To appear in the *Journal of Nonparametric Statistics*
Vol. 00, No. 00, Month 20XX, 1–22

Two-step combined nonparametric likelihood estimation of misspecified semiparametric models

Francesco Bravo*

Department of Economics, University of York, York YO10 5DD, UK

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

This paper proposes to estimate possibly misspecified semiparametric estimating equations models using a two-step combined nonparametric likelihood method. The method uses in the first step the plug in principle and replaces the infinite dimensional parameter with a consistent estimator. In the second step an estimator for the finite dimensional parameter is obtained by combining exponential tilting with a another member of the empirical Cressie-Read discrepancy. The resulting class of semiparametric estimators are robust to misspecification and have the same asymptotic variance as that of the efficient semiparametric generalised method of moment estimator under correct specification. It is also shown that the asymptotic distributions of the proposed estimators can be consistently estimated by a multiplier bootstrap procedure. The results of the paper are illustrated with a quadratic inference function model and an instrumental variable partially linear additive model. Monte Carlo evidence suggests that the proposed estimators have competitive finite sample properties.

Keywords:
Exponential tilting, Stochastic equicontinuity, Series estimation

AMS Subject Classification: 62G05; 62G09; 62G20

1. Introduction

In this paper we propose a novel estimation method for semiparametric estimating equations models, also known as moment conditions models in the econometric literature, where the parameter of interest is finite dimensional and the nuisance parameter is infinite dimensional. These models are rather general and include, for example, semiparametric extensions to generalised instrumental variables models that are often used in the economic and financial literature - see for example Hansen and Singleton (1982) - and to generalised estimating equations and quadratic inference functions models that are very popular in the statistical literature - see for example Liang and Zeger (1986) and Qu, Lindsay, and Li (2000). One important feature of the method we propose is its robustness to misspecification of the estimating equations themselves. By robust to misspecification we mean that the proposed estimators are characterised by the standard $n^{1/2}$ convergence rate. Note also that the misspecification considered here is a global one (see (1) below) and not the local one recently investigated by Kitamura, Otsu, and Evdokimov (2013), who proposed an estimation method that is robust to local misspecification and yields estimators with an asymptotic minimax property.

Misspecified estimating equations models are theoretically interesting and empirically

* Email: francesco.bravo@york.ac.uk, web page: https://sites.google.com/a/york.ac.uk/francescobravo

relevant. For example, many asset pricing models are likely to be misspecified and, as shown for example by Gospodinov, Kan, and Robotti (2013) and Gospodinov, Kan, and Robotti (2014), using standard statistical inferences in such models could be very misleading. Hall and Inoue (2003) and Grace and Reid (2010) provide other examples of misspecified estimating equations models.

The method we propose is an alternative to generalised method of moments (GMM) (Hansen 1982) that is based on a two-step version of the empirical Cressie-Read (ECR) discrepancy approach, originally introduced by Baggerly (1998) as a generalisation of Owen (1988)'s empirical likelihood, which includes other well-known nonparametric likelihoods including exponential tilting (see for example Kitamura and Stutzer (1997)), Pearson's χ^2 and Euclidean likelihood (see for example Owen (1991)). In the context of correctly specified estimating equations, estimators based on the ECR approach are asymptotically equivalent to those based on the efficient GMM approach (see for example Newey and Smith (2004)). This asymptotic equivalence is however lost with misspecified estimating equations, since misspecification affects ECR and GMM estimators in a very different way: GMM estimators are robust to misspecification but their asymptotic distribution depends on both the weighting matrix used in the estimation process and the gradient of the estimating equations - see for example Hall and Inoue (2003) and Theorem 3.3 below. On the other hand, the robustness of ECR estimators crucially depends on which nonparametric likelihood is used in the estimation process; for example Schennach (2007) showed that the empirical likelihood estimator is not robust to misspecification. This unusual behaviour can be explained informally by noting that in the case of empirical likelihood the so-called implied probabilities - see (5) below for a definition, can diverge causing a singularity in the first order conditions defining the estimator. Since empirical likelihood corresponds to the ECR discrepancy with the user specific real valued parameter $\gamma = -1$, the same informal argument can be used to deduce that no ECR estimators based on any $\gamma < 0$ is in fact robust to misspecification. On the other hand, ECR estimators based on $\gamma \geq 0$ (for example Kitamura and Stutzer (1997)'s exponential tilting and Owen's (1991) Euclidean likelihood) are robust to misspecification, but can result (for $\gamma > 0$) in implied probabilities that are not range preserving (i.e. they might be negative or bigger than 1) - see Baggerly (1998), which is clearly not a desirable property for inference. These two facts naturally lead to the combined ECR (CECR) discrepancy approach considered in this paper. The estimator we propose combines exponential tilting and another member of the ECR discrepancy defined by a nonpositive γ; for example, for $\gamma = -1$ the CECR estimator corresponds to the exponential tilting empirical likelihood estimator of Schennach (2007), for $\gamma = -1/2$ the CECR estimator is the exponential tilting Hellinger estimator, and $\gamma = -2$ the CECR is the exponential tilting Pearson's χ^2 estimator. The estimator is computed using an iterative process based on a nested algorithm in which in the inner stage exponential tilting is used to estimate the implied probabilities, while in the outer stage the other chosen member of the ECR discrepancy is used to estimate the unknown parameter of interest.

To deal with the semiparametric nature of the estimating equations considered in this paper, we propose a two-step version of the CECR estimator: the first step is used to obtain a consistent nonparametric estimator of the infinite dimensional (nuisance) parameter, whereas the second step is used to estimate the finite dimensional parameter of interest. The resulting two-step semiparametric CECR estimator is robust to misspecification as is the two-step semiparametric GMM estimator, but as opposed to the latter its asymptotic distribution does not depend on any weighting matrix nor on the gradient of the estimating equations. At the same time, the two-step semiparametric CECR estimator is asymptotically equivalent to the efficient two-step semiparametric GMM estimator (see for example Ackerberg, Chen, Hahn, and Liao (2014)) under correct specification and an asymptotic orthogonality condition given in Assumption A3(ii) in Section 3 below. These two characteristics imply that the proposed estimation method has two important advantages over GMM: first, as opposed to the GMM estimates that under misspecification depend on the chosen weighting matrix, the CECR estimates are unique. Second, because estimation of the weighting matrix is a source of bias (Newey and Smith 2004), the CECR estimator should typically have better finite sample properties, a claim that will be confirmed in the simulation study reported in Section 5.

In this paper we make the following contributions: First, we establish the asymptotic normality of the proposed two-step semiparametric CECR and GMM estimators that have not been considered in the literature before. This result is rather general because it considers the cases of the first step estimation either affecting or not affecting the asymptotic variance of the finite dimensional parameter estimator. This result complements and/or extends results of Andrews (1994), Newey (1994), Chen, Linton, and van Keilegom (2003), Hjort, McKeague, and VanKeilegom (2009), Bravo (2009) and Bravo, Escanciano, and Van Keilegom (2020) among others. It is important to note that all of these papers consider correctly specified and exactly identified semiparametric estimating equations and thus their results cannot be directly applied to the misspecified estimating equations considered in this paper.

Second, we consider the same weighted bootstrap used by Jin, Ying, and Wei (2001), Ma and Kosorok (2005), and more recently by Lavergne and Patilea (2013) among others and propose a weighted bootstrap procedure that can be used as an alternative to the standard bootstrap to obtain the standard errors and more generally to approximate the distributions of all of the estimators considered. Bootstrapping the standard errors seems particularly useful in the context of the semiparametric estimators of this paper, given the complicated structure of their asymptotic variances - see Section 5 below and Sections 3.1 and 3.2 in the supplemental appendix for two illustrative examples. The proposed weighted bootstrap procedure, often called multiplier bootstrap in the statistical literature, does not require the semiparametric model to be in reduced form and it is easy to implement, as it is based on randomly perturbing the objective function and then recomputing the estimators.

Finally, we illustrate the results of the paper both theoretically and numerically by considering two models that have not been previously considered in the semiparametric literature: a semiparametric extension of a quadratic inference functions (QIF) model and an instrumental variable partially linear additive model. These results extend, among others, those obtained by Li (2000), Bai, Zhu, and Fung (2008) and Lai, Li, and Lian (2013).

The rest of the paper is structured as follows: next section introduces the statistical model and the estimators. Section 3 develops the asymptotic theory; Sections 4 and 5, respectively, illustrate the theory with the two examples and report the results of the Monte Carlo simulations. Section 6 contains some concluding remarks.

The following notation is used throughout the paper: "$'$" indicates transpose, "$^-$" denotes the generalised inverse of a matrix, "$\otimes$" denotes Kronecker product, "$\|\cdot\|$" and "$\|\cdot\|_{\mathcal{F}}$" denote, respectively, the standard Euclidean (Frobenius) norm for random vectors (matrices) and a functional norm such as the sup norm for a pseudo-metric space of functions $\mathcal{F}$, "tr", "vec" are the trace and vec operators, and finally for any vector v, $v^{\otimes 2} = vv'$.

2. The model and estimators

Let $\{Z_i\}_{i=1}^n$ denote a random sample from the distribution P of $Z \in \mathcal{Z} \subset \mathbb{R}^{d_z}$, $\theta \in \Theta \subset \mathbb{R}^k$ denote a vector of unknown finite dimensional parameters, Θ is a compact set, and $h(Z) := h \in \mathcal{H} = \mathcal{H}_1 \times .... \times \mathcal{H}_m$ is a vector of unknown functions and $\mathcal{H}$ is a pseudo-metric space of functions. The statistical model we consider is

$$E\left[g\left(Z, \theta, h_\bullet\right)\right] = \mu\left(\theta, h_\bullet\right) \quad \text{for all } \theta \in \Theta, \tag{1}$$

where $g(\cdot) : \mathcal{Z} \times \Theta \times \mathcal{H} \to \mathbb{R}^l$ $(l \geq k)$ is a vector-valued known function, $\mu(\theta, h_\bullet)$ is the indicator of misspecification such that $\inf_{\theta \in \Theta} \|\mu(\theta, h_\bullet)\| > 0$, and $h_\bullet$ is an element of $\mathcal{H}$ that can be either the true parameter h_0 or the pseudo-true parameter h_* in case of misspecification of h itself.

To introduce the two-step semiparametric estimators of this paper it is useful to assume that (1) is correctly specified, that is $\mu(\theta, h_0) = 0$ for $\theta = \theta_0$, and that there exists a preliminary nonparametric estimator $\widehat{h}$ of h_0. One possible way to estimate θ_0 is to use GMM, with the resulting two-step semiparametric GMM estimator defined as

$$\widehat{\theta}_{GMM} = \arg\min_{\theta \in \Theta} \widehat{Q}_{\widehat{W}}\left(\theta, \widehat{h}\right), \tag{2}$$

where

$$\widehat{Q}_{\widehat{W}}\left(\theta, \widehat{h}\right) = \frac{1}{n}\sum_{i=1}^n g\left(Z_i, \theta, \widehat{h}\right)' \widehat{W} \frac{1}{n}\sum_{i=1}^n g\left(Z_i, \theta, \widehat{h}\right),$$

and $\widehat{W}$ is a possibly random positive semidefinite weighting matrix.

