Efficient M-estimators with auxiliary information*

Francesco Bravo†
University of York

First version September 2008
Revised September 2009

Abstract

This paper introduces a new class of M-estimators based on generalised empirical likelihood (GEL) estimation with some auxiliary information available in the sample. The resulting class of estimators is efficient in the sense that it achieves the same asymptotic lower bound as that of the efficient generalised method of moment (GMM) estimator with the same auxiliary information. The paper also shows that in case of smooth estimating equations the proposed estimators enjoy a small second order bias property compared to both efficient GMM and full GEL estimators. Analytical formulae to obtain bias corrected estimators are also provided. Simulations show that with correctly specified auxiliary information the proposed estimators and in particular those based on empirical likelihood outperform standard M and efficient GMM estimators both in terms of finite sample bias and efficiency. On the other hand with moderately misspecified auxiliary information estimators based on the nonparametric tilting method are typically charactersed by the best finite sample properties.

Keywords and Phrases: Asymptotic efficiency, Generalised empirical likelihood, Generalised method of moments, Second order bias.

*This paper is dedicated to Jemima.

†I would like to thank Peter Phillips for helpful comments. Address correspondence to: Francesco Bravo, Department of Economics and Related Studies, University of York, York, Y010 5DD, United Kingdom. email:fb6@york.ac.uk

1 Introduction

Since the seminal paper of Huber (1964) M-estimators, which are generalisations of the usual maximum likelihood estimators, have played an important role in statistical theory; see for example Van der Vaart (1998, Chapter 5). In this paper we introduce a new class of M-estimators, which is motivated by the fact that in many situations of practical interest we may have some auxiliary information about the otherwise unknown distribution F of the sample. For example we might know the probability that the observed data belong to a certain part of the sample space, or that F has given known moments (joint or marginal), or that is symmetric around a certain constant. This information is often available from auxiliary data such as national statistics or the census. Alternatively the auxiliary information can be a direct by-product of a given theoretical model. In these situations we might expect that incorporating such information into the estimation process can reduce the bias and increase the efficiency of the parameter estimates. For example Imbens and Lancaster (1994) and Hellerstein and Imbens (1999) use auxiliary information within a generalised method of moments (GMM) regression framework, whereas Handcock, Houvilainen and Rendall (2000) combine sample and auxiliary information within generalised linear models. Imbens and Lancaster (1994) report substantial efficiency gains in the parameter estimates by incorporating marginal moments from Census data.

The main objective of this paper is to propose a simple two-step method to incorporate auxiliary information into an M-estimation process. The method is based on the generalised empirical likelihood (GEL) estimator developed by Smith (1997) (see also Newey and Smith (2004) and references therein). To be specific in the first step GEL is used to obtain an estimator of F that is consistent with the auxiliary information available in the sample. This estimator is typically more efficient than the empirical distribution function normally used in nonparametric settings and puts unequal weight on each of the observations. In the second step the parameters of interest are then estimated using the same estimating equations that would have been used if the auxiliary information was not available, but with the contribution of each observation multiplied by its corresponding weight. This weighted estimation procedure defines a new class of M-estimators (WM-estimators henceforth) that typically will be more efficient than usual M-estimators. Intuitively, the latter are based on an estimator of F - the empirical distribution function- that is not efficient in presence of auxiliary information, whereas the former are based on an estimator - the GEL distribution function- that by construction makes effective use of this information.

The two-step estimation method of this paper is a generalisation of that proposed by Zhang (1995) and Hellerstein and Imbens (1999) - see also Owen (2001, Chapter 3.11). These authors use empirical likelihood to obtain the weights to be used in the estimation. Empirical likelihood however is only one of the possible estimators that can be used; one could in fact use Owen's (1991) euclidean likelihood. Another possibility is to use Efron's (1981) nonparametric tilting, or the more general empirical Cressie-Read statistic as defined by Baggerly (1998). These methods differ from each other either in terms of computational complexity or in terms of enjoying desirable statistical properties. For example, Brown and Chen (1998) used Euclidean likelihood because of its computational simplicity, whereas Imbens, Spady and Johnson (1998) used nonparametric tilting because of its robustness and numerical stability. They also typically differ in terms of finite sample properties. On the other hand all of these methods share a common structure of being members of the general class of GEL. Thus GEL provides a general and convenient unifying method to obtain a large class weighted estimators.

The two-step estimation method of this paper can be related to other methods including GMM, full (or one-step) GEL (Parente and Smith, 2005), and parametric likelihood estimation. All of these methods include the auxiliary information directly into the estimation process and produce M-estimators that are asymptotically equivalent to those obtained in this paper (i.e. they have the same asymptotic variance). However the proposed two-step procedure seems to be preferrable to these alternatives for two reasons: First it is computationally simpler because it involves two separate optimisation problems, which are typically easier to solve numerically especially for highly nonlinear models. Second in the case of smooth estimating equations the resulting WM-estimators enjoy a small second order bias property, that is the biases have less components than those based on both GMM and full GEL estimators, which in fact tend to be more biased in finite samples - see the simulations presented in Section 4 for some evidence. This interesting property is a direct consequence of the different way the auxiliary information is incorporated into the estimation process (i.e. directly in the case of GMM and GEL, indirectly in the case of the two-step estimation), and of the fact that the auxiliary information does not contain nuisance parameters. Indeed with nuisance parameters the property would typically not hold. Perhaps more importantly the resulting WM-estimators would not be asymptotically equivalent to those based on either GMM or full GEL estimation and would be typically inefficient.

In this paper we make several contributions: first we establish consistency and asymptotic normality of the WM-estimators based on GEL estimation of auxiliary information. We show that they are efficient in the sense that they have the same asymptotic variance as that of the efficient GMM estimator with the same auxiliary information. Second we show how GEL can be used to consistently estimate the asymptotic variances of the WM-estimators. Third we consider the case where the auxiliary information is misspecified (i.e. it is inaccurate), and investigate the asymptotic properties of the WM-estimators under local misspecification. Fourth we obtain expressions for the second order biases of the WM-estimators and compare them with those of GMM and GEL estimators. These expressions can be used to obtain analytical bias corrected versions of all of these estimators. Finally we illustrate the results with two empirically relevant examples: an instrumental variable quantile regression model and a binary dependent variable regression model. for these two models we use simulations to assess and compare the finite sample performances of the WM, standard M and efficient GMM estimators with both correct and moderately misspecified auxiliary information.

The results of this paper are quite general and can be used in practice to improve the efficiency of a large number of M-estimators defined both by smooth and nonsmooth estimating equations, including the robust estimators of Huber (1973), the regression quantiles of Koenker and Basset (1978) and the trimmed least squares of Powell (1986) among others.

The rest of the paper is structured as follows: next section describes briefly GEL estimation. Section 3 contains the main results, whereas Section 4 illustrates the results of this paper with two examples, and reports the results of the simulations. Section 5 contains some concluding remarks. An appendix contains all the proofs.

The following notation is used throughout the paper: "$a.s.$" stands for almost surely, $\stackrel{a.s.}{\to}$, $\stackrel{p}{\to}$, $\stackrel{d}{\to}$ denote convergence almost surely, in probability and in distribution, respectively, and $\|\cdot\|$ denotes the Euclidean norm. Finally "^{T}" denotes transpose, while "$'$" denotes derivative.

2 GEL estimation with auxiliary information

We begin this section with a simple example which motivates the two-step estimation procedure proposed in this paper.

Example 1.(Hellerstein and Imbens, 1999) Let x denote a random variable with unknown distribution F, and suppose we want to estimate the population mean μ. Without any auxiliary information about F the sample mean $\overline{x} = \sum_{i=1}^{n} x_i/n$ is the efficient estimator for μ. Consider now estimation of μ knowing that $\Pr(x > 0) = p$. While $\overline{x}$ is still consistent, it is no longer efficient. The efficient estimator for μ is in fact the weighted average $\overline{x}_p = p\overline{x}_1 + (1-p)\,\overline{x}_0$ where $\overline{x}_1 = \sum_{i=1}^{n} x_i I\{x_i > 0\} / \sum I\{x_i > 0\}$ and $\overline{x}_0 = \sum_{i=1}^{n} x_i I\{x_i \leq 0\} / \sum I\{x_i \leq 0\}$. This estimator can also be written as $\overline{x}_p = \sum_{i=1}^{n} w_i x_i/n$ where $w_i = (p/\overline{p})^{I\{x_i>0\}} \left[(1-p)/(1-\overline{p})\right]^{I\{x_i \leq 0\}}$ and $\overline{p} = \sum I\{x_i > 0\}/n$ and note that the asymptotic normalised variance of $\overline{x}_p$ is $E[V(x|I\{x_i > 0\})]$ so that

$$n\left[V(\overline{x}) - V(\overline{x}_p)\right] = V(x) - E\left[V(x|I\{x_i > 0\})\right] = V\left[E(x|I\{x_i > 0\})\right] > 0$$

as $n \to \infty$.

Example 1 clearly shows that incorporating weights obtained from available auxiliary information into an estimation process can increase its precision. It is precisely this type of weighted estimation that we are going to focus on in this paper.

Let $\{x_i\}_{i=1}^{n}$ be a random sample from an unknown distribution F with support $X \subset \Re$. Suppose that there exists some auxiliary information about F that can be expressed as a "moment function"

$$\int g(x)\,dF(x) = E[g(x)] = 0, \tag{1}$$

where $g(x)$ is an $\Re^q$-valued vector of functionally independent measurable functions. To describe how GEL estimation can be used in (1), let $\rho(v)$ denote a function of a scalar v that is concave on its domain, an open interval V containing 0. Let $V_n := \{\lambda : \lambda^{\tau} g(x_i) \in \mathcal{V},\ i = 1, ..., n\}$ and define the GEL class of functions

$$\mathcal{G}_n(\lambda) = \sum_{i=1}^{n} \rho(\lambda^{\tau} g(x_i))/n$$

where λ is an $\Re^q$-valued vector of unknown parameters. $\mathcal{G}_n(\lambda)$ includes as special cases empirical likelihood (EL) with $\rho(v) = \log(1-v)$ and $V = (-\infty, 1)$, (NT) nonparametric tilting with $\rho(v) = -\exp(v)$, Euclidean likelihood (EU) with $\rho(v) = -(1+v)^2/2$ and the family of empirical Cressie-Read statistics (ECR) with $\rho(v) = -(1+v)^{(1+\delta)/\delta}/(1+\delta)$ where $\delta \in \Re$ is a user-specified constant. In the rest of the paper we impose the following normalisation on $\rho(v)$: let $\rho_j(v) = d^j\rho(v)/dv^j$ and $\rho_j := \rho_j(0)$ $(j = 1, 2, ...)$; we normalise so that $\rho_1 = \rho_2 = -1$.[1]

[1] As long as $\rho_1 \neq 0$ and $\rho_2 < 0$ (which we will assume to be true) this normalisation can always

Let $\widehat{\lambda} := \arg\max_{\lambda \in V_n} G_n(\lambda)$; the estimated weights

$$\widehat{w}_i = \rho_1\left(\widehat{\lambda}^\tau g(x_i)\right) / \sum_{j=1}^n \rho_1\left(\widehat{\lambda}^\tau g(x_j)\right), \tag{2}$$

sum to one by construction, satisfy the sample moment condition $\sum_{i=1}^n \widehat{w}_i g(x_i) = 0$ when the first order conditions for $\widehat{\lambda}$ hold (by the strong law of large numbers), and are positive when $\widehat{\lambda}^\tau g(x_i)$ is uniformly small in i. Thus they can be interpreted as implied probabilities which incorporate the auxiliary information as defined in (1). Given $\widehat{w}_i$ the GEL distribution function estimator of F is defined as

$$\widehat{F}_w(x) = \sum_{i=1}^n \widehat{w}_i I\{x_i \le x\}.$$

The following theorem summarises the basic asymptotic properties of $\widehat{\lambda}$ and $\widehat{F}_w(x)$; let $E[g(x) g(x)^\tau] := \Sigma$.

