{3.5}$$

where $\tilde{f}_j(\boldsymbol{Z}_j; \boldsymbol{\theta}) = f_j(\boldsymbol{Z}_j; \boldsymbol{\theta}, \hat{\xi}_j) = \frac{1}{2\pi\phi} \exp\left\{-\frac{1}{2c\phi}(Y_j - \alpha - \beta X_j)^2\right\}$. When the true density belongs to the model, that is, $g_j(.) = f_j(., \boldsymbol{\theta}^0, \xi_j^0)$, with $\boldsymbol{\theta}^0 = (\alpha^0, \beta^0, \phi^0)^T$, there exists a parameter $\boldsymbol{\theta}^\dagger = (\alpha^0, \beta^0, k\phi^0)^T$, where $k = q - \frac{1}{2}$, that maximizes $\frac{1}{n}\sum_{j=1}^n E_j[h_j(\boldsymbol{Z}_j; \boldsymbol{\theta})]$ for all large n. Furthermore, the parameter satisfies $E_j[\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j; \boldsymbol{\theta}^\dagger)] = \boldsymbol{0}$, for $j = 1, 2, \ldots$, where E_j denotes the expectation with respect to the model distribution $N_2(\boldsymbol{\mu}_j^0, \phi^0\boldsymbol{I}_2)$, with $\boldsymbol{\mu}_j^0 = \boldsymbol{a}^0 + \boldsymbol{b}^0\xi_j^0$, $\boldsymbol{a}^0 = (0, \alpha^0)^T$, and $\boldsymbol{b}^0 = (1, \beta^0)^T$. Derivations of these results are included in Lemmas 1 to 3 in Section S3.2 of the Supplementary Material.

Because the parameter $\phi^\dagger = k\phi^0$ is related to the error variance, it must be that $k > 0$, and then we have $\frac{1}{2} < q \leq 1$. Furthermore, it can be shown that the regularity conditions are satisfied, provided that the sequence of incidental parameters (ξ_j^0) verify $0 < \liminf \frac{1}{n}\sum_{j=1}^n(\xi_j^0 - \bar{\xi}_n^0)^2 \leq$

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

$\limsup \frac{1}{n}\sum_{j=1}^{n}(\xi_j^0 - \bar{\xi_n^0})^2 < \infty$, where $\bar{\xi_n^0} = \frac{1}{n}\sum_{j=1}^{n}\xi_j^0$ and $\lim_{n\to\infty}\frac{1}{n^{1+\frac{\gamma}{2}}}\sum_{j+1}^{n}|\xi_j^0|^{2+\gamma} = 0$, for some $\gamma > 0$ (see Mak (1982)). Because the sequence of incidental parameters is not observable, we cannot guarantee the validity of these assumptions. However, we can see that they are satisfied if $\bar{\xi_n^0}$ and $\frac{1}{n}\sum_{j=1}^{n}(\xi_j^0 - \bar{\xi_n^0})^2$ converge to finite limits. This latter condition is the most commonly adopted in the literature on asymptotics in functional measurement error models; see, for example, Gleser (1983).

Obtaining an explicit expression for the asymptotic covariance matrix and asymptotic relative efficiency with respect to the MLE involves extensive, though not complicated calculations, which we omit here for brevity. These results and an extension of the MLq approach to the multivariate linear model are deferred to future research.

A simple reweighting algorithm can be derived to compute the MLqE $\hat{\boldsymbol{\theta}}_n^* = (\hat{\alpha}_n^*, \hat{\beta}_n^*, \hat{\phi}_n^*)^T$, following the strategy described in Section 4. The resulting equations are a weighted version of the maximum likelihood equations (Kimura (1992); Gleser (1981)). A derivation of the algorithm is included in Section S3.1 of the Supplementary Material.

Note that the MLqE of $\boldsymbol{\theta}$ is consistent for α and β, but not for ϕ. We have that $\eta_q(\hat{\boldsymbol{\theta}}_n^*) = (\hat{\alpha}_n^*, \hat{\beta}_n^*, k^{-1}\hat{\phi}_n^*)^T$, with $k = q - \frac{1}{2}$, is consistent for $\boldsymbol{\theta}$, where η_q corresponds to the mapping defined in Remark 2 of Section 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

In the next subsections, we analyze the performance of the estimators by means of a small simulation study and the analysis of a real data set.