An alternative method of estimating θ_0 can be based on the nonparametric likelihood approach, which consists of finding among the set of multinomial distributions $\{\pi_i(\theta_0)\}_{i=1}^n$ supported on the sample $\{Z_i\}_{i=1}^n$ the one closest to the empirical distribution function, that is the nonparametric maximum likelihood estimator. In the case of the ECR approach, this amounts to use the Cressie-Read power divergence criterion and solve the generic program

$$\min_{\pi_i(\theta)} \left\{ \sum_{i=1}^n \frac{(n\pi_i(\theta))^{\gamma+1} - 1}{\gamma(\gamma+1)} | \sum_{i=1}^n \pi_i(\theta) = 1, \ \sum_{i=1}^n \pi_i(\theta) g\left(Z_i, \theta, \widehat{h}\right) = 0 \right\}, \quad \gamma \in \mathbb{R}, \tag{3}$$

where γ is a user-specific parameter and the values $\gamma = -1$ and $\gamma = 0$, corresponding to empirical likelihood and exponential tilting, should be interpreted as limits. By a Lagrange multiplier argument, it is possible to show that the solution to (3) results in the so-called profile ECR function

$$\Gamma^{ECR}\left(\theta, \widehat{\lambda}, \widehat{h}\right) = \sum_{i=1}^n \frac{\left(1 + \gamma \widehat{\lambda}' g\left(Z_i, \theta, \widehat{h}\right)\right)^{\frac{\gamma+1}{\gamma}}}{\gamma + 1}, \tag{4}$$

where the estimated Lagrange multiplier $\widehat{\lambda}$ is associated with the restriction $\sum_{i=1}^n \pi_i(\theta) g\left(Z_i, \theta, \widehat{h}\right) = 0$. The two-step semiparametric ECR estimator of $\theta = \theta_0$

is then defined as $\widehat{\theta}_{ECR} = \arg\min_{\theta\in\Theta} \Gamma^{ECR}\left(\theta, \widehat{\lambda}, \widehat{h}\right)$. Crucially, under correct specification the asymptotic distribution of $n^{1/2}\left(\widehat{\theta}_{ECR} - \theta_0\right)$ does not depend on the value of γ, albeit some care should be taken if using the implied probabilities

$$\pi_i\left(\widehat{\theta}\right) = \frac{\left(1+\gamma\widehat{\lambda}'g\left(Z_i,\widehat{\theta},\widehat{h}\right)\right)^{\frac{1}{\gamma}}}{\sum_{j=1}^{n}\left(1+\gamma\widehat{\lambda}'g\left(Z_j,\widehat{\theta},\widehat{h}\right)\right)^{\frac{1}{\gamma}}} \tag{5}$$

for inference, since they are range preserving only for nonpositive values of γ, see for example Baggerly (1998). On the other hand, as mentioned in the Introduction, under global misspecification the value of γ becomes crucial, because only nonnegative values of γ yield estimators that are robust to misspecification. Note also that we use the notation $\lambda(\theta)$ to emphasise the dependence of the Lagrange multiplier λ on θ. This dependence is not relevant in the case of correctly specified estimating equations models as $\widehat{\lambda}(\theta)$ converges to 0 uniformly in $\theta \in \Theta$, but becomes very important under global misspecification since $\widehat{\lambda}(\theta)$ converges to a nonzero limit that depends on the pseudo-true value $\theta_{\gamma*}$ defined in the next section.

The two-step semiparametric CERC estimator we propose is robust to misspecification and has range preserving implied probabilities. The estimator is defined as

$$\widehat{\theta}_{CECR} = \arg\min_{\theta\in\Theta}\sum_{i=1}^{n}\frac{\left(n\pi_i^e\left(\theta,\widehat{\lambda}(\theta),\widehat{h}\right)\right)^{\gamma+1}-1}{\gamma(\gamma+1)} \tag{6}$$

for a given γ chosen in the range $(-\infty, 0]$, where exponential tilting is used to obtain the implied probabilities

$$\pi_i^e\left(\theta,\widehat{\lambda}(\theta),\widehat{h}\right) = \frac{\exp\left(\widehat{\lambda}(\theta)'g\left(Z_i,\theta,\widehat{h}\right)\right)}{\sum_{j=1}^{n}\exp\left(\widehat{\lambda}(\theta)'g\left(Z_j,\theta,\widehat{h}\right)\right)}$$

with

$$\widehat{\lambda}(\theta) = \arg\max_{\lambda(\theta)\in\Lambda(\Theta)} -\sum_{i=1}^{n}\exp\left(\lambda(\theta)'g\left(Z_i,\theta,\widehat{h}\right)\right). \tag{7}$$

3. Asymptotic theory

It is important to define the so-called pseudo-true value θ_*, that is the parameter of interest in misspecified estimating equations models. For the semiparametric two-step GMM estimator (2), the pseudo-true value θ_* is defined as the unique minimiser of the probability limit of the objective function $\widehat{Q}_{\widehat{W}}\left(\theta,\widehat{h}\right)$, that is

$$\theta_*(W) := \theta_* = \arg\min_{\theta\in\Theta}\mu(\theta,h_\bullet)'W\mu(\theta,h_\bullet),$$

where the definition emphasises the fact that the pseudo-true value θ_* depends on the matrix W, the (assumed positive definite) probability limit of the matrix $\widehat{W}$ defined in (2) - see Assumption A5(i) in the supplemental appendix for details. For the semiparametric two-step CECR estimator (6), the pseudo-true value θ_* is defined as the unique minimiser of the expected Cressie-Read discrepancy, that is

$$\theta_{\gamma *} = \arg\min_{\theta\in\Theta} E\left[\frac{\left(n\pi^e\left(\theta, \lambda_*\left(\theta\right), h_\bullet\right)\right)^{\gamma+1} - 1}{\gamma\left(\gamma+1\right)}\right],$$

where the definition emphasises the fact that pseudo-true value $\theta_{\gamma *}$ depends on the chosen $\gamma \in (-\infty, 0]$, that is on the chosen member of the Cressie-Read divergence, and

$$\lambda_*\left(\theta\right) = \arg\max_{\lambda(\theta)\in\Lambda(\Theta)} -E\left[\exp\left(\lambda\left(\theta\right)' g\left(Z, \theta, h_\bullet\right)\right)\right], \tag{8}$$

which exists and is unique by the strict concavity of $-E\left[\exp\left(\lambda\left(\theta\right)' g\left(Z, \theta, h_\bullet\right)\right)\right]$ in $\lambda\left(\theta\right)$. Note also that both definitions of pseudo true-value implicitly depend on $h_\bullet$ - the probability limit of $\widehat{h}$ - and thus do not rely on the correct specification of h itself.

The main consequence of misspecification can be seen in the asymptotic variances of both the two-step CECR and GMM estimators, which are more complicated than the corresponding ones obtained under correct specification (compare (11) and (14) below with $(A.1)$ and $(A.2)$ in the supplemental appendix). In particular, for the CECR estimator the difference can be explained by the fact that under misspecification the Lagrange multiplier estimator converges to a nonzero vector that implicitly depends on the unknown parameter of interest, and this dependency introduces additional terms in the corresponding asymptotic variance - see for example Sections 3.1 and 3.2 in the supplemental appendix. For the GMM estimator the difference can be explained by the fact that under misspecification both the gradient of the estimating equations and the weighting matrix contributes to the asymptotic normality of the estimator - see Theorem 3.3 below for more details.

We introduce some additional notation: for any random scalar, vector or matrix $v\left(Z, \theta, h\right)$, let $v\left(Z, \theta, h\right) = v\left(\theta, h\right)$, $v\left(Z_i, \theta, h\right) = v_i\left(\theta, h\right)$ and $\widehat{v}\left(\theta, h\right) = \sum_{i=1}^n v_i\left(\theta, h\right)/n$, so for example $g\left(Z_i, \theta, h\right) = g_i\left(\theta, h\right)$, and $\widehat{g}\left(\theta, h\right) = \sum_{i=1}^n g_i\left(\theta, h\right)/n$. For any $\overline{h}$, $h \in \mathcal{H}$, we say that $v\left(\theta, h\right)$ is pathway differentiable at h in the direction of $\left[\overline{h} - h\right]$ if, for $\tau \in [0, 1]$ and $\left\{h + \tau\left(\overline{h} - h\right)\right\} \subset \mathcal{H}$,

$$\lim_{\tau\to 0} \frac{\left(v\left(\theta, h + \tau\left(\overline{h} - h\right)\right) - v\left(\theta, h\right)\right)}{\tau} := v_h\left(\theta, h\right)\left[\overline{h} - h\right] \tag{9}$$

exists. Let

$$s\left(\theta, \lambda, h\right) = \exp\left(\lambda\left(\theta\right)' g\left(\theta, h\right)\right) \text{ and } \rho_\gamma\left(\theta, \lambda, h\right) = \frac{\left(n\pi^e\left(\theta, \lambda\left(\theta\right), h\right)\right)^{\gamma+1} - 1}{\gamma\left(\gamma+1\right)}, \tag{10}$$

and

$$
\begin{aligned}
R_\gamma(\theta,\lambda,h) &= E\left\{\frac{\partial^2\left[\rho_\gamma(\theta,\lambda,h),s(\theta,\lambda,h)\right]'}{\partial(\theta',\lambda')'^{\otimes 2}}\right\},\\
R_\gamma(\theta,\lambda,h)^{-1} &:= Q_\gamma(\theta,\lambda,h) = \begin{bmatrix} Q^\gamma_{\theta\theta}(\theta,\lambda,h) & Q^\gamma_{\theta\lambda}(\theta,\lambda,h)\\ Q^s_{\lambda\theta}(\theta,\lambda,h) & Q^s_{\lambda\lambda}(\theta,\lambda,h)\end{bmatrix},
\end{aligned}
$$

$$
\begin{aligned}
\Xi_\gamma(\theta,\lambda,h) &= E\left\{\begin{bmatrix}\frac{\partial\rho_\gamma(\theta,\lambda,h)}{\partial\theta}\\ \frac{\partial s(\theta,\lambda,h)}{\partial\lambda}\end{bmatrix}^{\otimes 2}\right\} = \begin{bmatrix}\Xi^\gamma_{\theta\theta}(\theta,\lambda,h) & \Xi^\gamma_{\theta\lambda}(\theta,\lambda,h)\\ \Xi^s_{\lambda\theta}(\theta,\lambda,h) & \Xi^s_{\lambda\lambda}(\theta,\lambda,h)\end{bmatrix},\\
\Xi_\gamma(\theta,\lambda,h,\delta) &= E\left\{\begin{bmatrix}\frac{\partial\rho_\gamma(\theta,\lambda,h)}{\partial\theta}+\delta_\gamma\\ \frac{\partial s(\theta,\lambda,h)}{\partial\lambda}+\delta_s\end{bmatrix}^{\otimes 2}\right\} = \begin{bmatrix}\Xi^\gamma_{\theta\theta}(\theta,\lambda,h,\delta) & \Xi^\gamma_{\theta\lambda}(\theta,\lambda,h,\delta)\\ \Xi^s_{\lambda\theta}(\theta,\lambda,h,\delta) & \Xi^s_{\lambda\lambda}(\theta,\lambda,h,\delta)\end{bmatrix},
\end{aligned}
$$

where δ_γ and δ_s denote, respectively, the probability limits of the asymptotic representations of the pathwise derivatives of $\partial\rho_\gamma(\theta_{\gamma*},\lambda_*,h_\bullet)/\partial\theta$ and of $\partial s(\theta_{\gamma*},\lambda_*,h_\bullet)/\partial\lambda$ satisfying the regularity condition given in A4 below.

Theorem 3.1 establishes the asymptotic distribution of the two-step semiparametric CECR estimator (6) for $\theta_{\gamma*}$ under some general regularity conditions (A0-A3(i) stated in the supplemental appendix), Assumptions A3(ii) and A4 below and some further standard regularity conditions (i-iv) commonly assumed in the nonlinear statistical estimation literature (see for example Van der Vaart (1998)). Assumptions A3(ii) and A4 are important because they explain the different structures of the asymptotic variances matrices $\Phi_{CECR}(\theta_{\gamma*},\lambda_*,h_\bullet)$ and $\Phi_{CECR}(\theta_{\gamma*},\lambda_*,h_\bullet,\delta)$ appearing in Theorem 3.1. Assume that:

either

A3 (ii) $E\left[\partial\rho_\gamma(\theta_{\gamma*},\lambda_*,h)/\partial\theta',\partial s(\theta_{\gamma*},\lambda_*,h)/\partial\lambda'\right]'|_{h=\widehat{h}} = o_p\left(n^{-1/2}\right),$

or

A4 (i) the pathwise derivatives $\partial\rho_{\gamma h}(\theta_{\gamma*},\lambda_*,h_\bullet)/\partial\theta\,[h-h_\bullet]$ of $\partial\rho_\gamma(\theta_{\gamma*},\lambda_*,h_\bullet)/\partial\theta$ and $\partial s_h(\theta_{\gamma*},\lambda_*,h_\bullet)/\partial\lambda\,[h-h_\bullet]$ of $\partial s(\theta_{\gamma*},\lambda_*,h_\bullet)/\partial\lambda$ exist *a.s.* in all directions $[h-h_\bullet]$ and for small enough $\|h-h_\bullet\|_{\mathcal{H}}$

$$
\begin{aligned}
\left\|\frac{\partial\rho_\gamma(\theta_{\gamma*},\lambda_*,h)}{\partial\theta}-\frac{\partial\rho_\gamma(\theta_{\gamma*},\lambda_*,h_\bullet)}{\partial\theta}-\frac{\partial\rho_{\gamma h}(\theta_{\gamma*},\lambda_*,h_\bullet)}{\partial\theta}[h-h_\bullet]\right\| &\le b_\gamma(Z)\|h-h_\bullet\|^2_{\mathcal{H}},\\
\left\|\frac{\partial s(\theta_{\gamma*},\lambda_*,h)}{\partial\lambda}-\frac{\partial s(\theta_{\gamma*},\lambda_*,h_\bullet)}{\partial\lambda}-\frac{\partial s_h(\theta_{\gamma*},\lambda_*,h_\bullet)}{\partial\lambda}[h-h_\bullet]\right\| &\le b_s(Z)\|h-h_\bullet\|^2_{\mathcal{H}}
\end{aligned}
$$

with $E|b_\gamma(Z)|<\infty$ and $E|b_s(Z)|<\infty$, (ii) there exist two functions $\delta_\gamma(\cdot):\mathcal{Z}\to\mathbb{R}^k$ and $\delta_s(\cdot):\mathcal{Z}\to\mathbb{R}^l$ with $E[\delta_\bullet(Z)]=0$ and $E\left[\|\delta_\bullet(Z)\|^2\right]<\infty$ such that

$$
\begin{aligned}
\left\|\frac{\partial\widehat{\rho}_{\gamma h}(\theta_{\gamma*},\lambda_*,h_\bullet)}{\partial\theta}\left[\widehat{h}-h_\bullet\right]-\widehat{\delta}_\gamma(Z)\right\| &= o_p\left(n^{-1/2}\right),\\
\left\|\frac{\partial\widehat{s}_h(\theta_{\gamma*},\lambda_*,h_\bullet)}{\partial\lambda}\left[\widehat{h}-h_\bullet\right]-\widehat{\delta}_s(Z)\right\| &= o_p\left(n^{-1/2}\right),
\end{aligned}
$$

where $\widehat{\delta}_\bullet(Z) := \sum_{i=1}^n \delta_\bullet(Z_i)/n$ with $\widehat{\delta}_\bullet(Z) \xrightarrow{p} \delta_\bullet(Z)$ and "$\bullet$" is either γ or s.