Theorem 1 *Assume that $E\|g(x)\|^\beta < \infty$ for some $\beta > 2$, Σ is positive definite, and $\rho(v)$ is twice continuously differentiable in a neighbourhood of 0. Then $\widehat{\lambda} := \arg\max_{\lambda \in V_n} G_n(\lambda)$ exists a.s. and*

$$n^{1/2}\widehat{\lambda} \xrightarrow{d} N\left(0, \Sigma^{-1}\right). \tag{3}$$

Moreover

$$n^{1/2}\left(\widehat{F}_w(x) - F(x)\right) \xrightarrow{d} N\left(0, V_{F_w}(x)\right), \tag{4}$$

where $V_{F_w}(x) = F(x)(1 - F(x)) - E[g(x_i) I\{x_i \le x\}]^\tau \Sigma^{-1} E[g(x_i) I\{x_i \le x\}]$.

Equation (4) clearly shows that in presence of (1) the estimator $\widehat{F}_w(x)$ based on the implied probabilities (2) is more efficient than the empirical distribution function $\widehat{F}_n(x) = \sum_{i=1}^n I\{x_i \le x\}/n$. It is precisely this efficiency property of $\widehat{F}_w(x)$ that will be used in the rest of the paper to obtain more efficient M-estimators.

3 Main results

Let $\psi(x, \theta) : \Re^d \times \Theta \to \Re^k$ denote a known vector of functions up to θ_0 such that

$$\Psi(\theta) := E\psi(x, \theta) \tag{5}$$

---

be imposed by replacing $\rho(v)$ by $\left(-\rho_2/\rho_1^2\right)\rho[(\rho_1/\rho_2) v]$. It is satisfied by EL, NT and ECR among others.

and

$$\Psi(\theta) = 0 \text{ at } \theta = \theta_0$$

where $\theta_0 \in int\{\Theta\}$, and $\Theta \subset \Re^k$ is the parameter space. In a fully nonparametric setting, an M-estimator $\widehat{\theta}$ of θ_0 solves approximately

$$\left\|\Psi_n\left(\widehat{\theta}\right)\right\| \leq \inf_{\theta\in\Theta} \|\Psi_n(\theta)\| + o_{a.s.}(1), \tag{6}$$

where

$$\Psi_n(\theta) := \int \psi(x,\theta)\, d\widehat{F}_n(x) = \sum_{i=1}^{n} \psi(x_i,\theta)/n$$

is the sample analogue of (5). Note that in the most recent statistical literature an estimator that solves (6) is also referred as Z estimator (see Van der Vaart (1998)). Note also that if $\psi(x,\theta)$ is smooth (6) simplifies to the more familiar

$$\widehat{\theta} := \arg\min_{\theta\in\Theta} \|\Psi_n(\theta)\|.$$

3.1 Correctly specified auxiliary information

Suppose that there exists auxiliary information about F available in the moment form given in (1). In order to include such information into the estimation process, let

$$\Psi_w(\theta) := \int \psi(x,\theta)\, d\widehat{F}_w(x) = \sum_{i=1}^{n} \widehat{w}_i \psi(x_i,\theta)$$

denote the weighted sample analogue of (5) where the implied probabilities $\widehat{w}_i$ are as in (2). Then we define the class of WM-estimators $\widehat{\theta}_w$ as

$$\left\|\Psi_w\left(\widehat{\theta}_w\right)\right\| \leq \inf_{\theta\in\Theta} \|\Psi_w(\theta)\| + o_{a.s}(1). \tag{7}$$

The following theorem establishes the strong consistency of $\widehat{\theta}_w$.

Theorem 2 *Suppose that (I) the parameter space Θ is a compact set, (II) for all $\zeta > 0$ $\inf_{\|\theta-\theta_0\|>\zeta} \|\Psi(\theta)\| \geq \varepsilon(\zeta) > 0$, (III) $\sup_{\theta\in\Theta} \|\Psi_n(\theta) - \Psi(\theta)\| = o_{a.s.}(1)$. Then, under the assumptions of Theorem 1 $\widehat{\theta}_w \overset{a.s.}{\to} \theta_0$.*

The conditions of Theorem 2 are fairly standard in both the statistical and econometric literature on nonlinear models estimation. Sufficient conditions for the uniform convergence (III) to hold are: (I), $\psi(x,\theta)$ continuous at each $\theta \in \Theta$ $a.s.$, and $E\sup_{\theta\in\Theta} \|\psi(x,\theta)\| < \infty$. Note however that the (III) is often stronger than needed for consistency of the estimator. The following theorem replaces uniformity with monotonicity as in Huber (1964). Assume that $\psi(x,\theta) : \Re \times \Theta \to \Re$ and $\Theta \subset \Re$.

Theorem 3 *Suppose that (I) the parameter space* Θ *is an open interval, (II) for all* $\zeta > 0$ $\inf_{|\theta-\theta_0|>\zeta} |\Psi(\theta)| \geq \varepsilon(\zeta) > 0$, *(III) there exists a neighbourhood* N_0 *of* θ_0 *such that* $E\sup_{N_0} |\psi(x,\theta)| < \infty$, *(IV)* $\psi(x,\theta)$ *is continuous and monotone in* θ. *Then, under the assumptions of Theorem 1* $\widehat{\theta}_w \overset{a.s.}{\to} \theta_0$.

The following theorem establishes the asymptotic normality for the GEL-based WM-estimator $\widehat{\theta}_w$ satisfying (7) without assuming smoothness of $\psi(x,\theta)$.

Theorem 4 *Suppose that* $n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) = O_p(1)$, *and (I) there exists a finite non-singular matrix* Γ *such that* $\lim_{\|\theta-\theta_0\|\to 0} \|\Psi(\theta) - \Gamma(\theta-\theta_0)\| = o(\|\theta-\theta_0\|)$, *(II) for all positive* $\delta_n \to 0$ $\sup_{\|\theta-\theta_0\|\leq\delta_n} \|\Psi_n(\theta) - \Psi(\theta) - \Psi_n(\theta_0)\| = o_p\left(n^{-1/2}\right)$ *(III) (a)* $\psi(x,\theta)$ *is continuous at* θ_0 *a.s. (b) there exists a neighbourhood* N_0 *of* θ_0 *such that* $E\sup_{N_0} \|\psi(x,\theta) g(x)\| < \infty$ *(IV)* $n^{1/2}\Psi_n(x,\theta_0) \overset{d}{\to} N(0, V(\theta_0))$, *(V)* $\theta_0 \in int\{\Theta\}$. *Then under the assumptions of Theorem 1*

$$n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) \overset{d}{\to} N\left(0, \Sigma_{\Gamma g}(\theta_0)\right),$$

where

$$\Sigma_{\Gamma g}(\theta_0) = \Gamma^{-1}\left\{V(\theta_0) - E\left[\psi(x,\theta_0) g(x)^\tau\right]\Sigma^{-1}E\left[\psi(x,\theta_0) g(x)^\tau\right]^\tau\right\}\left(\Gamma^\tau\right)^{-1}.$$

As with Theorem 2, the conditions of Theorem 4 are fairly standard. Sufficient conditions for the $n^{1/2}$-consistency condition to hold are that $\widehat{\theta}_w$ satisfies $\left\|\Psi_w\left(\widehat{\theta}_w\right)\right\| \leq \inf_{\theta\in\Theta}\|\Psi_w(\theta)\| + o_p\left(n^{-1/2}\right)$ together with the local differentiability of $\Psi(\theta)$ (I), the local stochastic equicontinuity (II) and a central limit theorem (IV).

The following theorem establishes the asymptotic normality for the WM-estimator $\widehat{\theta}_w$ using conditions similar to those used by Huber (1964, Lemma 4).

Theorem 5 *Suppose that* $\widehat{\theta}_w$ *satisfies* (7), $\widehat{\theta}_w \overset{p}{\to} \theta_0$, *and (I)* $\Psi(\theta)$ *is differentiable at* $\theta = \theta_0$ *with* $\Psi'(\theta_0) \neq 0$ *(II)* $\psi(x,\theta)$ *is monotone in* θ, *(III)* $E\left[\psi^2(x,\theta)\right]$ *and* $E\left[\psi(x,\theta) g(x)\right]$ *are continuous at* $\theta = \theta_0$, *(IV) there exists a neighbourhood* N_0 *of* θ_0 *such that* $E\sup_{N_0}\left[\psi^2(x,\theta)\right] < \infty$ *and* $E\sup_{N_0}\left[|\psi(x,\theta)|\,\|g(x)\|\right] < \infty$. *Then, under the assumptions of Theorem 1,*

$$n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) \overset{d}{\to} N\left(0, \sigma^2_{\Psi' g}(\theta_0)\right),$$

where

$$\sigma^2_{\Psi' g}(\theta_0) = \left\{E\psi^2(x,\theta_0) - E\left[\psi(x,\theta_0) g(x)^\tau\right]\Sigma^{-1}E\left[\psi(x,\theta_0) g(x)\right]\right\}/\Psi'(\theta_0)^2. \quad (8)$$

The following theorem establishes the asymptotic normality for the WM-estimator $\widehat{\theta}_w$ assuming that $\psi(x,\theta)$ is differentiable; let $\psi'(x,\theta_0) = \partial\psi(x,\theta)/\partial\theta|_{\theta=\theta_0}$.

Theorem 6 *Suppose that $\widehat{\theta}_w$ satisfies $\Psi_w\left(\widehat{\theta}_w\right) = \inf_{\theta\in\Theta}\|\Psi_w(\theta)\|$, $\widehat{\theta}_w \xrightarrow{p} \theta_0$, and (I) $\psi(x,\theta)$ is continuously differentiable in a neighbourhood N_0 of θ_0, (II) $E[\psi'(x,\theta)]$ is continuous and nonsingular at θ_0, $E\left[\|\psi(x,\theta_0)\|\,\|g(x)\|^2\right] < \infty$, there exists a neighbourhood N_0 of θ_0 such that $E\sup_{N_0}[\|\psi'(x,\theta)\|\,\|g(x)\|] < \infty$ (III) $n^{1/2}\Psi_n(x,\theta_0) \xrightarrow{d} N(0,V(\theta_0))$, (IV) $\theta_0 \in int\{\Theta\}$ Then, under the assumptions of Theorem 1*

$$n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) \xrightarrow{d} N\left(0,\Sigma_{\psi' g}(\theta_0)\right),$$

where

$$\Sigma_{\psi' g}(\theta_0) = [E\psi'(x,\theta_0)]^{-1}\left\{V(\theta_0) - E[\psi(x,\theta_0)g(x)^\tau]\Sigma^{-1}E[\psi(x,\theta_0)g(x)^\tau]^\tau\right\}\times [E\psi'(x,\theta_0)^\tau]^{-1} \quad (9)$$

Theorems 4-6 show that in presence of auxiliary information (1) on F, the asymptotic variances of the weighted estimators $\widehat{\theta}_w$ are always smaller than or equal to the asymptotic variances of the corresponding WM-estimators (6), which are, respectively, $\Gamma^{-1}V(\theta_0)(\Gamma^\tau)^{-1}$, $E\psi^2(x,\theta_0)/\Psi'(\theta_0)^2$ and $[E\psi'(x,\theta_0)]^{-1}V(\theta_0)[E\psi'(x,\theta_0)^\tau]^{-1}$. The reduction in the asymptotic variance will depend on the relevance of the auxiliary information: the larger the correlation between $\psi(x,\theta)$ and $g(x)$ the greater the gain in precision.

Remark 1 Calculations show that $\Sigma_{\Gamma g}(\theta_0)$ corresponds to the asymptotic variance of the efficient GMM estimator is given by

$$I(\theta_0)^{-1} = \left\{[\Gamma,0][Eh(x,\theta_0)h(x,\theta_0)^\tau]^{-1}[\Gamma,0]^\tau\right\}^{-1},$$

where $h(x,\theta) = [\psi(x,\theta)^\tau, g(x)^\tau]^\tau$. Thus the estimators of this paper are efficient in the class of GMM estimators defined as

$$\left\|W_n^{1/2}H_n\left(\widehat{\theta}_{GMM}\right)\right\| \leq \inf_{\theta\in\Theta}\left\|W_n^{1/2}\sum_{i=1}^{n}h(x_i,\theta)/n\right\| + o_{a.s}(1), \quad (10)$$

where W_n is a (possibly random) positive semi-definite weighting matrix. Moreover if we assume that $\psi(x,\theta)$ is differentiable, it is well-known (Chamberlain, 1987) that $I(\theta_0)^{-1}$ is the lower bound for any $n^{1/2}$ consistent regular estimator of θ_0 under $E[h(x,\theta_0)] = 0$. Thus in this case the estimators of this paper are also efficient in the sense that they achieve the (semiparametric) information lower bound for models defined by $E\left[h\left(x,\theta_0\right)\right]=0$.