6.1 Simulation study

We perform a small simulation study in order to investigate the efficiency and robustness of the MLqE under the pure model and various levels and types of contamination. We focus mainly on analyzing the behavior of the estimators when q varies. Owing to space considerations we do not include a simulation here to explore the empirical performance of the data-driven selection criterion for the tuning parameter under the different scenarios of contamination.

We used 1000 replications to estimate the empirical bias, variance, and MSE of the MLqE of parameter $\boldsymbol{\theta} = (\alpha, \beta, \phi)^T$. The MSE is the sum of the MSEs of the three individual estimators. For each simulation experiment, we choose the true values of the parameters to be $\alpha^0 = 0$, $\beta^0 = 1$, and $\phi^0 = 0.1$. We take the sample size as $n = 50$. We choose true covariates ξ_j randomly from the $N(0,1)$ distribution. To create different scenarios of contamination, we choose the random errors u_j and e_j as follows:

$$u_j \sim (1-\delta_x)N(0,\phi) + \delta_x N(2,\phi) \quad \text{and} \quad e_j \sim (1-\delta_y)N(0,\phi) + \delta_y N(2,\phi).$$

This gives the case where approximately $100\delta_x\%$ of the observations repre-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

sent leverage points, and $100\delta_y\%$ of the residuals are large. The case of a fully pure model with no artificially introduced large residuals or leverage points is obtained by fixing $\delta_x = 0$ and $\delta_y = 0$. Choosing $\delta_x = 0$ or $\delta_y = 0$, we have just large artificial residuals or leverage points, respectively. We fixed a grid for $0.5 < q \leq 1$ with increments of 0.01.

For brevity, we present only some results for the sample size $n = 50$ and $\phi = 0.1$. The results for other sample sizes and values of ϕ are qualitatively similar. In Table 1, the overall Monte Carlo MSE of the corrected MLqE of $\boldsymbol{\theta}$ is presented for several values of q and different levels of contamination given by combinations of δ_x and δ_y. The minimum values of MSE in the table are shown in bold, and $q_{min\,MSE}$ denotes the value of q that minimizes the Monte Carlo mean squared error.

As expected, at the true model ($\delta_x = 0$ and $\delta_y = 0$), the MSE of the corrected MLE ($q = 1$) is smaller than that of the MLqE with $q < 1$. This situation changes progressively as the level of contamination increases. By setting $q < 1$, we can successfully trade bias for variance and obtain a better estimation. We see that as the contamination increases, the MLE performs worst and the MSE of the MLqE tends to be smaller for values of q increasingly far from one. In Table 2, we compare the behavior of the MLE and the MLqE of the parameter β for $q = q_{min\,MSE}$, in terms of

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

bias, variance, and MSE, for different levels of contamination. The MLqE has bias noticeably smaller than that of the MLE under all considered contamination scenarios. There is also a variance reduction in the MLqE with respect to the MLE, except for percentages of contamination between 15% and 20%. Figure 1 shows the bias distribution of the MLqE of the parameter β for percentages of contamination with leverage points 5%, 10%, 15%, and 20% for various values of q. The box plots represent distributions of 1000 simulated values $\hat{\beta}_r - \beta^0$, for $r = 1, \ldots, 1000$. We can see, for example, that the MLqE of β can withstand up to 5% contamination for $q = 0.85$, 15% for $q = 0.8$, and almost 20% for $q = 0.75$. The behavior for the other model parameters with different values of ϕ is similar, though the optimal values of q increase and are closer to one when ϕ increases.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

Table 1: Simulated MSE of the corrected MLqE of $\boldsymbol{\theta}$, with $\phi = 0.1$ and sample size $n = 50$.