Assumption A3(ii) can be interpreted as an asymptotic orthogonality condition implying that estimation of $h_\bullet$ does not affect the variance of the asymptotic distribution of the estimator of θ_*. With correctly specified estimating equations (see Section 1 in the supplemental appendix), it ensures that the proposed two-step estimators are efficient, in the sense that they achieve the same efficiency bound as that given by Chamberlain (1987) for GMM estimators of unknown finite dimensional parameters and by Ackerberg et al. (2014) for two-step semiparametric GMM estimators. Note that A3(ii) is satisfied by a number of important semiparametric models including partial linear, additive and single index models, see Newey (1994) for further examples. Assumption A4 is more technical: A4(i) implies effectively Frechet differentiability of $\partial\rho_\gamma(\theta_{\gamma*}, \lambda_*, h_\bullet)/\partial\theta$ and of $\partial s(\theta_{\gamma*}, \lambda_*, h_\bullet)/\partial\lambda$. A4(ii) characterises the estimation effect from the first step estimation in terms of the random vectors $\widehat{\delta}_\bullet(Z)$ and their limits $\delta_\bullet(Z)$. It can often be verified if h is an unknown density or regression function and $\widehat{h}$ is a kernel or a nonparametric series estimator, see for example Chen and Liao (2015), Bravo et al. (2020) and the proof of Proposition 4.1 in the supplemental appendix.

THEOREM 3.1 *Assume that for $\lambda_*(\theta)$ defined in (8) (i) $E[\rho_\gamma(\theta, \lambda_*, h_\bullet)]$ has a unique minimum at $\theta_{\gamma*} \in \Theta$ (ii) $\theta_{\gamma*} \in int(\Theta)$, $\lambda_*(\theta) \in int(\Lambda(\Theta))$, with both Θ and $\Lambda(\Theta)$ compact sets (iii) $R_\gamma(\theta_{\gamma*}, \lambda_*, h_\bullet)$ is nonsingular (iv) $\Xi_\gamma(\theta_{\gamma*}, \lambda_*, h_\bullet)$ and $\Xi_\gamma(\theta_{\gamma*}, \lambda_*, h_\bullet, \delta)$ are positive definite. Then, under either (v) A0 (with $\vartheta = 0$), A1-A3(i) (in the supplemental appendix) and A3(ii)*

$$n^{1/2}\left(\widehat{\theta}_{CECR} - \theta_{\gamma*}\right) \xrightarrow{d} N\left(0, \Phi_{CECR}(\theta_{\gamma*}, \lambda_*, h_\bullet)\right),$$

where

$$\begin{aligned}\Phi_{CECR}(\theta_{\gamma*}, \lambda_*, h_\bullet) = {} & Q^\gamma_{\theta\theta}(\theta_{\gamma*}, \lambda_*, h_\bullet)\,\Xi^\gamma_{\theta\theta}(\theta_{\gamma*}, \lambda_*, h_\bullet)\,Q^\gamma_{\theta\theta}(\theta_{\gamma*}, \lambda_*, h_\bullet) + \\ & Q^\gamma_{\theta\lambda}(\theta_{\gamma*}, \lambda_*, h_\bullet)\,\Xi^s_{\lambda\theta}(\theta_{\gamma*}, \lambda_*, h_\bullet)\,Q^\gamma_{\theta\theta}(\theta_{\gamma*}, \lambda_*, h_\bullet) + \\ & Q^\gamma_{\theta\theta}(\theta_{\gamma*}, \lambda_*, h_\bullet)\,\Xi^\gamma_{\theta\lambda}\rho(\theta_{\gamma*}, \lambda_*, h_\bullet)\,Q^\gamma_{\theta\lambda}(\theta_{\gamma*}, \lambda_*, h_\bullet)' + \\ & Q^\gamma_{\theta\lambda}(\theta_{\gamma*}, \lambda_*, h_\bullet)\,\Xi^s_{\lambda\lambda}(\theta_{\gamma*}, \lambda_*, h_\bullet)\,Q^\gamma_{\theta\lambda}(\theta_{\gamma*}, \lambda_*, h_\bullet)',\end{aligned} \quad (11)$$

or (vi) A0 (with $\vartheta = 1/4$), A1-A2 (in the supplemental appendix) and A4

$$n^{1/2}\left(\widehat{\theta}_{CECR} - \theta_{\gamma*}\right) \xrightarrow{d} N\left(0, \Phi_{CECR}(\theta_{\gamma*}, \lambda_*, h_\bullet, \delta)\right), \quad (12)$$

where $\Phi_{CECR}(\theta_{\gamma*}, \lambda_*, h_\bullet, \delta)$ is as that defined in (11) with $\Xi_\gamma(\theta, \lambda, h, \delta)$ replacing $\Xi_\gamma(\theta, \lambda, h)$.

We now consider the weighted bootstrap procedure mentioned in the Introduction. Let $\{\omega_i\}_{i=1}^n$ denote an i.i.d. sample of positive random weights independent of $\{Z_i\}_{i=1}^n$ with $E(\omega_i) = 1$ and $Var(\omega_i) = v_0 < \infty$. Let

$$\widehat{\theta}_{\omega,CECR} = \arg\min_{\theta\in\Theta} \sum_{i=1}^n \omega_i \rho_{\gamma i}\left(\theta, \widehat{\lambda}_\omega, \widehat{h}\right) \quad (13)$$

denote the weighted bootstrap version of (6), where $\rho_{\gamma i}(\theta, \lambda, h) =$

$\left(\left(n\pi_i^e\left(\theta,\lambda\left(\theta\right),h\right)\right)^{\gamma+1}-1\right)/\gamma\left(\gamma+1\right)$ and $\widehat{\lambda}_\omega\left(\theta\right)$ is the weighted bootstrap analogue of (7). The following theorem shows that by drawing B times $\{\omega_i\}_{i=1}^n$ samples, the weighted bootstrap distribution of $\widehat{\theta}_{\omega,CECR}$, conditional on the observations, imitates asymptotically the unconditional one and hence can be used for inference.

Theorem 3.2 *Under the same assumptions of Theorem* 3.1*, then conditionally on the sample* $\{Z_i\}_{i=1}^n$,

$$\sup_u\left|\Pr{}^\omega\left(\left(\frac{n}{\upsilon_0}\right)^{1/2}\left(\widehat{\theta}_{\omega,CECR}-\widehat{\theta}_{CECR}\right)\leq u\right)-\Pr\left(n^{1/2}\left(\widehat{\theta}_{CECR}-\theta_{\gamma*}\right)\leq u\right)\right|=o_p\left(1\right),$$

where we use the notation $\Pr^\omega$ *to emphasise that the resampling is over the* ω_i*'s.*

Remark 1. Theorem 3.1 defines a class of estimators indexed by the user-specific parameter $\gamma\in(-\infty,0]$ that are all robust to misspecification. The theorem however does not provide any guidance on the choice of γ. One possibility would be to use second order asymptotic analysis and compare estimators in terms of their second order bias and mean squared error. Unfortunately in the context of (possibly misspecified) semiparametric estimating equations models, second order analysis is complicated by two key facts: First, the nonparametric estimator used to estimate $h_\bullet$ becomes important because its second order properties (if at all known) affect in a different way those of the estimators of interest. For example Dalalyan, Golubev, and Tsybakov (2006) noted that in the context of penalised spline estimation of a (correctly specified) semiparametric Gaussian shift model, the magnitude of the second order terms is almost as big as that of first order terms[1]. Linton (2002) showed that for certain adaptive semiparametric models (including partial linear) kernel (or local polynomial) estimators affect the second order variance term but not the bias, whereas Ichimura and Linton (2006) noted that in the context of semiparametric average treatment effect models, the second order terms are mostly bias related and very large for the optimal choice of bandwidth. Secondly, because of the dependence on different $\theta_{\gamma*}$'s and on the nonzero probability limit $\lambda_*\left(\theta_{\gamma*}\right)$, as well as on the Jacobian $\partial\lambda_*\left(\theta\right)/\partial\theta'$ and its derivatives $\partial^2\lambda_*\left(\theta\right)/\partial\theta'\partial\theta_j$ $(j=1,...,k)$ evaluated at $\theta=\theta_{\gamma*}$, the expressions of the second order terms become very cumbersome and difficult to compare, if a comparison is at all possible. For these reasons we use simulation evidence, reported in Section 5, which suggests that actually no particular value of γ yields an estimator that clearly dominates in terms of its finite sample properties.

To consider the asymptotic distribution of the two-step semiparametric GMM estimator defined in (2), let

$$H\left(\theta,h,W\right)=G\left(\theta,h\right)'WG\left(\theta,h\right)+\mu_*'W\otimes I_k\left\{E\left[\frac{\partial}{\partial\theta'}vec\left(\frac{\partial g\left(\theta,h\right)}{\partial\theta'}\right)\right]\right\},$$

[1]In the typical case of twice differentiable unknown functions, the second order terms are of order $O_p\left(n^{-7/10}\right)$.

$$
\begin{aligned}
\Omega(\theta,h,W) &= E\left\{\left[\begin{array}{c}(g(\theta,h)-\mu_*)\\ \left(\frac{\partial g(\theta,h)}{\partial\theta'}-G(\theta,h)\right)'W\mu_*\\ \left(\widehat{W}-W\right)\mu_*\end{array}\right]^{\otimes 2}\right\}\\
&= \left[\begin{array}{ccc}\Omega_g(\theta_*,h_\bullet) & \Omega_{gG}(\theta_*,h_\bullet,W) & \Omega_{gW}(\theta_*,h_\bullet,W)\\ \Omega_{gG}(\theta_*,h_\bullet,W)' & \Omega_G(\theta_*,h_\bullet,W) & \Omega_{GW}(\theta_*,h_\bullet,W)\\ \Omega_{gW}(\theta_*,h_\bullet,W)' & \Omega_{GW}(\theta_*,h_\bullet,W)' & \Omega_W(\theta_*,h_\bullet,W)\end{array}\right],
\end{aligned}
$$

$$
\begin{aligned}
\Omega(\theta,h,W,\delta) &= E\left\{(g(\theta,h)-\mu_*+\delta_g)',\left[\left(\frac{\partial g(\theta,h)}{\partial\theta'}-G(\theta,h)+vec^{-1}(\delta_{\partial g})\right)'W\mu_*\right]',\right.\\
&\qquad\left.\left[\left(\widehat{W}-W\right)\mu_*\right]'\right\}^{\prime\otimes 2},
\end{aligned}
$$

where δ_g and $\delta_{\partial g}$ denote the asymptotic representations of the pathwise derivatives $g_h(\theta_*,h)[h-h_\bullet]$ of $g(\theta_*,h)$ and $\partial g_h(\theta_*,h)/\partial\theta'[h-h_\bullet]$ of $\partial g(\theta_*,h)/\partial\theta'$ satisfying the regularity condition A7 given below and $vec^{-1}(\delta_{\partial g})$ is the inverse vec operator that arranges (column-wise) the components of the $lk\times 1$ vector $\delta_{\partial g}$ into an $l\times k$ matrix. Assume that:

either

A6 (ii) $E[g(\theta_*,h)-\mu(\theta_*,h)]|_{h=\widehat{h}}=o_p\left(n^{-1/2}\right)$, $E[vec(\partial g(\theta_*,h)/\partial\theta')-vec(G(\theta_*,h))]|_{h=\widehat{h}}=o_p\left(n^{-1/2}\right)$,