We now consider the problem of estimating $\Sigma_{\Gamma g}\left(\theta_0\right)$. We propose to use the following GEL-based estimator

$$\widehat{\Sigma}_{\Gamma g}\left(\widehat{\theta}_w\right)=\widehat{\Gamma}_{\widehat{w}}^{-1}\left\{\widehat{V}_{\widehat{w}}\left(\widehat{\theta}_w\right)-\sum_{i=1}^{n}\widehat{w}_i\left[\psi\left(x_i,\widehat{\theta}_w\right)g\left(x_i\right)^{\tau}\right]\widehat{\Sigma}_{\widehat{w}}^{-1}\times\right. \quad (11)$$
$$\left.\sum_{i=1}^{n}\widehat{w}_i\left[\psi\left(x_i,\widehat{\theta}_w\right)g\left(x_i\right)^{\tau}\right]^{\tau}\right\}\left(\widehat{\Gamma}_w^{\tau}\right)^{-1},$$

where $\widehat{V}_{\widehat{w}}\left(\widehat{\theta}_w\right)=\sum_{i=1}^{n}\widehat{w}_i\psi\left(x_i,\widehat{\theta}_w\right)\psi\left(x_i,\widehat{\theta}_w\right)^{\tau}$, $\widehat{\Sigma}_{\widehat{w}}=\sum_{i=1}^{n}\widehat{w}_i g\left(x_i\right)g\left(x_i\right)^{\tau}$, and $\widehat{\Gamma}_{\widehat{w}}$ is an estimator of Γ whose form depends on the smoothness of $\psi\left(x,\theta\right)$. In the smooth case Γ contains ordinary derivatives which can be easily estimated (see Remark 2 below). In the nonsmooth case a general strategy to estimate Γ is to use numerical derivatives, that is

$$\left(\widehat{\Gamma}_{\widehat{w}}\right)_{jl}=\sum_{i=1}^{n}\widehat{w}_i\left[\psi_j\left(x_i,\widehat{\theta}_w+b_n e_l\right)-\psi_j\left(x_i,\widehat{\theta}_w\right)\right]/b_n \quad j,l=1,...,k$$

where $b_n\to 0$ at an appropriate rate as $n\to\infty$, and e_l is lth unit vector. The following theorem establishes the weak consistency of $\widehat{\Sigma}_{\Gamma g}\left(\widehat{\theta}_w\right)$.

Theorem 7 *Suppose that $b_n\to 0$, $b_n^2 n\to\infty$, there exists a neighbourhood N_0 of θ_0 such that $E\sup_{N_0}\left[\left\|\psi\left(x,\theta\right)\right\|^2\right]<\infty$, and that the conditions of Theorems 1, 2, and 4 hold. Then*

$$\widehat{\Sigma}_{\Gamma g}\left(\widehat{\theta}_w\right)\xrightarrow{p}\Sigma_{\Gamma g}\left(\theta_0\right).$$

Remark 2 A practical problem for the computation of (11) is the choice of the size of b_n used to form the numerical derivatives. This is in general a difficult problem, similar in fact to the choice of bandwidth in nonparametric density estimation. In specific cases it is possible to construct estimators that do not involve numerical differentiation. For example if Γ is proportional to (or features) the (unknown) density $f\left(x,\theta\right)$ an alternative estimator for Γ can often be based on kernel methods - see Example 2 below. Another case is when $\psi\left(x,\theta\right)$ is differentiable *a.s.* with derivative that is continuous in θ *a.s.* and dominated by an integrable function. On the other hand in the smooth case $\widehat{\Gamma}_{\widehat{w}}=\sum_{i=1}^{n}\psi_i'\left(\widehat{\theta}_w\right)\widehat{w}_i$, and it is easy to show that (11) is strongly consistent -see Example 4 below.

We finally consider one-step WM-estimators and show that they have the same asymptotic distribution as that of the "fully iterated" WM-estimator $\widehat{\theta}_w$ of Theorem 5. Consider solving $\Psi_w\left(\widehat{\theta}_w\right) = 0$ using the Newton's algorithm starting with $\widehat{\theta}_w$. The full GEL-based WM-estimator is defined as

$$\widehat{\theta}_w^1 = \widehat{\theta}_w - \left[\sum_{i=1}^n \widehat{w}_i \psi'\left(x_i, \widehat{\theta}_w\right)\right]^{-1} \sum_{i=1}^n \widehat{w}_i \psi\left(x_i, \widehat{\theta}_w\right).$$

Theorem 8 *Suppose that the conditions of Theorem 5 hold, and that* $n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) = O_p(1)$ *. Then,*

$$n^{1/2}\left(\widehat{\theta}_w^1 - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma_{\psi' g}(\theta_0)\right),$$

where $\Sigma_{\psi' g}(\theta_0)$ *is as in* (9).

Remark 3 All of the results of this section can be generalised by introducing a sequence of nonsingular random matrices $M_n(x_i, \theta)$ and considering $\|M_w(x_i, \theta)\Psi_w(\theta)\|$, as for example in the classical method of minimum χ^2. As long as $\sup_{\theta\in\Theta}\|M_n(x_i, \theta)\|$ is bounded and converges to a nonsingular asymptotic matrix $M(\theta_0)$ it is not difficult to show that the resulting WM-estimator is consistent and asymptotically normal with covariance

$$\Sigma_{\Gamma g}(\theta_0) = \Gamma^{-1}\left\{M(\theta_0)V(\theta_0)M(\theta_0)^\tau - E\left[M(\theta_0)\psi(x, \theta_0)g(x)^\tau\right] \times \right.$$
$$\left. \Sigma^{-1}E\left[M(\theta_0)\psi(x, \theta_0)g(x)^\tau\right]^\tau\right\}(\Gamma^\tau)^{-1}.$$

3.2 Misspecified auxiliary information

Thus far we assumed that the auxiliary information (1) is correctly specified (i.e. it is accurate, or at least accurate with a negligible sampling error). There are however empirically relevant situations in which this might not be necessarily the case. Therefore it is of interest to investigate what are the consequences of using misspecified (i.e. inaccurate) information on the estimation procedure of this paper. In this section we consider two types of misspecification: a global and a local one. The former can be parameterised as

$$E[g(x)] = \delta \neq 0. \qquad (12)$$

An example of (12) is the situation where the auxiliary information is obtained from a sample that is not compatible with the one used in the estimation, in the sense that the two samples are drawn from a different population. Another example is the situation where there is a measurement error in the auxiliary information. In both cases the function $g(x)$ needs not have zero expectation when the expectation is taken over the sample population.

Remark 4 The proofs of Theorems 1 and 2 show that when (12) is true the parameter estimator $\widehat{\theta}_w$ is in general inconsistent and $n^{1/2}\Psi_n\left(\widehat{\theta}_w\right)$ diverges. This follows because the almost sure limit of the estimator $\widehat{\lambda}$ is not zero, implying that the GEL weights (2) effectively introduce an almost sure non zero term which typically affects the asymptotics of the WM-estimator.

We can test directly whether $E[g(x)] = 0$ using, for example, a GEL or a Wald test statistic, that is

$$\begin{aligned} G_n &= 2\sum_{i=1}^{n}\left[\rho\left(\widehat{\lambda}^{\tau}g(x_i)\right) - \rho(0)\right], \\ W_n &= n\overline{g}^{\tau}\left(\sum_{i=1}^{n} g(x_i)\, g(x_i)^{\tau}/n\right)^{-1}\overline{g} \end{aligned} \tag{13}$$

where $\overline{g} = \sum_{i=1}^{n} g(x_i)/n$. The asymptotic distributions of G_n and W_n are χ^2_p. If the p-values of (13) are reasonably high we should be fairly confident that the auxiliary information available is accurate enough that possibly only a small error is introduced into the M-estimation via the constraint $\sum_{i=1}^{n}\widehat{w}_i g(x_i) = 0$. On the other hand if the p-values are relatively low then the auxiliary information might be moderately misspecified. This situation is empirically relevant, because it is likely that typical sources of auxiliary information such as the Census contain some form of mild misspecification (due for example to the presence of measurement error). In Section 4 we use simulations to investigate the finite sample effects of using this type of misspecified auxiliary information in the weighted estimation.

The second type of misspecification is a local one, that is

$$E[g(x)] = \delta/n^{1/2}. \tag{14}$$

This is a situation in between the assumption of knowledge of correctly specified auxiliary information and that of a globally misspecified information, because it captures the case where the auxiliary information is misspecified for any finite n but the size of the variation is $O\left(n^{-1/2}\right)$ so that it vanishes asymptotically.

Theorem 9 *Suppose that* (14) *holds. Then under the same assumptions of Theorem*

1

$$n^{1/2}\widehat{\lambda} \xrightarrow{d} N\left(\Sigma^{-1}\delta, \Sigma^{-1}\right),$$
$$n^{1/2}\left(\widehat{F}_w(x) - F(x)\right) \xrightarrow{d} N\left(\delta^*, V_{F_w}(x)\right),$$

where $\delta^* = E\left[g(x_i) I\{x_i \leq x\}\right]^\tau \Sigma^{-1}\delta$ *and* $V_{F_w}(x)$ *is the asymptotic variance as defined in* (4).

Theorem 10 *Suppose that* (14) *holds. Then under the same assumptions of Theorems 2 or 3* $\widehat{\theta}_w \xrightarrow{a.s.} \theta_0$.

Theorem 11 *Suppose that* (14) *holds. Then under the same assumptions of Theorems 4-6*

$$n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) \xrightarrow{d} N\left(\Delta_\delta(\theta_0), V_g(\theta_0)\right), \tag{15}$$

where $\Delta_\delta(\theta_0) = \Gamma^{-1}E\left[\psi(x,\theta_0) g(x)^\tau\right]\Sigma^{-1}\delta$ *with* $V_g(\theta_0) = \Sigma_{\Gamma g}(\theta_0)$ *in the case of Theorem 4,* $\Delta_\delta(\theta_0) = E\left[\psi(x,\theta_0) g(x)^\tau\right]\Sigma^{-1}\delta / \left[\Psi'(x,\theta_0)\right]^2$ *with* $V_g(\theta_0) = \sigma^2{}_{\Psi' g}(\theta_0)$ *in the case of Theorem 5, and* $\Delta_\delta(\theta_0) = \left[E\psi'(x,\theta_0)\right]^{-1} E\left[\psi(x,\theta_0) g(x)^\tau\right]\Sigma^{-1}\delta$ *with* $V_g(\theta_0) = \Sigma_{\psi' g}(\theta_0)$ *in the case of Theorem 6.*

Remark 5 Calculations show that the asymptotic distribution of the efficient GMM estimator (and hence that of the full GEL estimator) under the local misspecification (14) is (15). Thus the WM estimators of this paper are asymptotically equivalent to both GMM and GEL estimators under local misspecification.

The following figure shows the effect of local misspecification in terms of finite sample bias of a simple weighted least squares estimator for the regression parameters of $y_i = \theta_0^\tau x_i + \varepsilon_i$ where $\theta_0 = [0, 0.5]^\tau$, $x_i = [1, x_{1i}]^\tau$, and $[x_{1i}, \varepsilon_i]^\tau \sim N(0, I)$. The auxiliary information is parameterised as $E(y) = \delta / n^{1/2}$ where $\delta = 30$, and is estimated by empirical likelihood. Note that values closer to the origin correspond to bigger sample sizes.

Figure 1 approximately here

3.3 Higher order comparisons

The previous two sections showed that under both correct and locally misspecified auxiliary information WM, GMM and (hence) (full) GEL estimators are asymptotically equivalent. In this section we assume that $\psi(x,\theta)$ is smooth and investigate the higher order asymptotic properties of the WM-estimators.

The following theorem gives a third order stochastic expansion for WM-estimators under regularity conditions similar to those used for example by Newey and Smith (2004); let $\partial^k(\cdot) = \partial^k(\cdot)/\theta^{j_1}....\partial\theta^{j_k}$.