		q							
δ_x	δ_y	1	0.95	0.90	0.85	0.80	0.75	0.70	$q_{min\,MSE}$
0	0	**0.0092**	0.0094	0.0099	0.0109	0.0127	0.0169	0.0259	1
0.02	0	0.0185	0.0131	**0.0112**	0.0115	0.0131	0.0173	0.0266	0.89
0.05	0	0.0429	0.0277	0.0166	**0.0137**	0.0144	0.0191	0.0283	0.84
0.10	0	0.1007	0.0764	0.0472	0.0269	**0.0207**	0.0230	0.0328	0.79
0.15	0	0.1733	0.1493	0.1136	0.0703	0.0448	**0.0375**	0.0441	0.76
0.20	0	0.2465	0.2280	0.1989	0.1513	0.1001	0.0748	**0.0724**	0.71
0.02	0.02	0.0293	0.0187	0.0133	**0.0125**	0.0138	0.0186	0.0281	0.86
0.02	0.05	0.0612	0.0400	0.0227	0.0154	**0.0152**	0.0201	0.0289	0.83
0.05	0.05	0.0793	0.0558	0.0322	0.0192	**0.0177**	0.0223	0.0314	0.81
0.02	0.10	0.1500	0.1179	0.0772	0.0421	**0.0302**	0.0305	0.0402	0.78
0.05	0.10	0.1603	0.1324	0.0940	0.0539	0.0379	**0.0376**	0.0472	0.78
0.02	0.15	0.2799	0.2503	0.2045	0.1388	0.0917	**0.0779**	0.0791	0.73
0.05	0.15	0.2881	0.2647	0.2266	0.1627	0.1061	**0.0853**	0.0946	0.73

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

Table 2: Monte Carlo squared bias, variance, and MSE of the MLE and MLqE of β with $q = q_{min\,MSE}$, for $\phi = 0.1$ and sample size $n = 50$.

		MLE			MLqE			
δ_x	δ_y	bias^2	Var	MSE	bias^2	Var	MSE	q
0	0	0.0000	0.0049	0.0049	0.0000	0.0049	0.0049	1
0.02	0	0.0014	0.0075	0.0089	0.0000	0.0058	0.0058	0.89
0.05	0	0.0079	0.0101	0.0180	0.0000	0.0071	0.0071	0.84
0.10	0	0.0253	0.0126	0.0379	0.0001	0.0008	0.0109	0.79
0.15	0	0.0480	0.0128	0.0608	0.0007	0.0197	0.0204	0.76
0.20	0	0.0697	0.0126	0.0823	0.0031	0.0371	0.0402	0.71
0.02	0.02	0.0000	0.0126	0.0126	0.0000	0.0064	0.0064	0.86
0.02	0.05	0.0053	0.0195	0.0248	0.0001	0.0077	0.0078	0.83
0.05	0.05	0.0002	0.0228	0.0230	0.0000	0.0096	0.0096	0.81
0.02	0.10	0.0276	0.0315	0.0591	0.0003	0.0160	0.0163	0.78
0.05	0.10	0.0109	0.0352	0.0461	0.0004	0.0216	0.0220	0.78
0.02	0.15	0.0699	0.0400	0.1099	0.0020	0.0466	0.0486	0.73
0.05	0.15	0.0430	0.0517	0.0947	0.0022	0.0516	0.0538	0.73