or

A7 (i) the pathwise derivatives $g_h(\theta_*,h)[h-h_\bullet]$ of $g(\theta_*,h)$ and $\partial g_h(\theta_*,h)/\partial\theta'[h-h_\bullet]$ of $\partial g(\theta_*,h)/\partial\theta'$ exist $a.s.$ in all directions $[h-h_\bullet]$ and for small enough $\|h-h_\bullet\|_{\mathcal{H}}$

$$
\begin{aligned}
\|g(\theta_*,h)-g(\theta_*,h_\bullet)-g_h(\theta_*,h)[h-h_\bullet]\| &\le b_g(Z)\|h-h_\bullet\|^2_{\mathcal{H}},\\
\left\|vec\frac{\partial g(\theta_*,h)}{\partial\theta'}-vec\frac{\partial g(\theta_*,h_\bullet)}{\partial\theta'}-vec\frac{\partial g_h(\theta_*,h_\bullet)}{\partial\theta'}[h-h_\bullet]\right\| &\le b_{\partial g}(Z)\|h-h_\bullet\|^2_{\mathcal{H}}
\end{aligned}
$$

for small enough $\|h-h_\bullet\|_{\mathcal{H}}$, with $E|b_g(Z)|<\infty$ and $E|b_{\partial g}(Z)|<\infty$, (ii) there exist functions $\delta_g(\cdot):\mathcal{Z}\to\mathbb{R}^l$ and $\delta_{\partial g}(\cdot):\mathcal{Z}\to\mathbb{R}^{lk}$ with $E\left[\delta_g(Z)',\delta_{\partial g}(Z)'\right]'=0$ and $E\left[\left\|\delta_g(Z)',\delta_{\partial g}(Z)'\right\|^2\right]<\infty$ such that

$$
\begin{aligned}
\left\|\widehat{g}_h(\theta_*,h_\bullet)\left[\widehat{h}-h_\bullet\right]-\widehat{\delta}_g(Z)\right\| &= o_p\left(n^{-1/2}\right),\\
\left\|vec\frac{\partial\widehat{g}_h(\theta_*,h_\bullet)}{\partial\theta'}\left[\widehat{h}-h_\bullet\right]-\widehat{\delta}_{\partial g}(Z)\right\| &= o_p\left(n^{-1/2}\right),
\end{aligned}
$$

where $\widehat{\delta}_\bullet(Z)\xrightarrow{p}\delta_\bullet(Z)$ and "$\bullet$" is either g or ∂g. Assumptions A6(ii)-A7 are similar to A3(ii) and A4 and are important because as with the CECR extimator they explain the difference between the asymptotic variance matrices $\Phi_{GMM}(\theta_*,h_\bullet,W)$ and $\Phi_{GMM}(\theta_*,\lambda_*,h_\bullet,\delta)$ in Theorem 3.3.

Theorem 3.3 *Assume that (i) $Q_W(\theta,h_\bullet)$ is uniquely minimized at $\theta_*\in\Theta$ (ii) $\theta_*\in int(\Theta)$ (iii) $H(\theta_*,h_\bullet,W)$ is nonsingular (iv) $\Omega(\theta_*,h_\bullet,W)$ and $\Omega(\theta_*,h_\bullet,W,\delta)$ are* *positive definite. Then under either (v) A0 (with* $\vartheta = 0$*) A1(i), A2(i), A3', A5, A6(i) (in the supplemental appendix) and A6(ii)*

$$n^{1/2}\left(\widehat{\theta}_{GMM} - \theta_*\right) \xrightarrow{d} N\left(0, \Phi_{GMM}\left(\theta_*, h_\bullet, W\right)\right),$$

where

$$\Phi_{GMM}\left(\theta_*, h_\bullet, W\right) = H\left(\theta_*, W, h_\bullet\right)^{-1}\left[G\left(\theta_*, h_\bullet\right)' W\left(\Omega_g\left(\theta_*, h_\bullet\right) + \Omega_G\left(\theta_*, h_\bullet, W\right) + \Omega_W\left(\theta_*, h_\bullet, W\right)\right) G\left(\theta_*, h_\bullet\right) + G\left(\theta_*, h_\bullet\right)' W\Omega_{gG}\left(\theta_*, h_\bullet, W\right) + G\left(\theta_*, h_\bullet\right)' W\Omega_{gW}\left(\theta_*, h_\bullet, W\right) G\left(\theta_*, h_\bullet\right) + \Omega_{gG}\left(\theta_*, h_\bullet, W\right)' WG\left(\theta_*, h_\bullet\right) + G\left(\theta_*, h_\bullet\right)' \Omega_{gW}\left(\theta_*, h_\bullet, W\right)' WG\left(\theta_*, h_\bullet\right) + \Omega_{GW}\left(\theta_*, h_\bullet, W\right) G\left(\theta_*, h_\bullet\right) + G\left(\theta_*, h_\bullet\right)' \Omega_{GW}\left(\theta_*, h_\bullet, W\right)'\right]\left(H\left(\theta_*, W, h_\bullet\right)^{-1}\right)', \tag{14}$$

or (vi) A0 (with $\vartheta = 1/4$*), A1(i), A2(i), A3', A5 (in the supplemental appendix) and A7*

$$n^{1/2}\left(\widehat{\theta}_{GMM} - \theta_*\right) \xrightarrow{d} N\left(0, \Phi_{GMM}\left(\theta_*, h_\bullet, W, \delta\right)\right), \tag{15}$$

where $\Phi_{GMM}\left(\theta_*, \lambda_*, h_\bullet, \delta\right)$ *is as that defined in* (14) *with* $\Omega\left(\theta_*, \lambda_*, h_\bullet, \delta\right)$ *replacing* $\Omega\left(\theta_*, \lambda_*, h_\bullet\right)$.

Theorem 3.3 shows that the asymptotic variance of the two-step semiparametric GMM estimator is fairly complicated but its structure is easy to interpret since it reflects the fact that the centred derivatives $vec\left(\partial g\left(\theta_*, h_*\right)/\partial\theta'\right) - vec\left(G\left(\theta_*, h_*\right)\right)$ and weight matrices $\widehat{W} - W$ contribute to the asymptotic normality of $n^{1/2}\left(\widehat{\theta}_{GMM} - \theta_*\right)$.

The same weighted bootstrap approach used to approximate the asymptotic distribution of $\widehat{\theta}_{CECR}$ can be used to approximate that of $\widehat{\theta}_{GMM}$. Let $\{\omega_i\}_{i=1}^n$ denote an i.i.d. sample of positive random weights independent of $\{Z_i\}_{i=1}^n$ with $E\left(\omega_i\right) = 1$ and $Var\left(\omega_i\right) = \upsilon_0 < \infty$ and let

$$\widehat{\theta}_{\omega,GMM} = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^n g_{\omega i}\left(\theta, \widehat{h}\right)' \widehat{W}_\omega \frac{1}{n}\sum_{i=1}^n g_{\omega i}\left(\theta, \widehat{h}\right)$$

denote the weighted bootstrap two-step semiparametric GMM estimator, where $g_{\omega i}\left(\theta, h\right) = \omega_i g_i\left(\theta, h\right)$ and $\widehat{W}_\omega$ is a possibly random matrix that might also depend on the ω_i's.

Theorem 3.4 *Under the same assumptions of Theorem* 3.3*, with* $\widehat{W}_\omega$ *replacing* $\widehat{W}$ *in assumption A5 (in the supplemental appendix), then conditionally on the sample* $\{Z_i\}_{i=1}^n$*,*

$$\sup_u \left| \Pr{}^\omega\left(\left(\frac{n}{\upsilon_0}\right)^{1/2}\left(\widehat{\theta}_{\omega,GMM} - \widehat{\theta}_{GMM}\right) \le u\right) - \Pr\left(n^{1/2}\left(\widehat{\theta}_{GMM} - \theta_*\right) \le u\right)\right| = o_p\left(1\right).$$

Remark 2 It is important to note that the proposed two-step semiparametric CECR estimator can also be used with correctly specified semiparametric models, that is for models where $\mu\left(\theta, h_0\right) = 0$ has a unique solution for $\theta = \theta_0$. See Section 1 in the supplemental appendix for further details and results.

4. Two examples

In this section we illustrate the results of the previous section by considering two semiparametric models with a separable additive nonparametric component. In the first case, the first step estimation has an effect on the asymptotic variance of the finite dimensional parameter estimator, whereas in the second case it has not. In both cases the nonparametric parameter is assumed to be correctly specified and is estimated with (nonparametric) series (see for example Stone (1985) and Newey (1997)). Suppose that $Z = [Z_1', Z_2']'$ and $h_0 = h_0(Z_2)$; for $m = 1$ the series estimator of h_0 is $\widehat{h} = p^K(Z_2)'\beta$, where

$$p^K(Z_2) = [p_1(Z_2), ..., p_K(Z_2)]'$$

is a K dimensional vector of approximating functions such as power series or splines and β is a vector of unknown parameters that can be estimated from the data. Let $|h(Z_2)|_d = \max_{|\lambda|\leq d} \sup_{Z_2 \in \mathcal{Z}_2} \left|\partial^{|\lambda|} h(Z_2)/\partial Z_{21}^{\lambda_1} ... \partial Z_{2k_2}^{\lambda_{k_2}}\right|$ for $|\lambda| := \sum_{j=1}^{k_2} \lambda_j$ and $k_2 = \dim(Z_2)$. Assume that:

- S1 (i) the support $\mathcal{Z}_2$ of Z_2 is the Cartesian product of compact subsets of $\mathbb{R}$ (ii) $h_0(Z_2)$ is d-times continuously differentiable on $\mathcal{Z}_2$ with uniformly bounded derivatives,
- S2 (i) for every K, there is a nonsingular matrix J such that for $P^K(Z_2) := Jp^K(Z_2)$ the smallest eigenvalue of $E\left[P^K(Z_2)^{\otimes 2}\right]$ is bounded away from zero uniformly in K (ii) there exists a sequence of constants $\xi(K)$ satisfying $\sup_{Z_2 \in \mathcal{Z}_2} \left\|P^K(Z_2)\right\| \leq \xi(K)$ and $K = K(n)$ such that $\xi(K)^2 K/n \to 0$ as $n \to \infty$,
- S3 (i) for $d \geq 0$ $\left|h_0(Z_2) - p^K(Z_2)'\beta\right|_d = O\left(K^{-\delta}\right)$ as $K \to \infty$ (ii) $n^{1/2}K^{-2\delta} \to 0$ as $n \to \infty$.

Assumptions S1 and S3 are standard in the literature of nonparametric series estimation, see for example Newey (1997); S2 usually implies that the density function of Z_2 is bounded below by a positive constant. S1 implies S3(i) with the rate $\delta := \delta(d, k_2)$; S3(ii) implies that the estimation bias from the nonparametric component is smaller than the parametric convergence rate. It is equivalent to the standard undersmoothing assumption often assumed in kernel estimation of semiparametric models. Examples of nonparametric series that satisfy S2 and S3 include power series and splines.

4.1. *Quadratic inference functions (QIF) models with nonparametric generated regressors*

QIF models were introduced by Qu et al. (2000) as an alternative to the generalised estimating equations (GEE) approach of Liang and Zeger (1986), which avoids estimating the nuisance correlation structure parameters by assuming that the inverse of the working correlation matrix can be approximated by a linear combination of several known basis matrices. The QIF model considered here allows for general misspecification of the estimating equations and is defined as

$$E(Y|X) = m\left(X_1'\theta_{10} + X_2^{*\prime}\theta_{20}\right), \tag{16}$$

where $m(\cdot) : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}^{d_y}$ is a vector of known functions with the same dimension as that of the vector valued response Y and $X_2^* \in \mathbb{R}^{k_2}$ are latent covariates that can be expressed as conditional expectations $E(X_2|X_1)$ of the observable covariates X_2. In this case $h_0(X_1) = E(X_2|X_1) := h_0$ and the series estimator is $\widehat{h}(X_{1i}) = p^K(X_{1i})'(PP')^{-}PX_2$, where $P = \left[p^K(X_{11}), ..., p^K(X_{1n})\right]$. Under correct specification, we assume that there exist unique θ_{10} and θ_{20} such that (16) holds *a.s.*, whereas under misspecification we have

$$\Pr\left(E(Y|X) \neq m\left(X_1'\theta_1 + X_2^{*\prime}\theta_2\right)\right) > 0,$$

$\forall\ \theta_1,\ \theta_2 \in \Theta_1 \times \Theta_2$. In this case estimation of the pseudo-true parameter of interest $\theta_* = [\theta_{1*}', \theta_{2*}']'$ or $\theta_{\gamma*} = \left[\theta_{\gamma1*}', \theta_{\gamma2*}'\right]'$ can be based on the profile quadratic inference function

$$g_i\left(\theta, \widehat{h}\right) = I_s \otimes \left(\frac{\partial m\left(v_i\left(\theta, \widehat{h}(X_{1i})\right)\right)}{\partial\theta'}\right)' [B_{1i}, ..., B_{si}]'\left(Y_i - m\left(v_i\left(\theta, \widehat{h}(X_{1i})\right)\right)\right) \quad (17)$$

$$B_{ji} = A_i^{-1/2} M_j A_i^{-1/2}\ j = 1, .., s,$$

where $v_i\left(\theta, \widehat{h}(X_{1i})\right) = X_{\widehat{h}i}'\theta$, $X_{\widehat{h}i} = \left[X_{1i}', \widehat{h}(X_{1i})'\right]'$, A is an $\mathbb{R}^{d_y} \times \mathbb{R}^{d_y}$ diagonal matrix containing the marginal variances of Y and M_j $(j = 1, ...s)$ are known basis matrices used to approximate the inverse of the working correlation matrix R, the number of which s depends on the assumed structure of R. For example if we assume an exchangeable working correlation matrix where all pairs of observations share the same correlation coefficient, then $s = 2$ and M_1 is the identity matrix while M_2 has 0's on the diagonal and 1's elsewhere - see Qu et al. (2000) for more examples.