Theorem 12 *Suppose that $\widehat{\theta}_w$ satisfies the conditions of Theorem 6, and that (I) $\psi(x,\theta)$ is four times continuously differentiable in a neighbourhood N_0 of θ_0, (II) there exists a neighbourhood N_0 of θ_0 such that for $k = 1, ..., 4$ (a) $E\sup_{\theta\in N_0}\left[\left\|\partial^k\psi(x,\theta)\right\|\right] < \infty$, (b) $E\sup_{\theta\in N_0}\left[\left\|\partial^k\psi(x,\theta)\right\|\left\|g(x)\right\|^k\right]$, (c) $E\left[\left\|\partial^{k-1}\psi(x,\theta_0)\right\|\left\|g(x)\right\|^k\right] < \infty$ (d) $E\left\|g(x)\right\|^k < \infty$, (III) $\rho(v)$ is four times continuously differentiable in a neighbourhood of 0. Then*

$$n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) = Q_1 + Q_2 + Q_3 + O_p\left(n^{-3/2}\right), \tag{16}$$

where $Q_1 \xrightarrow{d} N\left(0, \Sigma_{\psi' g}(\theta_0)\right)$, $\Sigma_{\psi' g}(\theta_0)$ is as in (9), Q_2 and Q_3 are, respectively, an $O_p\left(n^{-1/2}\right)$ quadratic and $O_p\left(n^{-1}\right)$ cubic polynomial in $\psi(x,\theta_0)$ and $g(x)$ whose exact expressions are given in (32) and (33) in the Appendix.

The following corollary gives an explicit expression for the second order bias of $n^{1/2}\left(\widehat{\theta}_w - \theta_0\right)$. Let $tr(\cdot)$ denote the trace operator, $\theta^{(j)}$ denote the jth $(j = 1, ..., k)$ component of θ, and let $g^{\Sigma}(x) = \Sigma^{-1/2}g(x)$ denote the standardised auxiliary information.

Corollary 13 *Under the assumptions of Theorem 12 the second order bias for $n^{1/2}\left(\widehat{\theta}_w - \theta_0\right)$ is given by*

$$Bias\left[n^{1/2}\left(\widehat{\theta}_w - \theta_0\right)\right] = \left[B_{w_1} + (1 + \rho_3/2)\, B_{w_2}\right]/n^{1/2}, \tag{17}$$

where

$$\begin{aligned}
B_{w_1} = {} & \left[E\left(\psi'(x,\theta_0)\right)\right]^{-1}\left\{E\left[\psi'(x,\theta_0)\left(E\left[\psi'(x,\theta_0)\right]\right)^{-1}\psi(x,\theta_0)\right] - \right.\\
& \left. E\left[\sum_{j=1}^{k}\partial\psi'(x,\theta_0)/\partial\theta^{(j)}\right]\left[E\left(\psi'(x,\theta_0)\right)\right]^{-1}V(\theta_0)\left[E\left(\psi'(x,\theta_0)\right)\right]^{-1}/2\right\} - \\
& \left[E\left(\psi'(x,\theta_0)\right)\right]^{-1}E\left[\psi'(x,\theta_0)\left(E\left[\psi'(x,\theta_0)\right]\right)^{-1}E\left(\psi(x,\theta_0)\, g^{\Sigma}(x)^{\tau}\right)g^{\Sigma}(x)\right],
\end{aligned}$$

$$\begin{aligned}
B_{w_2} = {} & \left[E\left(\psi'(x,\theta_0)\right)\right]^{-1}\left\{E\left[\psi(x,\theta_0)\, tr\left(g^{\Sigma}(x)\, g^{\Sigma}(x)^{\tau}\right)\right] - \right.\\
& \left.\left[E\left(\psi(x,\theta_0)\, g^{\Sigma}(x)^{\tau}\right)\right]E\left[g^{\Sigma}(x)\, tr\left(g^{\Sigma}(x)\, g^{\Sigma}(x)^{\tau}\right)\right]\right\}.
\end{aligned}$$

Corollary 13 shows that the bias of the WM-estimators depends on the expected first and second derivative of the estimators, as well as on the (higher order) correlation between the estimating equations (and their derivatives) and the auxiliary information. Corollary 13 also shows that among all of the WM-estimators those based on empirical likelihood (or any other estimator with $\rho_3 = -2$) are the least biased in the sense that their bias is given only by B_{w_1} as opposed to $B_{w_1} + (1 + \rho_3/2)\, B_{w_2}$. Interestingly the same result holds if the higher order correlation between the estimating equation and the auxiliary information and the third moment of the latter are simultaneously zero. Note also that the small bias property of empirical likelihood based WM estimators mirrors that obtained by Newey and Smith (2004) in the case of full GEL estimation of overidentified moment conditions models.

Let

$$\widehat{Bias}\left[n^{1/2}\left(\widehat{\theta}_w - \theta_0\right)\right] = \left[\widehat{B}_{w_1} + (1 + \rho_3/2)\, \widehat{B}_{w_2}\right]/n^{1/2}, \tag{18}$$

where

$$\widehat{B}_{w_1} = \left[\sum_{i=1}^{n} \widehat{w}_i \psi'\left(x_i, \widehat{\theta}_w\right)\right]^{-1} \left\{\sum_{i=1}^{n} \widehat{w}_i \left[\psi'\left(x_i, \widehat{\theta}_w\right)\left(\sum_{i=1}^{n} \widehat{w}_i \psi'\left(x, \widehat{\theta}_w\right)\right)^{-1} \psi\left(x_i, \widehat{\theta}_w\right)\right] - \right.$$
$$\sum_{i=1}^{n}\sum_{j=1}^{k} \widehat{w}_i \partial \psi'\left(x, \widehat{\theta}_w\right)/\partial\theta^{(j)} \left(\sum_{i=1}^{n} \widehat{w}_i \psi'\left(x_i, \widehat{\theta}_w\right)\right)^{-1} \sum_{i=1}^{n} \widehat{w}_i \psi\left(x_i, \widehat{\theta}_w\right)\psi\left(x_i, \widehat{\theta}_w\right)^{\tau} \times$$
$$\left.\left(\sum_{i=1}^{n} \widehat{w}_i \psi'\left(x_i, \widehat{\theta}_w\right)\right)^{-1}/2\right\} - \left[\sum_{i=1}^{n} \widehat{w}_i \psi'\left(x, \widehat{\theta}_w\right)\right]^{-1} \times$$
$$\sum_{i=1}^{n} \widehat{w}_i \left[\psi'\left(x, \widehat{\theta}_w\right)\left(\sum_{i=1}^{n} \widehat{w}\left[\psi'\left(x, \widehat{\theta}_w\right)\right]\right)^{-1} \sum_{i=1}^{n} \widehat{w}_i \left(\psi\left(x, \widehat{\theta}_w\right) g^{\widehat{\Sigma}}(x_i)^{\tau}\right) g^{\widehat{\Sigma}}(x_i)\right],$$

$$B_{w_2} = \left[\sum_{i=1}^{n} \widehat{w}_i \psi'\left(x, \widehat{\theta}_w\right)\right]^{-1} \left\{\sum_{i=1}^{n} \widehat{w}_i \left[\psi\left(x_i, \widehat{\theta}_w\right) tr\left(g^{\widehat{\Sigma}}(x_i)\, g^{\widehat{\Sigma}}(x_i)^{\tau}\right)\right] - \right.$$
$$\left.\left[\sum_{i=1}^{n} \widehat{w}_i \left(\psi\left(x_i, \widehat{\theta}_w\right) g^{\widehat{\Sigma}}(x_i)^{\tau}\right)\right] \sum_{i=1}^{n} \widehat{w}_i \left[g^{\Sigma}(x_i)\, tr\left(g^{\widehat{\Sigma}}(x_i)\, g^{\widehat{\Sigma}}(x_i)^{\tau}\right)\right]\right\},$$
$$\widehat{\Sigma} = \sum_{i=1}^{n} \widehat{w}_i g(x_i)\, g(x_i)^{\tau},$$

denote an estimator of (17). The following corollary shows its strong consistency.

Corollary 14 *Under the same assumptions of Theorem 12*

$$\widehat{Bias}\left[n^{1/2}\left(\widehat{\theta}_w - \theta_0\right)\right] \overset{a.s}{\rightarrow} Bias\left[n^{1/2}\left(\widehat{\theta}_w - \theta_0\right)\right].$$

Remark 6 Given the asymptotic equivalence between the GEL based WM-estimators of this paper and those based on either the efficient GMM or the full GEL methods for the augmented moment condition $h(x,\theta) = \left[\psi(x,\theta)^\tau, g^\Sigma(x)^\tau\right]^\tau$ it seems interesting to make a higher order comparison between them. Using the results of Newey and Smith (2004) some calculations show that the second-order bias of the efficient GMM estimator $\widehat{\theta}_{GMM}$ is

$$Bias\left[n^{1/2}\left(\widehat{\theta}_{GMM} - \theta_0\right)\right] = \left(B_{w_1} + B_{w_2} + B_{h_1} + B_{h_1}^\tau + B_{h_2} + B_{h_2}^\tau\right)/n^{1/2} \tag{19}$$

where B_{w_1}, B_{w_2} are as in (17) and

$$B_{h_1} = -\left[E\left(\psi'(x,\theta_0)\right)\right]^{-1} E\left\{\psi(x,\theta_0)\psi(x,\theta_0)^\tau\left[V(\theta_0) - E\left(\psi'(x,\theta_0)\right)E\left(\psi'(x,\theta_0)\right)\right]^{-1}\times\right.$$
$$\left.E\left(\psi'(x,\theta_0)\right)g^\Sigma(x)\right\} - E\left\{\psi(x,\theta_0)g^\Sigma(x)^\tau\left[V(\theta_0) - E\left(\psi'(x,\theta_0)\right)E\left(\psi'(x,\theta_0)\right)\right]^{-1}\times\right.$$
$$\left.E\left(\psi'(x,\theta_0)\right)g^\Sigma(x)\right\},$$

$$B_{h_2} = \left[E\left(\psi'(x,\theta_0)\right)\right]^{-1}\left[V(\theta_0) - E\left(\psi'(x,\theta_0)\right)E\left(\psi'(x,\theta_0)\right)^\tau\right]^{-1}\left[E\left(\psi'(x,\theta_0)\right)\right]^{-1}\times$$
$$E\left\{\psi'(x,\theta_0)\left[V(\theta_0) - E\left(\psi'(x,\theta_0)\right)E\left(\psi'(x,\theta_0)\right)^\tau\right]^{-1}E\left(\psi'(x,\theta_0)\right)g^\Sigma(x)\right\},$$

whereas the bias for the full GEL estimator $\widehat{\theta}_{GEL}$ is

$$Bias\left[n^{1/2}\left(\widehat{\theta}_{GEL} - \theta_0\right)\right] = \left(B_{w_1} + B_{w_2} + B_{h_1} + B_{h_1}^\tau\right)/n^{1/2}. \tag{20}$$

A simple comparison between (17) with (19) (20) clearly shows that both the efficient GMM and the full GEL estimators have an additional bias terms B_{h_1} and B_{h2}, which arise from the computation of $\sum_{i=1}^n \partial h\left(x_i,\widehat{\theta}\right)/\partial\theta/n$ and of the optimal weight matrix $\left[\sum_{i=1}^n h\left(x_i,\widehat{\theta}\right)h\left(x_i,\widehat{\theta}\right)/n\right]^{-1}$. Thus for the type of auxiliary information considered in this paper WM estimators compare favourably with respect to both efficient GMM and full GEL estimators in terms of second order bias.

Remark 7 Expansion (16) can be used to compute the higher order variance (and/or the mean squared error) of the original and bias corrected version WM-estimators. The resulting expression is extremely complicated and unfortunately does not give any clear indication in terms of which estimator is characterised by the smallest variance (albeit the Monte Carlo evidence presented in the next section seems to favour those based on empirical likelihood when the auxiliary information is correctly specified). On the other hand Newey and Smith (2004) show that among the class of the bias corrected full GEL estimators the empirical likelihood one enjoys the same third order efficiency property as that of the maximum likelihood estimator. They use an indirect argument in which they first show that the empirical likelihood estimator effectively coincides with a multinomial maximum likelihood estimator restricted to satisfy the moment condition, and then use the arguments of Pfanzagl and Wefelmeier (1978) to infer the third order efficiency of the bias corrected empirical likelihood estimator. However the same indirect argument cannot be applied to the weighted estimation procedure proposed in this paper because it is based on a two-step estimator that uses a restricted multinomial estimator that cannot be embedded in Newey and Smith's (2004) general argument.