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

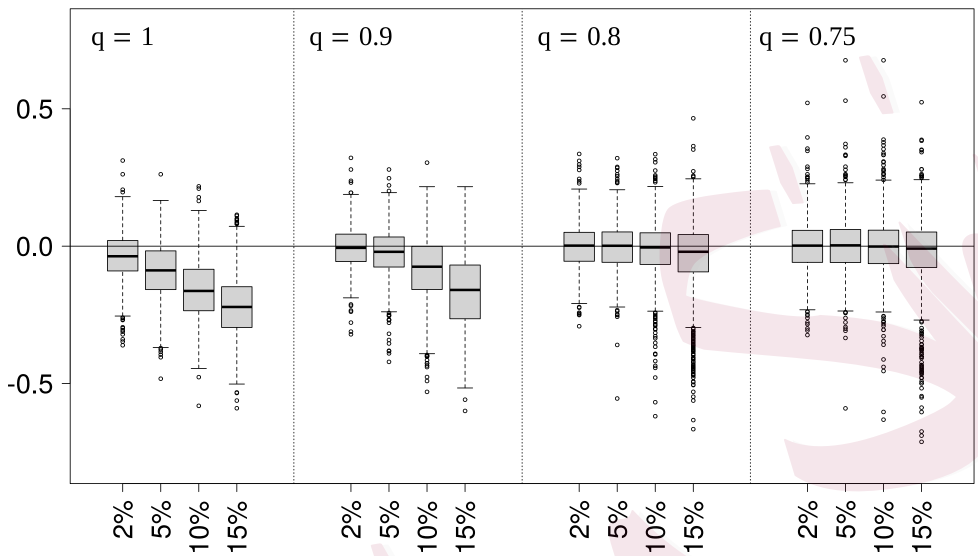

Figure 1: Bias distributions of the MLqE of β for values of q = 0.75, 0.80, 0.90, and 1 under a contaminated normal distribution with percentages of contamination 2%, 5%, 10%, and 15%, for $\phi = 1$ and $n = 50$.

6.2 Numerical example

In this section, we explore the numerical performance of the MLqE by analyzing a real data set. For all numerical calculations, the stopping rule for the reweighting algorithm is taken as $\delta = 10^{-6}$. The algorithm converges quickly, typically around 15 or 20 iterations, for values of q near the optimal.

We compare the robustness performance of the MLqE with respect to

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

that of the MLE, considering the consistent estimator of the parameter ϕ. We present a data-driven way of choosing the tuning parameter q by minimizing a bootstrap estimate of the MSE, as described in Section 5. We fixed a grid for $0.5 < q \leq 1$ with increments of 0.01, and based the calculations on 500 bootstrap repetitions.

Example: Hertzsprung–Russell data of the star cluster.

These data come from astronomy and are taken from Rousseeuw and Leroy (1987), who studied the robust least median of squares (LMS) estimator in the context of an ordinary simple linear regression. The Hertzsprung–Russell diagram of the star cluster CYG OB1 contains 47 stars in the direction of Cygnus. For these data, x is the logarithm of the effective temperature at the surface of the star (T_e), and y is the logarithm of its light intensity (L/L_0). Rousseeuw and Leroy (1987) inferred that there are two groups of data points; the four stars known as red giants (with indices 11, 20, 30, and 34) in the upper-left corner of the scatter plot (Fig. 2) represent a huge leverage point, and clearly form a separate group. The data were also analyzed by Ghosh and Basu (2013, 2015) for robust regression using the density power divergence approach. Here, we consider a simple linear regression with measurement errors (which corresponds to the case of model (3.4), with $p = 1$), taking into account that, in general, linear regression in

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

astronomy is characterized by measurement errors in the variables (Kelly (2007)). On the other hand, because the only stars of interest are those represented in the sample, we can consider that the true covariates are not a random sample from a large population. Then, it seems appropriate to consider a functional model. The parameters of the model are α, β, and ϕ. The assumption that both measurements are subject to random errors with equal variances seems reasonable. The MLqE of the model parameters for some values of q are presented in Table 3. It is clear that the MLE corresponding to $q = 1$ is highly sensitive to the presence of the four leverage points. We can see that for $0.7 \leq q \leq 0.95$, the MLqE is quite close, can ignore the outliers, and provides satisfactory fits. The optimal q obtained using the parametric bootstrap is $q = 0.88$, and the corresponding MLqE estimates are $\hat{\alpha} = -21.09$, $\hat{\beta} = 5.90$, and $\hat{\phi} = 0.01$. For comparison, we find the MLE and the MLqE after removing the four extreme outliers (cases 11, 20, 30, and 34) and the outlying point 7. The resulting estimates are presented in Table 4, and the fitted lines along with the MLE and the MLqE fits for the full data are plotted in Figure 2. We can see that the MLE fits without the outliers are very close to the MLqE fit for the full data with $q = 0.88$. Moreover, as expected, the MLqE fits after removing the outliers are very close to the MLE fits, with optimal values of q closer to one.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