Proposition 4.1 *Under QIF1-QIF4 (in the supplemental appendix) and if, for $\lambda_*(\theta)$ defined in* (8) *with $g_i\left(\theta, \widehat{h}\right)$ given in* (17), *(i) $E[\rho_\gamma(\theta, \lambda_*, h_0)]$ has a unique minimum at $\theta_{\gamma*} \in \Theta$ and (ii) $\theta_{\gamma*} \in int(\Theta)$ and $\lambda_*(\theta) \in int(\Lambda(\Theta))$, then the asymptotic distribution of $n^{1/2}\left(\widehat{\theta}_{CECR} - \theta_{\gamma*}\right)$ is as that given in* (12) *with the pathwise derivatives $\delta_\gamma(X_1)$ and $\delta_s(X_1)$ given in* (A.4) *in the supplemental appendix. In addition under QIF5 (in the supplemental appendix) a consistent estimator for $\Phi_{CECR}(\theta_{\gamma*}, \lambda_*, h_0, \delta)$ is given by its sample analogue $\widehat{\Phi}_{CECR}\left(\widehat{\theta}_{CECR}, \widehat{\lambda}, \widehat{h}, \widehat{\delta}\right)$.*
Under QIF1, QIF2, QIF6–QIF9 (in the supplemental appendix) and if (i) $Q_W(\theta, h_0)$ is uniquely minimised at $\theta_* \in \Theta$ and (ii) $\theta_* \in int(\Theta)$, then the asymptotic distribution of $n^{1/2}\left(\widehat{\theta}_{GMM} - \theta_*\right)$ based on (17) *is as that given in* (15) *with $\delta_g(X_1)$ and $\delta_{\partial g}(X_1)$ given in* (A.5) *in the supplemental appendix. In addition under QIF10 (in the supplemental appendix) a consistent estimator for $\Phi_{GMM}(\theta_*, h_0, W, \delta)$ is given by its sample analogue $\widehat{\Phi}_{GMM}\left(\widehat{\theta}_{GMM}, \widehat{h}, \widehat{W}, \widehat{\delta}\right)$.*

4.2. *Instrumental variables partially linear additive models*

The statistical model is

$$Y = X_1'\theta_0 + \sum_{j=1}^{m} h_{0j}(X_{2j}) + \varepsilon, \quad (18)$$

where $\theta_0 \in \Theta$ is a k-dimensional vector of unknown parameters, $h_0(X_2) = [h_{01}(X_{21}), ..., h_{0m}(X_{2m})]'$ is a vector of unknown real valued functions and ε is an unobservable error. Li (2000) considered the case of a correctly specified (18) in the sense that $E(\varepsilon|X) = 0$ $a.s.$, where $X = [X_1', X_2']'$. We assume instead that the X_1 vector of covariates is endogenous (that is it is correlated with the errors) and that there exists an $\mathbb{R}^l$-valued $(l \geq k)$ vector V of instruments, which is however misspecified in the sense that $\Pr(E(\varepsilon|V) \neq 0) > 0$ for all $\theta \in \Theta$. Let $\mathcal{A}$ denote the space of additive continuous functions h satisfying $h_j(0) = 0$ and $\sum_{j=1}^{m} E[h_j(X_{2j})]^2 < \infty$ for $j = 1, ..., m$ and let $\Pi(\cdot|\mathcal{A})$ denote the mean square projection of $\cdot$ on $\mathcal{A}$. In this example $h_0(X_2) = [\Pi(Y|\mathcal{A}), \Pi(X_1|\mathcal{A})']' := h_0$ and the series estimator of the projections is $\widehat{h}(X_{2i}) = \left[p^K(X_{2i})'\widehat{\beta}_Y, p^K(X_{2i})'\widehat{\beta}_{X_1}'\right]'$ where $\widehat{\beta}_Y = (P'P)^- P'Y$, $\widehat{\beta}_{X_1} = (P'P)^- P'[X_{11}, X_{12}, ..., X_{1k}]'$ with $P = \left[p^K(X_{21}), ..., p^K(X_{2n})\right]$ and

$$\begin{aligned} p^K(X_{2i}) &= \left[p_1^K(X_{21i})', ..., p_m^K(X_{2mi})'\right]', \\ p_j^K(X_{2j}) &= \left[p_{j1}(X_{2j}), ..., p_{jK_j}(X_{2j})'\right], \quad j = 1, ..., m. \end{aligned}$$

In this case estimation of the pseudo true parameter of interest θ_* or $\theta_{\gamma*}$ can be based on the profile moment indicator

$$g_i\left(\theta, \widehat{h}(X_{2i})\right) = V_i\left(\widehat{Y}_i - \widehat{X}_{1i}'\theta\right), \tag{19}$$

where $\widehat{Y}_i = Y_i - p^K(X_{2i})'\widehat{\beta}_Y$ and $\widehat{X}_{1i} = X_{1i} - p^K(X_{2i})'\widehat{\beta}_{X_1}$.

Proposition 4.2 *Under PLA1-PLA4 (in the supplemental appendix) and if, for* $\lambda_*(\theta)$ *defined* (8) *with* $g_i\left(\theta, \widehat{h}\right)$ *given in* (19)*, (i)* $E[\rho_\gamma(\theta, \lambda_*, h_0)]$ *has a unique minimum at* $\theta_{\gamma*} \in \Theta$ *and (ii)* $\theta_{\gamma*} \in int(\Theta)$ *and* $\lambda_*(\theta) \in int(\Lambda(\Theta))$*, then the asymptotic distribution of* $n^{1/2}\left(\widehat{\theta}_{CECR} - \theta_{\gamma*}\right)$ *is as that given in* (11)*. In addition, a consistent estimator for* $\Phi_{CECR}(\theta_{\gamma*}, \lambda_*, h_0)$ *is given by its sample analogue* $\widehat{\Phi}_{CECR}\left(\widehat{\theta}_{CECR}, \widehat{\lambda}, \widehat{h}\right)$*. Under PLA1-PLA2 and PLA5-PLA7 (in the supplemental appendix) and if (i)* $Q_W(\theta, h_0)$ *is uniquely minimized at* $\theta_* \in \Theta$ *and (ii)* $\theta_* \in int(\Theta)$*, then the asymptotic distribution of* $n^{1/2}\left(\widehat{\theta}_{GMM} - \theta_*\right)$ *based on* (19) *is as that given in* (14) *with* $H(\theta_*, W, h_0) = E\left(\overline{X}_1 V'WV\overline{X}_1'\right)$*. In addition, a consistent estimator for* $\Phi_{GMM}(\theta_*, h_0, W)$ *is given by its sample analogue* $\widehat{\Phi}_{GMM}\left(\widehat{\theta}_{GMM}, \widehat{h}, \widehat{W}\right)$*.*

5. Monte Carlo evidence

In this section we use the two examples of the previous section to illustrate the finite sample properties of the estimators considered in Section 3. We consider six estimators for the case of correctly specified estimating equations and seven estimators for the case of misspecified estimating equations. The estimators are $\widehat{\theta}_{EL}$ (two-step semiparametric empirical likelihood), $\widehat{\theta}_{ET}$ (two-step semiparametric exponential tilting, that is CECR with $\gamma = 0$), $\widehat{\theta}_{ETEL}$ (two-step semiparametric combined exponential tilting empirical likelihood, that is CECR with $\gamma = -1$), $\widehat{\theta}_{ELEU}$ (two-step semiparametric combined empirical

likelihood Euclidean likelihood), $\widehat{\theta}_{ETEU}$ (two-step semiparametric combined exponential tilting Euclidean likelihood, that is CECR with $\gamma = 2$) and two $\widehat{\theta}_{GMM}$s (two-step semiparametric GMM) computed using two different weighting matrices $\widehat{W}$.
Under correct specification and the asymptotic orthogonality condition A2'(ii) (in the supplemental appendix) all of the above estimators (including $\widehat{\theta}_{GMM}$ with the optimal weighting matrix $\widehat{W} = \left(\sum_{i=1}^{n} g_i\left(\widetilde{\theta}, \widehat{h}\right)^{\otimes 2}/n\right)^{-1}$ (Ackerberg et al. 2014), where $\widetilde{\theta}$ is a preliminary $n^{1/2}$ consistent estimator) are asymptotically equivalent (see for example Theorem 5 (a)-(b) in the supplemental appendix), but only $\widehat{\theta}_{ET}$, $\widehat{\theta}_{ETEL}$, $\widehat{\theta}_{ETEU}$ and $\widehat{\theta}_{GMM}$ are robust to misspecification. Note that while robust to misspecification, $\widehat{\theta}_{ETEU}$ is defined for $\gamma = 1$, which, as mentioned in Section 2, can yield negative implied probabilities.
Under misspecification, we consider two specifications of $\widehat{W}$, namely the same optimal weighting matrix used in the case of correct specification and the identity matrix I ($\widehat{\theta}_{GMM_I}$). The latter choice is motivated by the fact that under misspecification there is no optimal choice of $\widehat{W}$, hence any weighting matrix satisfying assumption A5 (in the supplemetal appendix) can be used in the estimation process. Indeed the choice of $\widehat{W} = I$ is often used with misspecified models, see for example Cochrane (2001) and Ai and Chen (2007).
For the weighted bootstrap we use the same two points distribution as that used by Lavergne and Patilea (2013) defined by $\Pr\left(\omega = \left(3-\sqrt{5}\right)/2\right) = \left(5+\sqrt{5}\right)/10$ and $\Pr\left(\omega = \left(3+\sqrt{5}\right)/2\right) = \left(5-\sqrt{5}\right)/10$ for which $E(\omega) = Var(\omega) = 1$.