4 Monte Carlo evidence

In this section we illustrate the theory developed in the paper with three examples: estimation of the slope parameters in an instrumental variable quantile regression model, robust estimation of location, and M-estimation of a binary dependent variable regression model.. The finite sample performance of the usual M-estimator (6), efficient GMM estimator (i.e. as defined in (10) with $W_n = \sum_{i=1}^{n} h\left(z_i, \widetilde{\theta}\right) h\left(z_i, \widetilde{\theta}\right)^{\tau} /n$ and $\widetilde{\theta}$ is a $n^{1/2}$-consistent preliminary estimator of θ_0) and the WM-estimators (7) for all of the examples is assessed by simulations. In addition the simulations are also used to assess the robustness of the WM and efficient GMM estimators to using moderately misspecified auxiliary information, which is identified by a p-value of an empirical likelihood ratio test used to assessed its correctness between approximately 0.10 and 0.25[2].

In the simulations we generate 5000 independent Monte Carlo random samples of sizes $n = 50$ and 100 from a $N(0, 1)$ (standard normal distribution) population, a $t(4)$ (t distribution with four degrees of freedom), $\chi^2(4) - 4$ (centred chi-squared distribution with four degrees of freedom), and $\left(\chi^2(4) - 4\right)/\sqrt{8}$ (standardised chi-squared

[2]With p-values less than 0.10 one would typically reject the hypothesis of correctly specified auxiliary information. With p-values higher than around 0.25 preliminary simulations results suggested that the finite sample behaviour of both WM and efficient GMM estimators is very similar to the case of correctly specified auxiliary information.

distribution with four degrees of freedom). All the computations were carried out in R. For each sample we evaluate biases (B), variances (V) and relative efficiencies (E)[3] of the usual M, GMM and the three WM-estimators that are most used in practice, namely Euclidean likelihood (EU) , nonparametric tilting (NT) and empirical likelihood (EL). The three corresponding implied probabilities (2) to be used in (7) are given, respectively, by

$$\begin{aligned} \widehat{w}_i^{EU} &= 1 - \overline{g}^\tau \overline{\Sigma}^{-1} g(x_i) / \left[n \left(1 - \overline{g}^\tau \overline{\Sigma}^{-1} \overline{g} \right) \right], \\ \widehat{w}_i^{NT} &= \exp\left(\widehat{\lambda}^\tau g(x_i) \right) / \sum_{i=1}^n \exp\left(\widehat{\lambda}^\tau g(x_i) \right), \\ \widehat{w}_i^{EL} &= 1 / \left[n \left(1 - \widehat{\lambda}^\tau g(x_i) \right) \right], \end{aligned} \tag{21}$$

where $\overline{g} := \sum_{i=1}^n g(x_i)/n$, $\overline{\Sigma} := \sum_{i=1}^n g(x_i) g(x_i)^\tau / n$, $\widehat{\lambda} := \arg\max \left[-\sum_{i=1}^n \exp(\lambda^\tau g(x_i)) \right]$ in $\widehat{w}_i^{NT}$ and $\widehat{\lambda} := \arg\max \sum_{i=1}^n \log(1 - \lambda^\tau g(x_i))$ in $\widehat{w}_i^{EL}$.

Remark 8. In general to compute $\widehat{\lambda}$ one can apply the multivariate Newton's algorithm to $\sum_{i=1}^n \rho(\lambda^\tau g(x_i))$; this amounts to Newton's method for solving the nonlinear system of q first-order conditions $\sum_{i=1}^n \rho_1(\lambda^\tau g(x_i)) g(x_i) = 0$ with starting point in the iterative process set to $\lambda_0 = 0^\tau$. For such choice of starting point, the convergence of the algorithm is typically quadratic. Note also that the case of EU there is no need to use any numerical optimisation method to find the maximiser $\widehat{\lambda}$ since the latter can be obtained in closed form and is given by $\widehat{\lambda} = \overline{\Sigma}^{-1} \overline{g}$.

Example 2 Let $x = [y, z_1^\tau, z_2^\tau]^\tau$ and let $q_p(y|z_2) := \inf\{y : F(y|z_2) \geq p\} = z_1^\tau \theta_{p0}$ denote the pth $(0 < p < 1)$ quantile of y conditional on z_2 assumed to have the same dimension of z_1. The instrumental variables quantile regression estimator $\widehat{\theta}_p$ for θ_p solves $\sum_{i=1}^n \psi\left(x_i, \widehat{\theta}_p\right)/n = 0$, where $\psi(x_i, \theta_p) := z_{2i} sign_p\{y_i - z_{1i}^\tau \theta_p\}$, and $sign_p\{\cdot\} = pI\{\cdot \leq 0\} - (1-p) I\{\cdot \geq 0\}$. Let $\varepsilon = y - z_1^\tau \theta_{p0}$ and $z = [z_1^\tau, z_2^\tau]^\tau$; the following proposition establishes the asymptotic distribution of the weighted instrumental variables quantile regression estimator $\widehat{\theta}_{pw}$ solving $\sum_{i=1}^n \widehat{w}_i \psi\left(x_i, \widehat{\theta}_p\right) = 0$.

Proposition 15 *Suppose that* (1) *holds, and (I)* $F_\varepsilon(0|z) = p$, *(II)* Θ *compact, (III)* $E\|z_2\|^2 < \infty$, $E\|z_1 z_2^\tau\| < \infty$, *(IV)* $F_\varepsilon(\cdot|z)$ *is differentiable at 0 with* $F_\varepsilon'(0|z) = f_\varepsilon(0|z) > 0$, *(V)* $E[f_\varepsilon(0|z) z_1 z_2^\tau]$ *is nonsingular, (VI) there exists a neighbourhood* N_0 *of* θ_0 *such that* $E\sup_{N_0} \|\psi(x, \theta) g(x)\| < \infty$, *(VII)* $\theta_{p0} \in int\{\Theta\}$. *Then*

$$n^{1/2}\left(\widehat{\theta}_{pw} - \theta_{p0}\right) \xrightarrow{d} N(0, \Sigma_{\zeta g}(\theta_0)),$$

[3] The relative efficiency of two asymptotically normal estimators say $\widehat{\theta}_a$ and $\widehat{\theta}_b$ is defined as $E = V\left(\widehat{\theta}_b\right)/V\left(\widehat{\theta}_a\right)$.

where

$$\Sigma_{\zeta g}(\theta_0) = \Gamma^{-1}\left\{p(1-p)E\left(z_2 z_2^\tau\right) - E\left[sign_p\{\varepsilon\} z_2 g(x)^\tau\right]\Sigma^{-1}E\left[sign_p\{\varepsilon\} z_2 g(x)^\tau\right]^\tau\right\}\Gamma^{-1},$$

$\Gamma = -E\left[f_\varepsilon(0|z) z_1 z_2^\tau\right]$. *Moreover suppose that (VIII)* $b_n \to 0$, $b_n^2 n \to \infty$, *(III')* $E\|z\|^3 < \infty$,*(IX) there exists a constant such that* $f_\varepsilon(\cdot|z) \leq \overline{f}_\varepsilon$ *for all* z. *Then* $\widehat{\Sigma}_{\zeta g}\left(\widehat{\theta}_{pw}\right) \xrightarrow{p} \Sigma_{\zeta g}(\theta_{p0})$, *where*

$$\widehat{\Sigma}_{\zeta g}\left(\widehat{\theta}_{pw}\right) = \widehat{\Gamma}_{\widehat{w}}^{-1}\left\{p(1-p)\left[\sum_{i=1}^n \widehat{w}_i z_{2i} z_{2i}^\tau\right] - \left[\sum_{i=1}^n \widehat{w}_i sign_p\{\widehat{\varepsilon}_i\} z_{2i} g(x_i)^\tau\right]\right. \quad (22)$$

$$\left.\widehat{\Sigma}_{\widehat{w}}^{-1}\left[\sum_{i=1}^n \widehat{w}_i sign_p\{\widehat{\varepsilon}_i\} z_{2i} g(x_i)^\tau\right]^\tau\right\}\widehat{\Gamma}_{\widehat{w}}^{-1},$$

$\widehat{\Gamma}_{\widehat{w}} = \sum_{i=1}^n \widehat{w}_i I\{|\widehat{\varepsilon}_i| \leq 2b_n\} z_{1i} z_{2i}^\tau / b_n$, $\widehat{\Sigma}_{\widehat{w}}$ *is as in* (11), *and* $\widehat{\varepsilon}_i = y_i - z_{1i}^\tau \widehat{\theta}_{pw}$.

In the simulations we consider median regression estimation of $\theta_0 = [1, 0.5]^\tau$ in $y = z_1^\tau \theta_0 + \varepsilon$ where $z_1 = [z_{11}^*, z_{12}]^\tau$ $z_{11}^* = z_{11} + \varepsilon$ and z_{1j} $(j = 1, 2)$ and ε are $N(0,1)$. The instruments are specified as $z_2 = [z_{21}, z_{22}]^\tau$ and z_{2j} $(j = 1, 2)$ are $N(0,1)$. The auxiliary information consists of the knowledge of two quantiles for the instrument z_{21}, that is $E[g(x)] = [I(z_{21} \leq q) - p]^\tau = 0$ with $p = [0.1, 0.4]^\tau$. For the correctly specified case $g(x)^{cs}$ the values of the quantiles are $q^{cs} = [-1.28, -0.25]^\tau$. For the two moderately misspecified cases $g_1(x)^{ms}$ and $g_2(x)^{ms}$ we use the same random seed 123 and specify for $n = 50$ $q_1^{ms} = [-0.70, -0.06]^\tau$ and $q_2^{ms} = [-0.55, -0.04]^\tau$ which yield average p-values (based on 5000 replications) of the EL ratio test for the hypothesis $E[g(x)^{cs}] = 0$ of 0.200 and 0.114, respectively. For $n = 100$ we specify $q_1^{ms} = [-0.90, -0.10]^\tau$ and $q_2^{ms} = [-0.76, -0.11]^\tau$ which yield average p-values (based on 5000 replications) of the EL ratio test for the hypothesis $E[g(x)^{ms}] = 0$ of 0.207 and 0.117, respectively. Tables 1a and 1b report also the point estimates of (22) with bandwidth b_n chosen by the "Hall-Sheather" rule (Hall and Sheather, 1988).

Tables 1a,b approximately here

Example 3. Huber's (1964) location estimator $\widehat{\theta}$ for θ_0 solves $\sum_{i=1}^n \psi\left(x_i - \widehat{\theta}\right)/n = 0$ where $\psi(\cdot) = \cdot$ for $|\cdot| \leq k$ and $\psi(\cdot) = k sign(\cdot)$ for finite k, or is simply the sample mean for $k = \infty$. The following proposition establishes the asymptotic distribution of the weighted location estimator $\widehat{\theta}_w$ solving $\sum_{i=1}^n \widehat{w}_i \psi\left(x_i, \widehat{\theta}\right) = 0$.