Table 3: MLqE and number of iterations of the algorithm for the Hertzsprung–Russell data using several values of q.

q	1	0.95	0.90	0.85	0.80	0.75	0.70
$\hat{\alpha}$	35.43	-20.46	-21.01	-21.13	-21.03	-20.78	-20.44
$\hat{\beta}$	-7.06	5.77	5.88	5.91	5.88	5.83	5.75
$\hat{\phi}$	0.08	0.01	0.01	0.01	0.01	0.01	0.01
iter.	1	28	17	17	14	19	26

Table 4: MLE and MLqE for the Hertzsprung–Russell data using all data and with outliers removed.

	Using all data		*Removing cases 11, 20, 30, 34*		*Removing cases 7, 11, 20, 30, 34*	
Estimates	MLE	MLqE $q = 0.88$	MLE	MLqE $q = 0.90$	MLE	MLqE $q = 0.93$
$\hat{\alpha}$	35.43	-21.09	-18.26	-21.01	-20.75	-20.95
$\hat{\beta}$	-7.06	5.90	5.28	5.88	5.84	5.87
$\hat{\phi}$	0.08	0.01	0.01	0.01	0.01	0.01

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

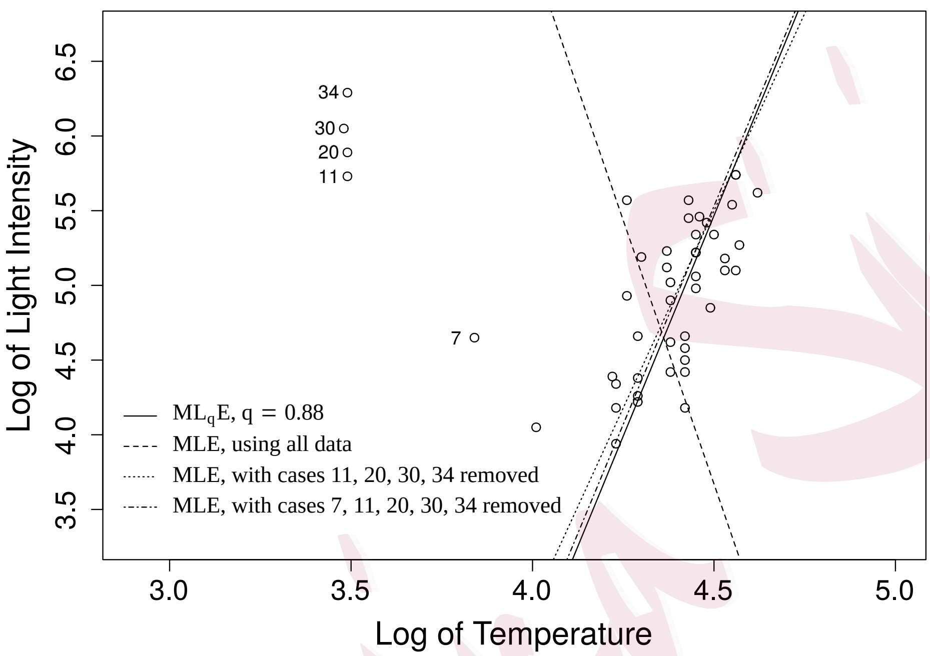

Figure 2: Scatter plot and four fitted regression lines for Hertzsprung–Russell data. MLE using all data and with outliers removed and MLqE using all data.

7. Conclusion

We have provided a robust fully parametric estimation procedure that extends maximum likelihood estimation in functional measurement error models. The method also extends the MLq approach to the more gen-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

eral problem of estimation where we have independent, but not identically distributed observations and the presence of incidental parameters. The method applies generally to any functional measurement error model, as long as the model family assumed is closed under a power transformation. We have established asymptotic properties of the MLqE under appropriate regularity conditions. The estimation procedure can be implemented easily by a simple and fast reweighting algorithm with well-established convergence behavior. The MLqE adapts according to the discrepancy between the data and the postulated model by tuning a single parameter q, which for $0 < q < 1$ controls the balance between robustness and efficiency. Choices of q near one afford considerable robustness, while retaining efficiency close to that of the MLE.