We first consider a semiparametric QIF model with identity link

$$Y = X_{11}\theta_{10} + \kappa X_{12} + E\left(X_2|X_{11}\right) + \varepsilon, \tag{20}$$

where X_{11} and X_{12} are two correlated $U(0,1)$ random variables, $E\left(X_2|X_{11}\right) = \sin\left(X_{11}\right)$ and $\varepsilon \sim N_c\left(0, R(\alpha)\right)$, where $N_c(\cdot)$ denotes a c-dimensional multivariate normal, $R(\alpha)$ stands for the covariance matrix and c is the dimension of the cluster (which implies that the dimension of Y is also c). Note that by Proposition 2 of Newey (1994) the pathwise derivatives appearing in the asymptotic variances $\Phi_{CECR}\left(\theta_*, \lambda_*, h_0, \delta\right)$ and $\Phi_{GMM}\left(\theta_*, h_0, W, \delta\right)$ are all identically equal to 0. In the simulations we set $\theta_{10} = 2$ with the dimension of the cluster $c = 10$; the working correlation matrix $R(\alpha)$ has an AR(1) structure with parameter $\alpha = 0.5$, so that M_1 is the identity matrix and for $j = 2, 3$ the structure of M_j is known (Qu et al. 2000). The parameter κ takes two values, 0 or $\kappa^* \neq 0$, which correspond, respectively, to the correctly specified case ($\kappa = 0$) and to the misspecified one (κ^*), since we specify the estimating equations (17) as

$$g_i\left(\theta_1, \widehat{h}\right) = I_3 \otimes X_{11i}\left[B_1, B_2, B_3\right]'\left(Y_i - 1_c X_{11i}\theta_1 - 1_c \widehat{E}\left(X_{2i}|X_{11i}\right)\right), \tag{21}$$

where $\widehat{E}\left(X_{2i}|X_{11i}\right)$ is a nonparametric series estimator of $E\left(X_2|X_{11}\right)$. For $\kappa = 0$ (21) is correctly specified since $E\left(g\left(\theta_{10}, h_0\right)\right) = 0$ in view of (20), whereas for $\kappa = \kappa^*$, (21) is misspecified since there is no θ_1 that satisfies $E\left(g\left(\theta_1, h_0\right)\right) = 0$ simultaneously for B_1, B_2 and B_3, given the correlation $\rho_{X_{11}X_{12}}$ between X_{11} and X_{12} and $\kappa^* \neq 0$. To construct the estimator $\widehat{E}\left(X_{2i}|X_{1i}\right)$ we use cubic splines to generate the approximating functions

$p^K(\cdot)$ with the term K determined by the generalised cross-validation criterion[2]

$$\widehat{K}_{GCV}(K) = \arg\min_K \frac{1}{n}\sum_{i=1}^{n} \frac{\left\| I_3 \otimes X_{11i}\left[B_{1i}, B_{2i}, B_{3i}\right]'\left(Y_i - 1_c X_{11i}\widehat{\theta}_1 - 1_c \widehat{E}\left(X_{2i}|X_{11i}\right)\right)\right\|^2}{\left(1-(K/n)\right)^2}.$$

To compute the pseudo-true values θ_{1*} of the two-step semiparametric empirical likelihood and four two-step semiparametric CECR estimators using numerical integration and a two-step procedure. In the first step, for a fixed θ_1, the maximiser $\lambda_*(\theta_1)$ of (8) with $h_\bullet = h_0$ is computed. In the second step, given $\lambda_*(\theta_1)$ obtained in the first step, the minimiser $\theta_{\gamma 1*}$ of $E\left[\rho_\gamma\left(\theta_1, \lambda_*(\theta_1), h_0\right)\right]$ is found over the interval $\theta_1 \in [-5, 5]$. This procedure yields, for a correlation coefficient $\rho_{X_{11}X_{12}}$ of 0.5 and $\kappa^* = 2$, the pseudo-true value θ_{1*} of -2.06 for the empirical likelihood, of -1.03 for the exponential tilting, of -0.918 for the combined exponential tilting empirical likelihood, of -2.77 for the combined empirical likelihood Euclidean likelihood and of -1.98 for the combined exponential tilting Euclidean likelihood specifications. For the two-step semiparametric GMM (QIF) objective function, the pseudo-true value θ_{1*} is computed using only the second step of the procedure, which gives a value of -2.35.

Tables 1 and 2 report the finite sample bias, standard error, coverage of a nominal 95% confidence interval and its average length calculated using the normal approximation and the weighted bootstrap approximation for the six estimators under correct specification and seven estimators under misspecification. The results are for two sample sizes $n = 100$ and $n = 400$ and are based on 1000 replications with the number B of bootstrap replications set to $B = 300$.

Tables 1 and 2 approx. here

Figure 1 shows the cumulative distributions of $\widehat{\theta}_{EL}$, $\widehat{\theta}_{ET}$, $\widehat{\theta}_{ELEU}$ and $\widehat{\theta}_{ETEU}$ (centred at θ_0 under correct specification and at the corresponding θ_*'s under misspecification) for $n = 100$. Figure 2 shows the mean squared error (MSE), coverage and average length of the nominal 95% level confidence intervals for the six estimators considered in Tables 1 and 2 and seventeen additional two-step semiparametric CECR estimators corresponding to γ =[-16,-15,-12,-8,-7,-5,-4,-3,-2,-3/2,-3/4,-1/2,-1/4,1/2,1,2,3,5] under misspecification (note that the estimators corresponding to the positive values of γ can yield implied probabilities that are not positive by construction).

Figures 1 and 2 approx. here

Next we consider a partially linear additive (PLA) model which extends the one considered by Li (2000) by allowing endogeneity of the regressors and misspecification of the instruments. The model is

$$\begin{aligned} Y &= \theta_{10} + \theta_{20}X_1 + h_{10}(X_{21}) + h_{20}(X_{22}) + \varepsilon_1, \\ X_1 &= X_3 + 3X_4 + \varepsilon_2, \end{aligned} \tag{22}$$

where $[\theta_{10}, \theta_{20}]' = [1, 0.5]'$, $X_j \sim U(0,2)$ $(j = 1,3,4)$, $X_{2k} \sim U(0,2)$ $(k = 1,2)$,

[2]As discussed in Bickel and Kwon (2002) estimation of the finite dimensional parameter is not very sensitive in general to the choice of K (which can be interpreted as a smoothing parameter) as long as the selected K does not result in a large bias for the infinite dimensional parameter estimate. In this respect cross-validation typically performs well, regardless of possible misspecification, and this is why we use it in the simulations.

$h_{10}(X_{21}) = \exp(-X_{21}) - 1$, $h_{20}(X_{22}) = \sin(\pi X_{22})$,

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}\right),$$

$\rho_{12} = corr(\varepsilon_1, \varepsilon_2) = 0.4$. In the case of correct specification, we use as instruments $V = [1, X_{21}, X_{22}, X_3, X_4]'$, while in the misspecified case we use as instruments $V = [1, X_{21}, X_{22}, X_3^*, X_4]'$ with $X_3^* = 10\varepsilon_2$. As with the previous example, the various pseudo-true values $\theta_{\gamma *} = [\theta_{\gamma 1*}, \theta_{\gamma 2*}]'$ are calculated using numerical integration and the same two-step procedure described above, with the only difference that the minimisation of $E[\rho_\gamma(\theta, \lambda_*, h_0)]$ is carried over the rectangle $[-2, 2] \times [-1, 1]$. The resulting pseudo-true values are $[1.71, 0.31]'$ for the empirical likelihood, $[1.54, 0.22]'$ for the exponential tilting, $[1.64, 0.32]'$ for the combined exponential tilting empirical likelihood, $[1.63, 0.42]'$ for the combined empirical likelihood Euclidean likelihood, $[1.56, 0.21]'$ for the combined exponential tilting Euclidean likelihood and finally $[1.34, 0.81]$ for the GMM specification. As with the previous example, we use cubic splines to estimate the nonparametric components with the term K determined by the generalised cross-validation criterion

$$\widehat{K}_{GCV}(K) = \arg\min_K \frac{1}{n} \sum_{i=1}^{n} \frac{\left(Y_i - [1, X_{1i}]\,\widehat{\theta} - \widehat{h}\right)^2}{(1 - (K/n))^2}.$$

Tables 3 and 4 report the finite sample bias, standard errors, coverage of a nominal 95% confidence interval and average length for the slope parameter θ_2 calculated using the normal approximation and the weighted bootstrap approximation for the same six estimators, sample sizes and bootstrap replications as those considered in Tables 1 and 2.

Tables 3 and 4 approx. here

Figure 3 shows the cumulative distribution of $\widehat{\theta}_{2EL}$, $\widehat{\theta}_{2ET}$, $\widehat{\theta}_{2ELEU}$ and $\widehat{\theta}_{2ETEU}$ (centred at θ_{20} under correct specification and at the corresponding θ_{2*} under misspecification) for $n = 100$. Figure 4 shows the MSE, coverage and average length of the nominal 95% level confidence intervals for the six estimators considered in Tables 3 and 4 and the same seventeen additional two-step semiparametric CECR estimators under misspecification considered in Figure 3.

Figures 3 and 4

The results of Tables 1-4 can be summarised as follows: under correct specification (Tables 1 and 3) all estimators are characterised by good finite sample properties. The finite sample biases are statistically insignificant for both sample sizes, with the biases of the two-step semiparametric empirical likelihood estimator $\widehat{\theta}_{EL}$ and of the two-step semiparametric combined empirical likelihood euclidean likelihood estimator $\widehat{\theta}_{ELEU}$ being the smallest and those of the GMM estimators being the largest. The magnitude of the standard errors indicates that the estimators are fairly precise, with the magnitude shrinking, as expected, by a factor of 2 as the sample size increases from 100 to 400. The confidence intervals have good coverage accuracy that gets closer to the nominal 95% level as the sample size increases. Similarly the average lengths of the confidence intervals decrease as the sample size increases. Finally, the proposed weighted bootstrap procedure yields confidence intervals that have better coverage accuracy and are slightly shorter than those based on the normal approximation. Under misspecification (Tables 2 and 4) the biases of $\widehat{\theta}_{EL}$ and $\widehat{\theta}_{ELEU}$ become statistically significant and can be more than four times as large as that of $\widehat{\theta}_{ET}$ or of $\widehat{\theta}_{ETEL}$ (see for example Table 4). Also $\widehat{\theta}_{ETEU}$ has a larger bias than that of either $\widehat{\theta}_{ET}$ or $\widehat{\theta}_{ETEL}$. The standard errors of $\widehat{\theta}_{ET}$, $\widehat{\theta}_{ETEL}$ and $\widehat{\theta}_{ETEU}$ are considerably smaller than those of $\widehat{\theta}_{EL}$ and $\widehat{\theta}_{ELEU}$, which do not decrease as the sample size increases. Note, however, that the standard error of $\widehat{\theta}_{ETEU}$ is larger than that of $\widehat{\theta}_{ET}$ and $\widehat{\theta}_{ETEL}$. As expected, misspecification negatively affects the coverage and especially the length of the confidence intervals based on $\widehat{\theta}_{EL}$ and $\widehat{\theta}_{ELEU}$ and their standard errors, which can be up to more than twice wider than those based on the robust estimators and standard errors. Tables 2 and 4 also show that $\widehat{\theta}_{ET}$, $\widehat{\theta}_{ETEL}$ and $\widehat{\theta}_{ETEU}$ compare favourably with respect to the GMM estimators based on both $\widehat{W} = \left(\sum_{i=1}^{n} g_i\left(\widetilde{\theta}, \widehat{h}\right)^{\otimes 2}/n\right)^{-1}$ and $\widehat{W} = I$. Interestingly, among the two GMM estimators, the one based on the identity matrix (GMM_I) performs better than the one based on the optimal weighting matrix. Finally, the weighted bootstrap seems to work well under misspecification, delivering confidence intervals closer to their nominal level and shorter than those based on the normal approximation.

Figures 1 and 3 confirm the findings of Tables 1-4 as they show that under correct specification the finite sample distribution of the centred four estimators $\widehat{\theta}_{EL}$, $\widehat{\theta}_{ELEU}$, $\widehat{\theta}_{ET}$ and $\widehat{\theta}_{ETEU}$ are very similar, whereas under misspecification there is a marked difference between the finite sample distributions of $\widehat{\theta}_{ET}$ and $\widehat{\theta}_{ETEL}$ with those of $\widehat{\theta}_{EL}$ and $\widehat{\theta}_{ELEU}$. In particular the much larger variability of $\widehat{\theta}_{EL}$ and $\widehat{\theta}_{ELEU}$, as reported by the standard errors in Tables 2 and 4, is clearly shown. Figures 2 and 4 show that all of the additional two-step semiparametric CECR estimators with $\gamma < 0$ have finite sample properties similar to those of $\widehat{\theta}_{ET}$, $\widehat{\theta}_{ETEL}$ with no one clearly dominating in terms of MSE, coverage accuracy and/or average length of the confidence intervals. Figures 2 and 4 also show that the additional two-step semiparametric CECR estimators with $\gamma > 0$, while robust to misspecification, are characterised by a larger MSE (about 32% for the QIF model and 35% for the PLA model), larger average length (about 12% for the QIF model and 8% for the PLA model) and a slightly lower coverage (about 1% for the QIF model and about 4% for the PLA model). The larger MSE is due mainly to larger standard errors than those obtained for $\gamma \leq 0$ (about 18% for the QIF model and 16% for the PLA model), which also explains the fact that the average length of the confidence intervals is longer. Thus Figures 2 and 4 seem to suggest that two-step semiparametric CECR estimators defined by nonpositive γ's are characterised by better sample properties than those defined by a positive γ.

6. Conclusion

In this paper we propose a general estimation method that can be used in the context of possibly misspecified semiparametric estimating equations models, where a preliminary estimator of the infinite dimensional parameter is available. The method is an alternative to generalised method of moments estimation and combines exponential tilting with another member of the empirical Cressie-Read discrepancy defined by a nonpositive value of the user specific parameter γ. The resulting two-step semiparametric CECR estimators are robust to misspecification and have implied probabilities that can are range preserving. Their asymptotic distribution is normal with a variance that depends on whether there is an estimation effect arising from the preliminary nonparametric estimator or not. To illustrate these results we consider two models that have not been previously investigated in the semiparametric literature: a quadratic inference functions model with nonparametric generated regressors and a partially linear additive model with endogenous regressors. We use a simulation study to investigate the finite sample properties of a number of estimators and test statistics. Taken together the results of the simulation study suggest that the proposed two-step combined nonparametric likelihood estimation can be a useful and valid alternative to generalised method of moment estimation in the context of possibly misspecified semiparametric estimating equations models.