Proposition 16 *Suppose that* (1) *holds, and (I) x is symmetrically distributed around θ_0, (II) Θ is an open interval, (III) $E|x|^2 < \infty$, (IV) $E|x|\,\|g(x)\| < \infty$. Then*

$$n^{1/2}\left(\widehat{\theta}_w - \theta_0\right) \xrightarrow{d} N\left(0, \sigma^2_{\theta g}(\theta_0)\right),$$

where

$$\sigma^2_{\theta g}(\theta_0) = \left\{\int_{\theta_0-k}^{\theta_0+k} x^2 dF(x) + k^2\left(\int_{-\infty}^{\theta_0-k} + \int_{\theta_0+k}^{\infty}\right) dF(x) - \sigma^{\tau}_{\psi g}\Sigma^{-1}\sigma_{\psi g}\right\} / \left[\int_{\theta_0-k}^{\theta_0+k} F(x)\right]^2,$$

and $\sigma_{\psi g} = \left[\int_{\theta_0-k}^{\theta_0+k} x + k\left(\int_{\theta_0+k}^{\infty} - \int_{-\infty}^{\theta_0-k}\right)\right] g(x)\, dF(x)$. *Moreover* $\widehat{\sigma}^2_{\theta g}\left(\widehat{\theta}_w\right) \xrightarrow{a.s} \sigma^2_{\theta g}(\theta_0)$ *where*

$$\widehat{\sigma}^2_{\theta g}\left(\widehat{\theta}_w\right) = \left[\sum_{i=1}^{n} \widehat{w}_i I\left\{\left|x_i - \widehat{\theta}_w\right| \leq k\right\}\right]^{-2} \left\{\sum_{i=1}^{n} \widehat{w}_i x_i^2 I\left\{\left|x_i - \widehat{\theta}_w\right| \leq k\right\} + \right. \tag{23}$$

$$\left. k^2\left(\sum_{i=1}^{n} \widehat{w}_i I\left\{x_i \leq \widehat{\theta}_w - k\right\} + \sum_{i=1}^{n} \widehat{w}_i I\left\{x_i \geq \widehat{\theta}_w + k\right\}\right) - \widehat{\sigma}^{\tau}_{\psi g}\widehat{\Sigma}^{-1}_{\widehat{w}}\widehat{\sigma}_{\psi g}\right\},$$

$$\widehat{\sigma}_{\psi g} = \sum_{i=1}^{n} \widehat{w}_i x_i g(x_i) I\left\{\left|x_i - \widehat{\theta}_w\right| \leq k\right\} + k\sum_{i=1}^{n} \widehat{w}_i g(x_i) I\left\{x_i \geq \widehat{\theta}_w + k\right\} -$$

$$k\sum_{i=1}^{n} \widehat{w}_i g(x_i) I\left\{x_i \leq \widehat{\theta}_w - k\right\}.$$

and $\widehat{\Sigma}_{\widehat{w}}$ *is as in* (11).

In the simulations we consider estimating the location θ when the pth population quantile q is known, so that $E[g(x)] = E(I\{x \leq q\}) - p = 0$. Table 2 reports the finite sample properties of $\widehat{\theta}$ and $\widehat{\theta}_w$ for $p = [0.25, 0.40, 0.60, 0.75]$ for the case $k = 1.5$, including the point estimates $\left(\widehat{V}\right)$ of the variance $\sigma^2_{\theta g}(\theta_0)$ obtained using (23).

Table 2 approximately here

Example 4 Let $x = [y, z^{\tau}]^{\tau}$ for a binary variable $y \in \{0,1\}$ and $F(\cdot)$ denote the cumulative density function with $f(\cdot)$ and $f'(\cdot)$ to denote its density and first derivative. For example for $F(\cdot) = \Phi(\cdot)$ that is the cumulative distribution of a standard normal we have the standard probit model. An M-estimator (optimally weighted Z estimator) $\widehat{\theta}$ for θ_0 solves $\sum_{i=1}^{n} \psi\left(x_i, \widehat{\theta}\right)/n = 0$, where $\psi(x_i, \theta) := (y_i - F(z_i^{\tau}\theta)) f(z_i^{\tau}\theta) z_i / F(z_i^{\tau}\theta) F(-z_i^{\tau}\theta)$, where, with a slight abuse of notation, $F(-z_i^{\tau}\theta) = 1 - F(z_i^{\tau}\theta)$ The following proposition establishes the asymptotic distribution of the weighted M-estimator $\widehat{\theta}_w$ solving $\sum_{i=1}^{n} \widehat{w}_i \psi\left(x_i, \widehat{\theta}_w\right) = 0$.

Proposition 17 *Suppose that* (1) *holds and (I)* Θ *compact,* $E\sup_{\theta\in F}\left\|f\left(z^{\tau}\theta\right)/F\left(z^{\tau}\theta\right)\times F\left(-z^{\tau}\theta\right)\right\|<\infty$*, (II)* $E\left\|z\right\|^{2}<\infty$*, (III)* $E\left(zz^{\tau}\right)$ *is nonsingular, (IV) there exists a neighbourhood* N_0 *of* θ_0 *such that* $E\sup_{N_0}\left\|\psi\left(x,\theta\right)g\left(x\right)\right\|<\infty$*, (V)* $\theta_{p0}\in int\left\{\Theta\right\}$*. Then*

$$n^{1/2}\left(\widehat{\theta}_{w}-\theta_0\right)\xrightarrow{d}N\left(0,\Sigma_{Fg}\left(\theta_0\right)\right),$$

where

$$\Sigma_{Fg}\left(\theta_0\right)=\left[E\left(\psi'\left(x,\theta_0\right)\right)\right]^{-1}\left\{E\left(\psi'\left(x,\theta_0\right)\right)-E\left[\left(y-F\left(z^{\tau}\theta_0\right)\right)\lambda\left(z^{\tau}\theta_0\right)/F\left(-z^{\tau}\theta_0\right)zg\left(x\right)^{\tau}\right]\right.$$
$$\left.\Sigma^{-1}E\left[\left(y-F\left(z^{\tau}\theta_0\right)\right)\lambda\left(z^{\tau}\theta_0\right)/F\left(-z^{\tau}\theta_0\right)z_{i}g\left(x\right)^{\tau}\right]^{\tau}\right\}\left[E\left(\psi'\left(x,\theta_0\right)\right)\right]^{-1},$$

$E\left(\psi'\left(x,\theta_0\right)\right)=E\left[\lambda\left(z^{\tau}\theta_0\right)\lambda\left(-z^{\tau}\theta_0\right)zz^{\tau}\right]$ *and* $\lambda\left(\cdot\right)=f\left(\cdot\right)/F\left(\cdot\right)$*. Moreover* $\widehat{\Sigma}_{Fg}\left(\widehat{\theta}_{w}\right)\xrightarrow{a.s.}\Sigma_{Fg}\left(\theta_0\right)$*, where*

$$\widehat{\Sigma}_{Fg}\left(\widehat{\theta}_{w}\right)=\widehat{\psi}\left(x_i,\widehat{\theta}_{w}\right)_{\widehat{w}}^{-1}\left\{\widehat{\psi}\left(x_i,\widehat{\theta}_{w}\right)_{\widehat{w}}-\left[\sum_{i=1}^{n}\widehat{w}_i\left(y_i-F\left(z_i^{\tau}\widehat{\theta}_{w}\right)\right)\lambda\left(z_i^{\tau}\widehat{\theta}_{w}\right)/F\left(-z_i^{\tau}\widehat{\theta}_{w}\right)\right.\right.$$
$$\left.\left.z_ig\left(x_i\right)^{\tau}\right]\widehat{\Sigma}_{\widehat{w}}^{-1}\left[\sum_{i=1}^{n}\widehat{w}_i\left(y_i-F\left(z_i^{\tau}\widehat{\theta}_{w}\right)\right)\lambda\left(z_i^{\tau}\widehat{\theta}_{w}\right)/F\left(-z_i^{\tau}\widehat{\theta}_{w}\right)z_ig\left(x_i\right)^{\tau}\right]^{\tau}\right\}\widehat{\psi}\left(x_i,\widehat{\theta}_{w}\right)_{\widehat{w}}^{-1}, \quad (24)$$

and $\widehat{\psi}\left(x_i,\widehat{\theta}_{w}\right)_{\widehat{w}}=\sum_{i=1}^{n}\widehat{w}_i\lambda\left(z_i^{\tau}\widehat{\theta}_{w}\right)\lambda\left(-z_i^{\tau}\widehat{\theta}_{w}\right)z_iz_i^{\tau}$*.*

In the simulations we consider estimating $\theta_0=[1,0.5]^{\tau}$ with $z=[1,z_1]^{\tau}$, and z_1 is $N(0,1)$. The auxiliary information consists of the knowledge of the conditional mean of y given $z\geq 0$, that is $E\left[g\left(x\right)\right]=\left[E\left(y|z\geq 0\right)-\mu_{+},E\left(y|z<0\right)-\mu_{-}\right]^{\tau}=0$. For the correctly specified case $g\left(x\right)^{cs}$ the approximate values of $\left[\mu_{+}^{cs},\mu_{-}^{cs}\right]^{\tau}$ are $[0.91,0.71]^{\tau}$ for $N(0,1)$ errors (i.e. standard probit), $[0.87,0.70]^{\tau}$ for $t(4)$ errors, $[0.62,0.49]^{\tau}$ for centred $\chi^2(4)$ errors, and $[0.97,0.68]^{\tau}$ for standardised $\chi^2(4)$ errors. For the two moderately misspecified cases $g_1\left(x\right)^{ms}$ and $g_2\left(x\right)$ we use use the same random seed 123 and specify for $n=50$ $[0.74,0.62]^{\tau}$ and $[0.71,0.59]^{\tau}$ for the $N(0,1)$ case, $[0.71,0.60]^{\tau}$ and $[0.67,0.57]^{\tau}$ for the $t(4)$ case, $[0.45,0.35]^{\tau}$ and $[0.41,0.33]^{\tau}$ for the centred $\chi^2(4)$ case, and finally $[80,60]^{\tau}$ and $[77,56]^{\tau}$ for the standardised $\chi^2(4)$ case. With these values the average p-values (based on 5000 replications) of the EL ratio test for the hypothesis $E\left[g\left(x\right)^{cs}\right]=0$ are, respectively, 0.195 and 0.111, 0.216 and 0.114, 0.197 and 0.115 and finally 0.216 and 0.112. For $n=100$ we specify $\left[\mu_{+1}^{ms},\mu_{-1}^{ms}\right]^{\tau}=[0.78,0.67]^{\tau}$ and $[0.76,0.65]^{\tau}$ for the $N(0,1)$ case, $[0.75,0.64]^{\tau}$ and $[0.73,0.61]^{\tau}$ for the $t(4)$ case, $[0.50,0.40]^{\tau}$ and $[0.47,0.39]^{\tau}$ for the centred $\chi^2(4)$ case, and finally $[0.83,0.65]^{\tau}$ and $[0.81,0.62]^{\tau}$for the standardised $\chi^2(4)$ case. With these values the average p-values (based on 5000 replications) of the EL ratio test for the hypothesis $E\left[g\left(x\right)^{ms}\right]=0$ are, respectively, 0.211 and 0.119, 0.221 and 0.111, 0.204 and 0.123 and finally 0.219 and 0.112.

We also consider the bias corrected M and WM estimators, that is $n^{1/2}\left(\widehat{\theta}_w-\widehat{Bias}\left(\widehat{\theta}_w\right)\right)$ where $\widehat{Bias}\left(\widehat{\theta}_w\right)$ is a consistent estimator (see (18)) of

$$\begin{aligned}
&Bias\left[n^{1/2}\left(\widehat{\theta}_w-\theta_0\right)\right]=\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\left\{E\left[-\left(y-F\left(v_0\right)\right)^2\lambda\left(v_0\right)\times\right.\right. \quad (25)\\
&\lambda\left(-v_0\right)zz^{\tau}/F\left(v_0\right)F\left(-v_0\right)+\left(y-F\left(v_0\right)\right)f'\left(v_0\right)zz^{\tau}/F\left(v_0\right)F\left(-v_0\right)\right]\times\\
&\left.\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\left[f\left(v_0\right)\left(y-F\left(v_0\right)\right)z\right]/F\left(v_0\right)F\left(-v_0\right)\right\}/n^{1/2}+\\
&E\left\{3\sum_{j=1}^{k}\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)f\left(v_0\right)f'\left(v_0\right)zz^{\tau}z^{(j)}\right]\right\}\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}/2n^{1/2}-\\
&\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\left\{E\left[-\left(y-F\left(v_0\right)\right)^2\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}/F\left(v_0\right)F\left(-v_0\right)+\right.\right.\\
&\left.\left(y-F\left(v_0\right)\right)f'\left(v_0\right)zz^{\tau}/F\left(v_0\right)F\left(-v_0\right)\right]\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\times\\
&\left.E\left(\left(y-F\left(z^{\tau}\theta_0\right)\right)\lambda\left(z^{\tau}\theta_0\right)/F\left(-z^{\tau}\theta_0\right)zg^{\Sigma}\left(x\right)^{\tau}\right)g^{\Sigma}\left(x\right)\right\}/n^{1/2}+\\
&\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\left\{E\left[f\left(v_0\right)\left(y-F\left(v_0\right)\right)ztr\left(g^{\Sigma}\left(x\right)g^{\Sigma}\left(x\right)^{\tau}\right)\right]-\right.\\
&\left.E\left[\left(y-F\left(z^{\tau}\theta_0\right)\right)\lambda\left(z^{\tau}\theta_0\right)/F\left(-z^{\tau}\theta_0\right)zg^{\Sigma}\left(x\right)^{\tau}\right]E\left[g^{\Sigma}\left(x\right)tr\left(g^{\Sigma}\left(x\right)g^{\Sigma}\left(x\right)^{\tau}\right)\right]\right\}/n^{1/2},
\end{aligned}$$

where $v_0=z^{\tau}\theta_0$, $z^{(j)}$ is the jth component of z, and

$$g^{\Sigma}\left(x\right)=\left[\left(yI\left(z>0\right)-\mu_0^{+}\right)/\left[\mu_0^{+}\left(1-\mu_0^{+}\right)\right]^{1/2},\left(yI\left(z\leq 0\right)-\mu_0^{-}\right)/\left[\mu_0^{-}\left(1-\mu_0^{-}\right)\right]^{1/2}\right]^{\tau}.$$