Our illustration of the methodology using a simple linear model showed that the MLqE is appealing in that it provides practitioners with a simple and fast estimation strategy with satisfactory fit to real-world data, while keeping a simple normal model. A simulation study and real-data analysis showed the satisfactory behavior of the MLqE and its advantages over the MLE in the presence of outlying observations and/or deviations from the assumed model. There can be no universal way of selecting an appropriate value of q in a given situation. We consider a data-based choice of an

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0414

appropriate value of q by minimizing a bootstrap estimate of the mean squared error of the estimators. However, other strategies for the optimal selection of q should be explored as well.

For the sake of brevity, we illustrated the proposed approach using a normal simple linear model. However, the results can be extended without difficulty to a multivariate functional linear model. In this case, the steps of the reweighting algorithm reduce to a simple variable transformation of the algorithm proposed by Gleser (1981).

Estimation for functional generalized linear measurement error models in canonical form when the explanatory vector is measured with an independent normal error could be addressed using the proposed methodology. For some of these models, the family of densities of $\boldsymbol{Z}_j = (\boldsymbol{X}_j^T, Y_j)^T$, given by (1.1), will be closed under the power transformation (2.3), as mentioned in Section 2.2. Moreover, for these models, we might consider replacing the incidental parameters $\boldsymbol{\xi}_j$ with uniform minimum variance unbiased estimators that depend on the structural parameters, as in Stefanski and Carroll (1987).

Furthermore, because the proposed approach uses a quite general incidental parameter framework, it can be applied in settings beyond functional measurement error models, such as panel, longitudinal, or clustered data models, when the likelihood can be fully specified.

In this study, we replace the incidental parameters in the Lq-likelihood with their estimates, which depend on the structural parameters. Moreover, in some models, the incidental parameters can be eliminated by conditioning on certain sufficient statistics (Andersen (1970)). Then, based on a conditional density independent of the incidental parameters, an MLq approach can be explored.

Another case of interest is the hypothesis testing problem. Asymptotics on MLqE can be used to develop a robust and efficient test of a hypothesis involving the structural parameter. Wald-type tests and Lq-likelihood-ratio-type tests, as introduced by Qin and Priebe (2017), can be considered.

Supplementary Material

The online Supplementary Material includes (i) the regularity conditions used to prove the asymptotic properties, (ii) proofs of Theorem 1 and Propositions 1 and 2, and (iii) derivations of some results in the simple linear functional model.

Acknowledgments

The authors acknowledge the partial financial support from the project

FONDECYT 1150325, Chile. We thank the referee for the many helpful suggestions.

References

Abdullah, M.B. (1989). Robust estimation of a linear functional relationship. *Journal of Statistical Computation and Simulation* **33**, 101–123.

Ammann, L. and Van Ness, J. (1988). A routine for converting regression algorithms into corresponding orthogonal regression algorithms. *ACM Transactions on Mathematical Software* **14**, 76–87.

Ammann, L. and Van Ness, J. (1989). Standard and robust orthogonal regression. *Communication in Statistics - Simulation and Computation* **18**, 145–162.

Andersen, E.B. (1970). Asymptotic Properties of Conditional Maximum-likelihood Estimators. *Journal of the Royal Statistical Society, B* **32**, 283–301.

Arslan, O. (2004). Convergence behavior of an iterative reweighting algorithm to compute m-estimates for location an scatter. *Journal of Statistical Planning and Inference* **118**, 115–128.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics* **5**, 445–463.

Buonaccorsi J.P. (2010). *Measurement Error: Models, Methods, and Applications*. Chapman and Hall, London.

Carroll, R.J. and Gallo, P.P. (1982). Some aspects of robustness in functional errors-in-variables regression models. *Communications in Statistics - Theory and Methods* **11**, 2573–2585.

Carroll, R., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd Ed. Chapman and Hall/CRC.

Cheng, C-L. and Van Ness, J.W. (1999). *Statistical Regression with Measurement Error*. Arnold, London.

Cressie, N. and Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, B* **46**, 440–464.