Acknowledgement(s)

I would like to thank an Associate Editor and especially two referees for very useful comments and suggestions that improved considerably the paper.

Supplemental material

A supplemental appendix contains all the proofs and some additional material.

References

Ackerberg, D., Chen, X., Hahn, J., and Liao, Z. (2014), 'Asymptotic efficiency of semiparametric two-step GMM', *Review of Economic Studies*, 81, 919–943.

Ai, C., and Chen, X. (2007), 'Estimation of Possibly Misspecified Semiparametric Conditional Moment Restrictions Models with Different Conditioning Variables', *Journal of Econometrics*, 141, 5–43.

Andrews, D.W.K. (1994), 'Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity', *Econometrica*, 62, 43–72.

Baggerly, K.A. (1998), 'Empirical Likelihood as a Goodness of Fit Measure', *Biometrika*, 85, 535–547.

Bai, Y., Zhu, Z., and Fung, W. (2008), 'Partial linear models for longitudinal data based on quadratic inference functions', *Scandinavian Journal of Statistics*, 35, 104–118.

Bickel, P., and Kwon, J. (2002), 'Inference for Semiparametric Models: Some Current Frontiers and an Answer (with Discussion)', *Statistica Sinica*, 11, 863–960.

Bravo, F. (2009), 'Two-step generalised empirical likelihood inference for semiparametric models', *Journal of Multivariate Analysis*, 100, 1412–1431.

Bravo, F., Escanciano, J.C., and Van Keilegom, I. (2020), 'Two-step semiparametric empirical likelihood inference', *Annals of Statistics*, 48, 1–26.

Chamberlain, G. (1987), 'Asymptotic efficiency in estimation with conditional moment restrictions', *Journal of Econometrics*, 34, 305–334.

Chen, X., and Liao, Z. (2015), 'Sieve semiparametric two-step GMM under weak dependence', *Journal of Econometrics*, 189, 163–189.

Chen, X., Linton, O., and van Keilegom, I. (2003), 'Estimation of Semiparametric Models When the Criterion Function is Not Smooth', *Econometrica*, 71, 1591–1608.

Cochrane, J. (2001), *Asset Pricing*, Princeton University Press.

Dalalyan, A., Golubev, G., and Tsybakov, A. (2006), 'Penalized maximum likelihood and semiparametric second-order efficiency', *Annals of Statistics*, 33, 169–201.

Gospodinov, N., Kan, R., and Robotti, C. (2013), 'Chi-squared tests for evaluation and comparison of asset pricing models', *Journal of Econometrics*, 173, 108–125.

Gospodinov, N., Kan, R., and Robotti, C. (2014), 'Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors', *Review of Financial Studies*, 27, 2139–2170.

Grace, Y., and Reid, N. (2010), 'A note on mis-specified estimating equations', *Statistica Sinica*, 20, 1749–1769.

Hall, A., and Inoue, A. (2003), 'The Large Sample Behaviour of the generalized method of moments Estimator in Misspecified Models', *Journal of Econometrics*, 114, 361–394.

Hansen, L. (1982), 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica*, 50, 1029–1054.

Hansen, L., and Singleton, K. (1982), 'Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models', *Econometrica*, 50, 1269–1286.

Hjort, N., McKeague, I., and VanKeilegom, I. (2009), 'Extending the Scope of Empirical Likelihood', *Annals of Statistics*, 39, 1079–1111.

Ichimura, H., and Linton, O. (2006), 'Asymptotic expansions for some semiparametric program evaluation estimators', in *Identification and inference for econometric models. Essays in honor of Thomas Rothemberg.*

Jin, Z., Ying, Z., and Wei, L. (2001), 'A simple resampling method by perturbing the minimand', *Biometrika*, 88, 381–390.

Kitamura, Y., and Stutzer, M. (1997), 'An Information Theoretic Alternative to Generalized Method of Moments Estimation', *Econometrica*, 65, 861–874.

Kitamura, Y., Otsu, T., and Evdokimov, K. (2013), 'Robustness, infinitesimal neighbourhoods and moment restrictions', *Econometrica*, 81, 1185–1201.

Lai, P., Li, G., and Lian, H. (2013), 'Quadratic inference functions for partially linear single index models with longitudinal data', *Journal of Multivariate Aanalysis*, 118, 115–127.

Lavergne, P., and Patilea, V. (2013), 'Smooth minimum distance estimation and testing with conditional estimating equations: Uniform in bandwidth theory', *Journal of Econometrics*, 177, 47–59.

Li, Q. (2000), 'Efficient estimation of additive partially linear models', *International Economic Review*, 41, 1073–1092.

Liang, K., and Zeger, S. (1986), 'Longitudinal Data Analysis Using Generalised Linear Models', *Biometrika*, 73, 12–22.

Linton, O. (2002), 'Edgeworth approximations for semiparametric instrumental variables estimators and test statistics', *Journal of Econometrics*, 106, 325–368.

Ma, S., and Kosorok, M. (2005), 'Robust semiparametric M-estimation and the weighted bootstrap', *Journal of Multivariate Analysis*, 96, 190–217.

Newey, W. (1994), 'The Asymptotic Variance of Semiparametric Estimators', *Econometrica*, 62, 1349–1382.

Newey, W. (1997), 'Convergence Rates and Asymptotic Normality for Series Estimators', *Journal of Econometrics*, 79, 147–168.

Newey, W.K., and Smith, R.J. (2004), 'Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators', *Econometrica*, 72, 219–256.

Owen, A. (1988), 'Empirical Likelihood Ratio Confidence Intervals for a Single Functional', *Biometrika*, 36, 237–249.

Owen, A. (1991), 'Empirical Likelihood for Linear Models', *Annals of Statistics*, 19, 1725–1747.

Qu, A., Lindsay, B.G., and Li, B. (2000), 'Improving Generalised Estimating Equations Using Quadratic Inference Functions', *Biometrika*, 87, 823–836.

Schennach, S. (2007), 'Point Estimation with Exponentially Tilted Empirical Likelihood', *Annals of Statistics*, 35, 634–672.

Stone, C. (1985), 'Additive Regression and Other Nonparametric Models', *Annals of Statistics*, 36, 228–260.

Van der Vaart, A. (1998), *Asymptotic Statistics*, Cambridge University Press.

7. Tables and figures

Table 1. Finite sample bias (B), standard error (SE), 0.95 coverage probability (COV), average length (AL), their bootstrap versions (BCOV) and (BAL) for estimators and confidence intervals of θ_{10} in the QIF model under correct specification

	B $(\text{x}10^{-2})$	SE$(\times 10^{-2})$	COV	AL	BCOV	BAL
$n = 100$						
EL	2.892	9.264	0.912	0.373	0.910	0.327
ET	3.113	8.112	0.904	0.380	0.909	0.332
$ETEL$	3.110	9.143	0.908	0.386	0.909	0.343
$ELEU$	2.866	8.044	0.904	0.356	0.908	0.329
$ETEU$	3.056	8.054	0.906	0.354	0.910	0.322
GMM	3.919	10.354	0.907	0.366	0.909	0.323
$n = 400$						
EL	2.086	2. 632	0.921	0.165	0.919	0.153
ET	2.122	2. 556	0.914	0.175	0.920	0.156
$ETEL$	2.194	2. 571	0.912	0.175	0.918	0.158
$ELEU$	2.090	2. 522	0.908	0.163	0.912	0.146
$ETEU$	2.099	2.563	0.914	0.162	0.916	0.151
GMM	2.501	2. 677	0.913	0.183	0.918	0.156

Table 2. Finite sample bias (B), standard error (SE), 0.95 coverage probability (COV) average length (AL), their bootstrap versions (BCOV) and (BAL) for estimators and confidence intervals of θ_{1*} in the QIF model under misspecification

	B$(\text{x}10^{-2})$	SE$(\text{x}10^{-1})$	COV	AL	BCOV	BAL
$n = 100$						
EL	9. 637	12.299	0.827	0.457	0.899	0.401
ET	4. 564	2. 706	0.895	0.389	0.904	0.354
$ETEL$	5. 103	3.116	0.900	0.398	0.903	0.345
$ELEU$	9. 899	11. 832	0.826	0.475	0.891 .	0.398
$ETEU$	5.897	3.564	0.895	0.406	0.899	0.378
GMM	9.886	8.225	0.870	0.428	0.899	0.384
GMM_I	9.407	8.126	0.889	0.416	0.895	0.382
$n = 400$						
EL	9.132	11. 450	0.841	0.450	0.897	0.398
ET	3. 005	1. 534	0.912	0.199	0.912	0.210
$ETEL$	3. 451	1. 675	0.913	0.194	0.908	0.195
$ELEU$	8. 899	9.877	0.841	0.469	0.894	0.410
$ETEU$	3.798	2.104	0.912	0.228	0.912	0.185
GMM	7.432	7.543	0.856	0.326	0.886	0.312
GMM_I	7. 019	6.896	0.897	0.282	0.899	0.289

Table 3. Finite sample bias (B), standard error (SE), 0.95 coverage probability (COV) average length (AL), their bootstrap versions (BCOV) and (BAL) for estimators and confidence intervals of θ_{20} in the IV partially linear additive model under correct specification

	B$\left(\times 10^{-2}\right)$	SE$\left(\times 10^{-1}\right)$	COV	AL	BCOV	BAL
$n = 100$						
EL	-2.152	1.903	0.901	0.745	0.923	0.731
ET	-2.323	1.877	0.912	0.735	0.913	0.725
$ETEL$	-2.889	1.896	0.905	0.743	0.917	0.728
$ELEU$	-2.165	1.910	0.908	0.748	0.923	0.741
$ETEU$	-2.932	1.902	0.912	0.745	0.928	0.702
GMM	3.123	2.457	0.918	0.960	0.926	0.898
$n = 400$						
EL	-1.086	0.915	0.920	0.358	0.928	0.324
ET	-1.091	0.976	0.927	0.382	0.938	0.316
$ETEL$	-1.098	0.943	0.913	0.369	0.925	0.343
$ELEU$	-1.097	0.930	0.918	0.364	0.923	0.332
$ETEU$	-1.206	0. 932	0.918	0.365	0.927	0.351
GMM	1.466	1.062	0.931	0.416	0.937	0.391

Table 4. Finite sample bias (B),standard error (SE), 0.95 coverage probability (COV) average length (AL), their bootstrap versions (BCOV) and (BAL) for estimators and confidence intervals of θ_{2*} in the IV partially linear additive model under misspecification

	B$\left(\text{x}10^{-1}\right)$	SE$\left(\times 10^{-1}\right)$	COV	AL	BCOV	BAL
$n = 100$						
EL	-0.988	6.901	0.845	1.346	0.871	1.312
ET	-0.323	2.321	0.897	0.876	0.896	0.749
$ETEL$	-0.331	2.272	0.899	0.858	0.901	0.756
$ELEU$	-1.145	7.002	0.813	1.297	0.880	1.235
$ETEU$	-0.373	2.823	0.888	0.891	0.898	0.807
GMM	0.788	5.123	0.857	1.231	0.866	1.043
GMM_I	0.556	4.675	0.869	1.118	0.875	0.996
$n = 400$						
EL	-0.906	5.996	0.879	1.075	0.882	1.071
ET	-0.199	1.328	0.902	0.584	0.905	0.496
$ETEL$	-0.195	1.237	0.904	0.541	0.906	0.499
$ELEU$	-0.989	6.597	0.851	1.063	0.862	1.052
$ETEU$	-0.147	1.315	0.898	0.662	0.901	0.512
GMM	0.453	3.998	0.875	1.112	0.872	0.945
GMM_I	0.400	3.441	0.880	0.938	0.883	0.927