The GMM bias corrected estimator is $n^{1/2}\left(\widehat{\theta}_{GMM}-\widehat{Bias}\left(\widehat{\theta}_{GMM}\right)\right)$ where $\widehat{Bias}\left(\widehat{\theta}_{GMM}\right)$ is a consistent estimator of

$$\begin{aligned}
&Bias\left[n^{1/2}\left(\widehat{\theta}_{GMM}-\theta_0\right)\right]=Bias\left[n^{1/2}\left(\widehat{\theta}_w-\theta_0\right)\right]-\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\times\\
&E\left\{\left(y-F\left(v_0\right)\right)^2\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}/F\left(v_0\right)F\left(-v_0\right)\left[E\left(\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\times\right.\right.\\
&\left.\left(I-\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\right]^{-1}E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]g^{\Sigma}\left(x\right)\right\}/n^{1/2}-\\
&E\left\{\left[f\left(v_0\right)\left(y-F\left(v_0\right)\right)zg^{\Sigma}\left(x\right)^{\tau}\right]/F\left(v_0\right)F\left(-v_0\right)\left[E\left(\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\times\right.\right.\\
&\left.\left(I-\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\right]^{-1}E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]g^{\Sigma}\left(x\right)\right\}/n^{1/2}+\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\\
&\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\left[E\left(\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\left(I-\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\right]^{-1}\times\\
&\left\{E\left[\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right]\right\}^{-1}\left\{E\left[-\left(y-F\left(v_0\right)\right)^2\lambda\left(v_0\right)\lambda\left(-v_0\right)/zz^{\tau}/F\left(v_0\right)F\left(-v_0\right)+\right.\right.\\
&\left.\left(y-F\left(v_0\right)\right)f'\left(v_0\right)zz^{\tau}/F\left(v_0\right)F\left(-v_0\right)\right]\left[E\left(\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\left(I-\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)\right]^{-1}\times\\
&\left.E\left(\lambda\left(v_0\right)\lambda\left(-v_0\right)zz^{\tau}\right)g^{\Sigma}\left(x\right)\right\}/n^{1/2}.
\end{aligned}$$

Tables 3a,b and 4a,b report, respectively, the finite sample properties of $\widehat{\theta}$, $\widehat{\theta}_w$, $\widehat{\theta}_{GMM}$ and their bias corrected versions $\widehat{\theta}^{c}$[4], $\widehat{\theta}^c_w$, $\widehat{\theta}^c_{GMM}$ as well as the point estimates $\left(\widehat{V}\right)$ of the variance $\Sigma_{Fg}(\theta_0)$ obtained using (24) with correct and moderately misspecified auxiliary information.

Tables 3 a,b 4 a,b approximately here

We first discuss the results of Tables1a- 4b in the case of correctly specified auxiliary information. First all of the three WM-estimators have finite sample biases that are smaller than those of the original M and GMM estimators. The bias reduction seems to be a little more substantial in the case of symmetric distributions. Second, as clearly expected from Theorems 4-6, all of the three WM estimators have finite sample variances that are uniformly smaller than those of usual M-estimators, and are typically smaller than those of GMM estimators. The efficiency gain (i.e. the magnitude of the variance reduction) of the proposed estimators depends on the type of estimation considered, on the relevance of the auxiliary information and on the shape of the distribution of the observations. Third the variance estimators $(22)-(24)$ work remarkably well with symmetric distributions and both EL and NT weights. Fourth the bias correction is very effective and removes almost completely the finite sample bias for symmetric distributions and drastically reduces that for skewed distributions. The variances of the bias corrected estimators are also reduced. Fifth among the three WM estimators considered, those based on EL weights have an edge over those based on NT and EU weights in terms of efficiency. They also seem to have an edge in terms of finite sample bias. This result is interesting because not only confirms the small bias property of EL based WM-estimators for the case of smooth estimating equations (see Section 3.3), but also because it suggests that this property seems to be holding also for nonsmooth estimating equations. Finally these results hold for both sample sizes, suggesting that the asymptotic approximations are reliable for relatively small sample sizes. The only difference is that the biases and variances are slightly larger for $n = 50$.

We now discuss the results of Tables 1a-4b in the case of moderately misspecified auxiliary information. For the $g_1(x)^{ms}$ cases, that is for cases where the degree of misspecification is relatively low, the results are qualitatively very similar to those obtained with correctly specified auxiliary information, and indicate that although

[4] The bias corrected version $\widehat{\theta}^c$ of the M-estimator $\widehat{\theta}$ is $\widehat{\theta}^c = \widehat{\theta} - \widehat{Bias}\left(\widehat{\theta}\right)$ where $\widehat{Bias}\left(\widehat{\theta}\right)$ is a consistent estimator of the first four lines of (25).

the misspecification has some negative finite sample effects on both GMM and WM estimators, the WM estimators are still clearly superior to both M and GMM estimators in terms of finite sample bias and efficiency. However WM estimators based on EL weights seem to be affected by the misspecification comparatively more than those based on either EU or NT. The "sensitivity" to misspecification of EL is confirmed and emphasised in the second (stronger) case of misspecification (that is for $g_2(x)^{ms}$). In this case, as expected from the discussion in Remark 4, all of the WM estimators are characterised by bigger finite sample biases, but among them those based on NT weights seems to be less sensitive to the increase in the level of misspecification. The robustness of NT is also reflected in the variances, which are now typically smaller than those based on EL weights. Finally under misspecification the bias corrections are not as effective as in the case of correctly specified auxiliary information, but they are still useful to reduce the bias of the WM-estimators. As for the case of correctly specified information these results are robust to the sample size; in the case of $n = 50$ EL seems to be a little more sensitive to misspecification.

In sum the results of the simulations can be summarised as follows: if the auxiliary information is correctly specified (or the p-vales of a test statistic used to assess its correctness are above 0.20-0.25) WM-estimators (with or without bias correction) based on EL weights are characterised by the best finite sample performances both in terms of bias and efficiency. On the other hand if there are some doubts about the "correctness" of the auxiliary information (as suggested, for example, by p-values between 0.10-0.25), then WM estimators with NT weights have the best finite sample performance.

5 Conclusions

In this paper we have introduced a new class of weighted M-estimators where the weights are obtained from GEL estimation of some auxiliary information about the otherwise unknown distribution of the data. These estimators are efficient in the sense of having a smaller variance than that of standard M-estimators, and also in the sense of having the same asymptotic variance as that of efficient GMM estimators with the same auxiliary information. Compared to the latter however, the estimators of this paper are much simpler to compute. Furthermore in the case of smooth estimating equations the proposed estimators are characterised by a small second order bias property compared to efficient GMM estimators.

The finite sample behaviour of the weighted M-estimators based on the three most used GEL members (empirical likelihood, Euclidean likelihood and nonparametric tilting) has been investigated by means of simulations. The results of the latter suggest that when the auxiliary information is correctly specified the proposed estimators are typically less biased and can be notably more precise than those based on standard M and efficient GMM estimation, with those based on empirical likelihood being the least biased and more precise. On the other hand when there are some doubts about the accuracy of the auxiliary information weighted M-estimators based on nonparametric tilting seem to be preferrable.

References

Andrews, D. (1994), Empirical processes methods in econometrics, *in* R. Engle and D. Mcfadden, eds, 'Handbook of Econometrics', Vol. 4, Amsterdam: North-Holland.

Baggerly, K. A. (1998), 'Empirical likelihood as a goodness of fit measure', *Biometrika* **85**, 535–547.

Brown, B. M. and Chen, S. X. (1998), 'Combined and least squares empirical likelihood', *Annals of the Institute of Statistical Mathematics* **50**, 697–714.

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

Efron, B. (1981), 'Nonparametric standard errors and confidence intervals (with discussion)', *Canadian Journal of Statistics,* **9**, 139–172.

Hall, P. and Sheather, S. (1988), 'On the distribution of a studentized quantile', *Journal of the Royal Statistcal Society B* **50**, 381–391.

Handcock, M., Houvilainen, S. and Rendall, M. (2000), 'Combining registration system and survey data to estimate birth probabilities', *Demography* **37**, 187–192.

Hellerstein, J. and Imbens, G. (1999), 'Imposing moment restrictions from auxiliary data by weighting', *Review of Economics and Statistics* **81**, 1–14.

Huber, P. (1964), 'Robust estimation of a location parameter', *Annals of Mathematical Statistics* **18**, 121–140.

Huber, P. (1973), 'Robust regression: Asymptotics, conjectures and Monte Carlo', *Annals of Statistics* **1**, 799–821.

Imbens, G. and Lancaster, T. (1994), 'Combining micro and macro data in microeconometric models', *Review of Economic Studies* **61**, 655–680.

Imbens, G. W., Spady, R. H. and Johnson, P. (1998), 'Information theoretic approaches to inference in moment condition models', *Econometrica* **66**, 333–357.

Koenker, R. and Basset, G. (1978), 'Regression quantiles', *Econometrica* **46**, 33–50.

McCullagh, P. (1987), *Tensor Methods in Statistics*, London: Chapman and Hall.

Newey, W. K. and Smith, R. J. (2004), 'Higher order properties of GMM and generalized empirical likelihood estimators', *Econometrica* **72**, 219–256.

Owen, A. (1991), 'Empirical likelihood for linear models', *Annals of Statistics* **19**, 1725–1747.

Owen, A. (2001), *Empirical Likelihood*, Chapman and Hall.

Parente, P. and Smith, R. (2005), 'GEL methods for non-smooth moment indicators'. Working paper, University of Warwick.

Pfanzagl, J. and Wefelmeier, W. (1978), 'A third-order optimum property of the maximum likelihood estimator', *Journal of Multivariate Analysis* **8**, 1–29.

Powell, J. (1986), 'Symmetrically trimmed least squares estimators for Tobit models', *Econometrica* **54**, 1435–1460.

Smith, R. J. (1997), 'Alternative semi-parametric likelihood approaches to generalised method of moments estimation', *Economic Journal* **107**, 503–519.

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

Zhang, B. (1995), 'M-estimation and quantile estimation in the presence of auxiliary information', *Journal of Statistical Planning and Inference* **44**, 77–94.

Appendix

We use the following abbreviations and conventions: let $g(x_i) = g_i$, $M_n = \max_i \|g_i\|$, $\psi(x_i, \theta) = \psi_i(\theta)$, $\lim = \lim_{n\to\infty}$ and $\sum_{i=1}^n = \sum$; also CLT, CMT, LIL and (U)S(W)LLN stand for central limit theorem, continuous mapping theorem, law of iterated logarithm and (uniform) strong (weak) law of large numbers, respectively.