Croux, C., Fekri, M. and Ruiz-Gazen, A. (2010). Fast and robust estimation of the multivariate errors in variables model. *Test* **19**, 286–303.

Fekri, M. and Ruiz-Gazen, A. (2004). Robust weighted orthogonal regression in the errors-in-variables model. *Journal of Multivariate Analysis* **88**, 89–108.

Fekri, M. and Ruiz-Gazen, A. (2006). Robust estimation in the simple errors-in-variables models. *Statistics and Probability Letters* **70**, 1741–1747.

Ferrari, D. and La Vecchia, D. (2012). On robust estimation via pseudo-additive information. *Biometrika* **99**, 238–244.

Ferrari, D. and Yang, Y. (2010). Maximum Lq-likelihood estimation. *Annals of Statistics* **38**, 753–783.

Fuller, W.A. (1987). *Measurement error models*. Wiley, N.Y.

Galea, M. and de Castro, M. (2017). Robust inference in a linear functional model with replications using the t distribution. *Journal of Multivariate Analysis* **160**, 134–145.

Ghosh, A. and Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electronic Journal of Statistics* **7**, 2420–2456.

Ghosh, A. and Basu, A. (2015). Robust estimation for non-homogeneous data and the selection of the optimal tuning parameter: the density power divergence approach. *Journal of Applied Statistics* **42**, 2056–2072.

Giménez, P. and Bolfarine, H. (2000). Comparing consistent estimators in comparative calibration models. *Journal of Statistical Planning and Inference* **86**, 143–155.

Gleser, L.J. (1981). Estimation in a multivariate "error in variables" regression model: Large sample results. *The Annals of Statistics* **9**, 24–44.

Gleser, L.J. (1983). Functional, structural and ultrastructural errors-in-variables model. In *American Statistical Association Proceedings of the Business and Economic Statistics Section*, 24–44. American Statistical Assoc., Alexandria, VA.

Kelly, B.C. (2007). Some aspects of measurement error in linear regression of astronomical data. *The Astronomical Journal* **665**, 1489–1506.

Kimura, D.K. (1992). Functional comparative calibration using an EM algorithm. *Biometrics* **48**, 1263–1271.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics* **95**, 391–413.

Luong, A. and Mak, T.K. (1991). Robust estimation in a linear relationship model. *Communication in Statistics - Theory and Methods* **20**, 721–733.

Mak, T.K. (1982). Estimation in the presence of incidental parameters. *The Canadian Journal of Statistics* **10**, 121–132.

Nakamura, T. (1990). Corrected score function of errors-in-variables model: methodology and applications to generalized linear models. *Biometrika* **77**, 127–137.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.

Qin, Y. and Priebe, C.E. (2017). Robust hypothesis testing via Lq-likelihood. *Statistica Sinica* **27**, 1793–1013.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* John Wiley & Sons, New York.

Stefanski, L. (1985). The effects of measurement error on parameter estimation. *Biometrika* **72**, 583–592.

Stefanski, L. and Carroll, R.J. (1987). Conditional scores and optimal scores for generalized linear measurement error models. *Biometrika* **74**, 703–716.

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical*

doi:10.5705/ss.202019.0414

Physics **52**, 479–487.

Vilca-Labra, F.E., Bolfarine, H. and Arellano-Valle, R.B. (1998). Elliptical functional models. *Journal of Multivariate Analysis* **65**, 36–57.

Warwick, J. and Jones, M.C. (2005). Choosing a robustness tuning parameter. *Journal of the Statistical Computation and Simulation* **75**, 581–588.

Zamar, R.H. (1985). *Orthogonal Regression M-estimators*. Ph.D. Thesis, University of Washington, USA.

Centro Marplatense de Investigaciones Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, CIC, Funes 3350, Mar del Plata, Argentina.

E-mail: pcgimene@mdp.edu.ar

Centro Marplatense de Investigaciones Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, CIC, Funes 3350, Mar del Plata, Argentina.

E-mail: lucasguarracino@gmail.com

Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago, Chile.

E-mail: mgalea@mat.uc.cl