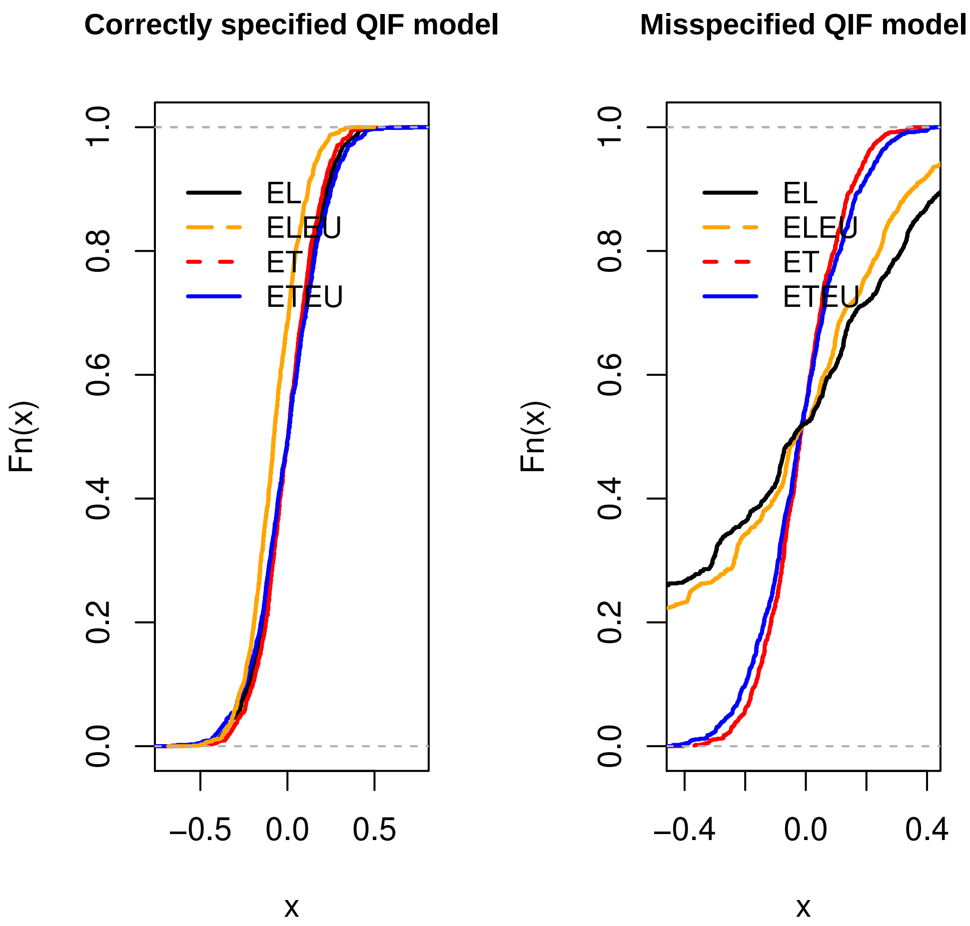

Figure 1. Distributions of the centered EL, ELEU, ET and ETEU estimators of the QIF model under correct specification (left panel) and misspecification (right panel).

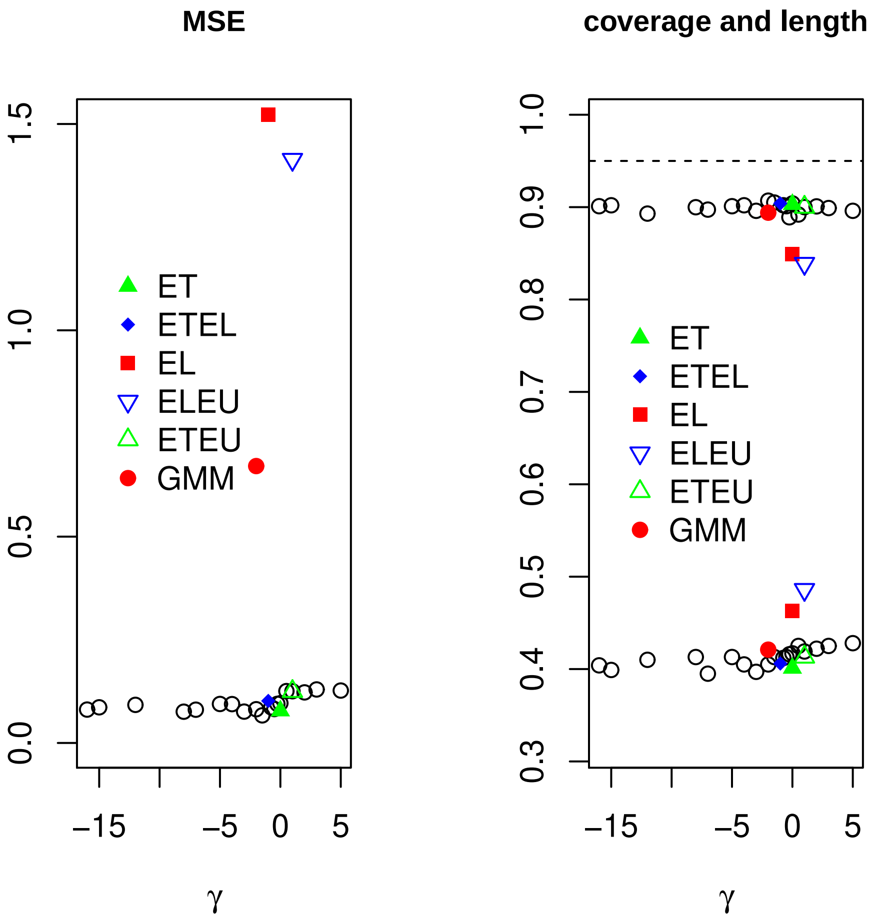

Figure 2. MSE, coverage and average length for the additional $\widehat{\theta}_{CECR}$ estimators and for the estimators of Tables 1 and 2 of the QIF model under misspecification. Note that the position on the γ -axis of the GMM estimator is arbitrarily set at $\gamma = -2$.

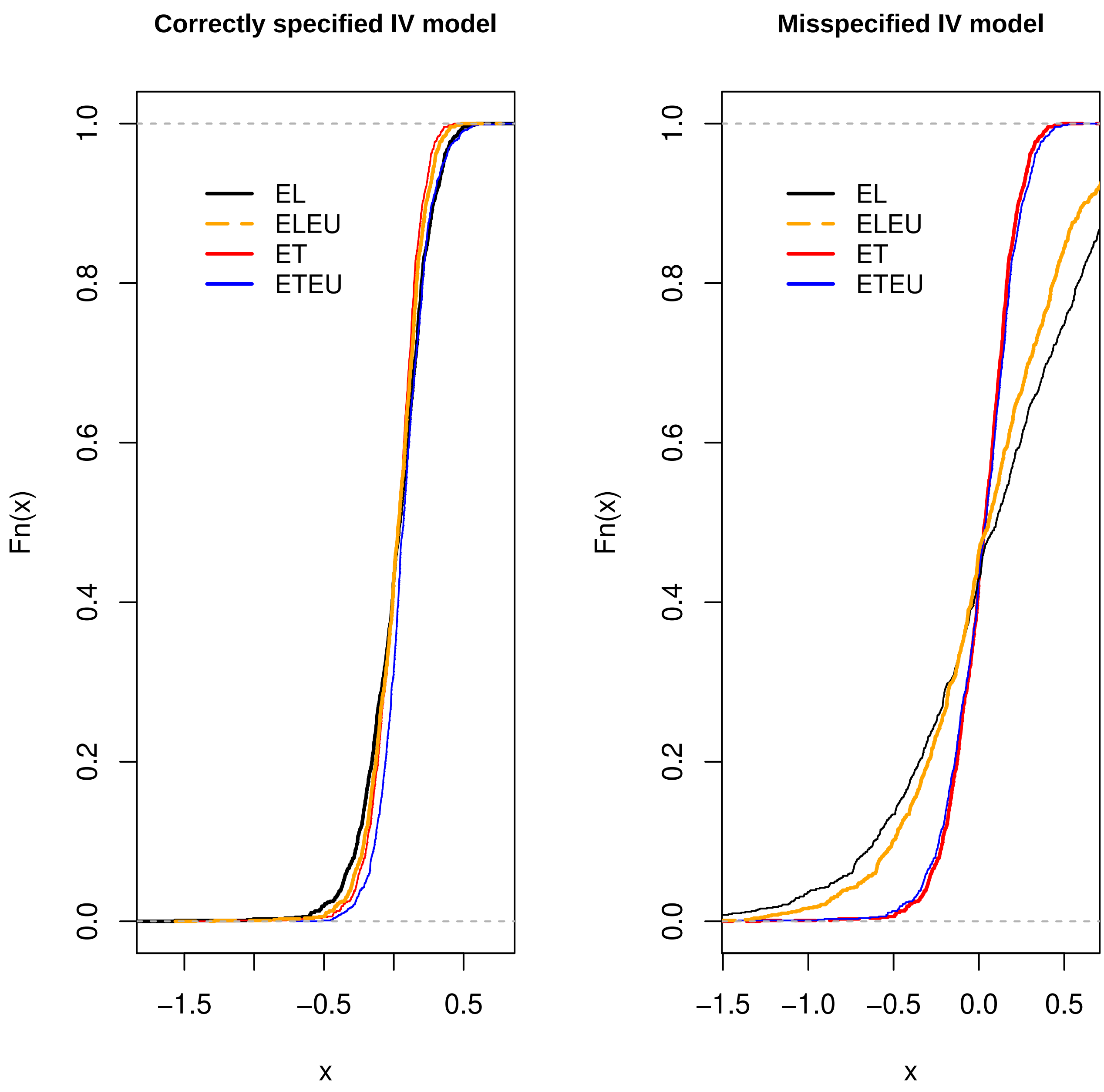

Figure 3. Distributions of the centered EL, ELEU, ET and ETEU estimators of the IV PLA model under correct specification (left panel) and misspecification (right panel).

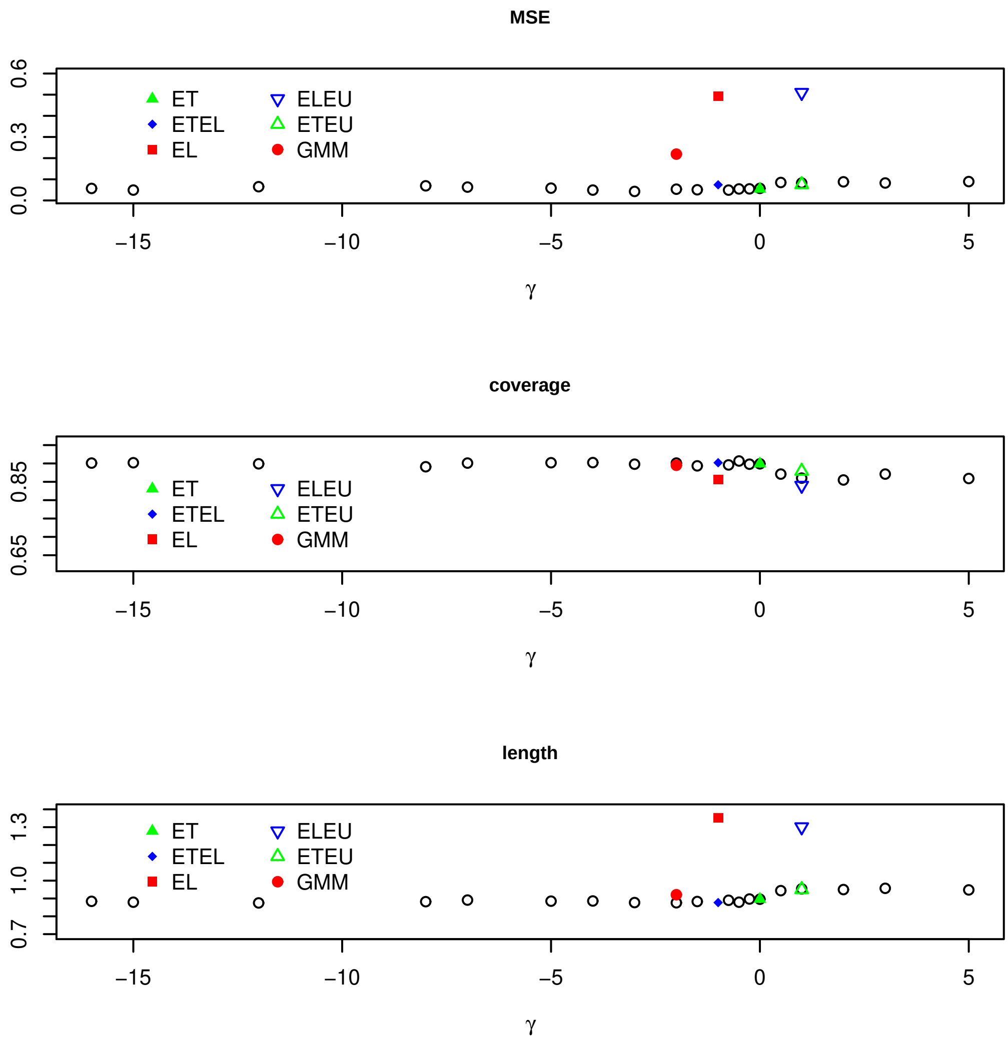

Figure 4. MSE, coverage and average length for the additional $\widehat{\theta}_{CECR}$ estimators and for the estimators of Tables 3 and 4 of the IV PLA model under misspecification. Note that the position on the γ -axis of the GMM estimator is arbitrarily set at $\gamma = -2$.