Proof of Theorem 1. By the first Borel-Cantelli lemma $M_n = o_{a.s.}\left(n^{1/\beta}\right)$ so that on $\Lambda_n := \left\{\lambda : \|\lambda\| \le n^{-\beta}\right\}$, $\lambda^\tau g_i = o_{a.s.}(1)$ and therefore $\Lambda_n \subseteq V_n$ $a.s.$ Since $G_n(\lambda)$ is strictly concave on Λ_n it follows that there exists $(a.s.)$ a unique $\widetilde{\lambda} := \arg\max_{\lambda\in\Lambda_n} G_n(\lambda)$. A Taylor expansion about 0 gives

$$\mathcal{G}_n(0) \le \mathcal{G}_n\left(\widetilde{\lambda}\right) = \sum\left[-\widetilde{\lambda}^\tau g_i + \rho_2\left(\lambda_*^\tau g_i\right)\widetilde{\lambda}^\tau g_i g_i^\tau \widetilde{\lambda}/2\right]/n \le -\left\|\widetilde{\lambda}\right\|\|\overline{g}\| - \sigma_s\left\|\widetilde{\lambda}\right\|^2$$

where $\|\lambda_*\| \le \left\|\widetilde{\lambda}\right\|$, $\overline{g} = \sum g_i/n$ and $\sigma_s > 0$ is the smallest eigenvalue of Σ. Subtracting $G_n(0) - \sigma_s\left\|\widetilde{\lambda}\right\|^2$, dividing by $\left\|\widetilde{\lambda}\right\|$ and finally using LIL one gets $\left\|\widetilde{\lambda}\right\| \le \|\overline{g}\| = O_{a.s.}\left(n^{-1/2}(\log\log n)^{1/2}\right)$. Since $\left\|\widetilde{\lambda}\right\| = o_{a.s.}\left(n^{-\beta}\right)$, $\widetilde{\lambda} \in int\{\Lambda_n\}$ $a.s.$ hence the first order condition for an interior maximum $\partial G_n\left(\widetilde{\lambda}\right)/\partial\lambda = 0$ is satisfied $a.s.$ Clearly $\widetilde{\lambda} \in V_n$ so by concavity of $G_n(\lambda)$ and convexity of V_n it follows that $G_n\left(\widetilde{\lambda}\right) = \sup_{\lambda\in V_n} G_n(\lambda)$ which implies the existence of a unique $\widehat{\lambda} := \arg\max_{\lambda\in V_n} G_n(\lambda)$. Next by Taylor expansion $\rho_1\left(\widehat{\lambda}^\tau g_i\right) = -1 + \rho_2\left(\lambda_*^\tau g_i\right)\widehat{\lambda}^\tau g_i$ where $\|\lambda_*\| \le \left\|\widehat{\lambda}\right\|$. Since $\left|\widehat{\lambda}^\tau M_n\right| = o_{a.s.}(1)$ $\max_i\left|\rho_2\left(\lambda_*^\tau g_i\right) + 1\right| = o_{a.s.}(1)$ uniformly in i so $\rho_1\left(\widehat{\lambda}^\tau g_i\right) = -1 - \widehat{\lambda}^\tau g_i + o_{a.s.}(1)$. Similarly

$$1/\sum\rho_1\left(\widehat{\lambda}^\tau g_i\right) = -1/n\left(1 + \sum\rho_2\left(\lambda_*^\tau g_i\right)g_i^\tau/n\right)\widehat{\lambda} = -1/n\left(1 + O_{a.s.}\left(n^{-1}\log\log n\right)\right),$$

by LIL and thus

$$\max_i\left|\widehat{w}_i - 1/n\left(1 + \widehat{\lambda}^\tau g_i\right)\right| = O_{a.s.}\left(n^{-1}\log\log n\right). \tag{26}$$

By construction $\sum\widehat{w}_i g_i = 0$ $a.s$, so by (26) $0 = \overline{g} + \sum g_i g_i^\tau\widehat{\lambda}/n + O_{a.s.}\left(n^{-1}\log\log n\right)$. By SLLN $\sum g_i g_i^\tau/n \overset{a.s}{\to} \Sigma$ and hence by CMT $n^{1/2}\widehat{\lambda} = -\Sigma^{-1}\sum g_i/n^{1/2} + O_{a.s.}\left(n^{-1/2}\log\log n\right)$. Applying CLT and CMT to the latter gives (3). Again by (26)

$$n^{1/2}\left(\widehat{F}_w(x) - F(x)\right) = n^{1/2}\left(\widehat{F}_n(x) - F(x)\right) - n^{1/2}\sum I\{x_i \le x\}\widehat{\lambda}^\tau g_i/n +$$
$$O_{a.s.}\left(n^{-3/2}\log\log n\right) = n^{1/2}\left(\widehat{F}_n(x) - F(x)\right) - n^{1/2}E\left[\widehat{\lambda}^\tau g_i I\{x_i \le x\}\right] + o_{a.s.}(1),$$

from which (4) follows by CLT, and CMT. ■

Proof of Theorem 2. Note that by (26)

$$\left\|\Psi_w(\theta)\right\| \le \left\|\Psi_n(\theta)\right\|\left(1+o_{a.s}(1)\right)$$

uniformly in Θ. By this, the definition of the estimator and standard arguments

$$\begin{aligned}
&\left\|\Psi\left(\widehat{\theta}_w\right)\right\| \le \left\|\Psi\left(\widehat{\theta}_w\right)-\Psi_w\left(\widehat{\theta}_w\right)\right\| + \left\|\Psi_w\left(\widehat{\theta}_w\right)\right\| \le \\
&\sup_{\theta\in\Theta}\left\|\Psi_n(\theta)-\Psi(\theta)\right\| + o_{a.s}(1)\left\|\Psi_n\left(\widehat{\theta}_w\right)\right\| + o_{a.s}(1) \le \\
&o_{a.s}(1)+o_{a.s}(1)\left\|\Psi(\theta_0)\right\| = o_{a.s}(1).
\end{aligned}$$

By (II) it then follows that $\widehat{\theta}_w \in \|\theta-\theta_0\| < \zeta$ $a.s.$ and since ζ is arbitrary $\widehat{\theta}_w \overset{a.s.}{\to} \theta_0$. ■

Proof of Theorem 3. Let $\Psi_w(\theta\pm\varepsilon) = \sum \widehat{w}_i\psi_i(\theta\pm\varepsilon)$ for some $\varepsilon > 0$. By (26) and SLLN we have that $\Psi_w(\theta\pm\varepsilon) \overset{a.s.}{\to} \Psi(\theta\pm\varepsilon)$. Then monotonicity of $\psi_i(\theta)$ implies monotonicity of $\Psi(\theta)$ and since θ_0 is the unique root of $\Psi(\theta_0\pm\varepsilon)$, $\Psi(\theta_0-\varepsilon) < 0 < \Psi(\theta_0+\varepsilon)$ for ε sufficiently small. It then follows that

$$\Psi_w(\theta_0-\varepsilon) < 0 < \Psi_w(\theta_0+\varepsilon) \quad a.s.$$

whence there exists a $\widehat{\theta}_w$ such that $\Psi_w\left(\widehat{\theta}_w\right) \overset{a.s.}{=} 0$ and $\widehat{\theta}_w \overset{a.s}{\to} \theta_0$ by the continuity of $\Psi_w(\theta)$. ■

Proof of Theorem 4. Let $d\widehat{G}_n(x) := d\widehat{F}_n(x) - dF(x)$. Note that

$$\begin{aligned}
&\Psi_w(\theta) = \Gamma(\theta-\theta_0) + o\left\|\theta-\theta_0\right\| + \Psi_n(\theta)\left(1+\widehat{\lambda}^\tau g_i\right) + o_{a.s.}(1) = \\
&H_w(\theta) + o\left(\left\|\theta-\theta_0\right\|\right) + \int\left(\psi(\theta)-\psi(\theta_0)\right)d\widehat{G}_n(x) + \\
&\widehat{\lambda}^\tau\sum\left[\psi(\theta)-\psi(\theta_0)\right]g_i/n + o_{a.s.}(1),
\end{aligned}$$

where

$$H_w(\theta) = \Gamma(\theta-\theta_0) + \Psi_n(\theta_0)\left(1+\widehat{\lambda}^\tau g_i\right).$$

Let $n^{1/2}\left(\overline{\theta}-\theta_0\right) = O_p(1)$; then

$$\begin{aligned}
&\left\|n^{1/2}\left(\Psi_w\left(\overline{\theta}\right)-H_w\left(\overline{\theta}\right)\right)\right\| \le o_p(1) + \\
&\sup_{\|\theta-\theta_0\|\le\delta_n}\left\|n^{1/2}\left[\int\left(\psi_i(\theta)-\psi_i(\theta_0)\right)d\widehat{G}_n(x)\right]\right\| + \\
&\left\|n^{1/2}\widehat{\lambda}\right\|\left\|\sum\left[\psi_i\left(\overline{\theta}\right)-\psi_i(\theta_0)\right]g_i/n\right\| = A_1 + A_2.
\end{aligned} \tag{27}$$

By (II) $A_1 = o_p(1)$ while by (III) (a) and the consistency of $\overline{\theta}$ there exists a $\delta_n \to 0$ such that $\sup_{\|\theta-\theta_0\|\le\delta_n} \|(\psi_i(\theta) - \psi_i(\theta_0)) g_i\| = o_p(1)$. Then by (III) (b) and dominated convergence $E\sup_{\|\theta-\theta_0\|\le\delta_n} \|(\psi(\theta) - \psi(\theta_0)) g_i\| \to 0$ so that by triangle and Markov inequalities

$$\left\|\sum \left[\psi_i\left(\overline{\theta}\right) - \psi_i(\theta_0)\right] g_i/n\right\| \le \sum \sup_{\|\theta-\theta_0\|\le\delta_n} \|(\psi_i(\theta) - \psi_i(\theta_0)) g_i\|/n = o_p(1),$$

and $A_2 = O_p(1) o_p(1) = o_p(1)$. Thus $n^{1/2}\Psi_w\left(\overline{\theta}\right)$ is asymptotically equivalent to $n^{1/2}H_w\left(\overline{\theta}\right)$. Let $\widetilde{\theta} := \arg\min_\theta \|H_w(\theta)\|$ and note that

$$\left\|n^{1/2}H_w\left(\overline{\theta}\right)\right\| = \left\|n^{1/2}H_w\left(\widetilde{\theta}\right)\right\| + o_p(1),$$

which implies that $\left\|n^{1/2}\Gamma\left(\overline{\theta} - \widetilde{\theta}\right)\right\| \le \|\Gamma\| \left\|n^{1/2}\left(\overline{\theta} - \widetilde{\theta}\right)\right\| = Co_p(1)$ and hence $\left\|n^{1/2}\left(\overline{\theta} - \widetilde{\theta}\right)\right\| = o_p(1)$. Thus the distribution of $\overline{\theta}$ is asymptotically equivalent to that of $\widetilde{\theta}$. Since $\widehat{\theta}_w$ is $n^{1/2}$-consistent by assumption, the conclusion follows by CLT and CMT. ■

Proof of Theorem 5. Assume that $\psi(\theta)$ is nonincreasing in θ, let $y_n = \theta_0 + y\sigma_{\Psi' g}/n^{1/2}$where $y \in \Re$, and $\Psi_w(y_n)$ denote the corresponding weighted estimating equation. Then by (26), (3) and SLLN

$$\begin{aligned}\Psi_w(y_n) &= \sum \psi_i(y_n)\left(1 + \widehat{\lambda}^\tau g_i\right)/n + o_{a.s.}(1) \\ &= \sum \left[\psi_i(y_n) - E\left[\psi(y_n) g^\tau\right]\Sigma^{-1} g_i\right]/n + o_{a.s.}(1) = \sum z_{in}/n + o_{a.s.}(1).\end{aligned}$$

As in Huber (1964) it suffices to show that $\lim \Pr\{\Psi_w(y_n) \le 0\} = \lim \Pr\{\sum z_{in}/n \le 0\} = F(y)$ for every y, where $F(\cdot)$ is the standard normal distribution. Let $Z_{in} := (z_{in} - Ez_{1n})/\sigma(z_{1n})$ where $\sigma^2(z_{1n}) = Var(z_{1n})$ and note that $\lim n^{1/2}E[z_{1n}/\sigma(z_{1n})] = -y$ (Huber, 1964, p. 78). Therefore $\lim \Pr\{\sum z_{in}/n \le 0\} = \lim \Pr\left\{n^{-1/2}\sum Z_{in} \le y\right\}$. Since the Lindeberg condition $\lim \int_{z_n > n^{1/2}\varepsilon} z_n^2 dF(z) = 0$ holds for

$$z_n := \left|\psi_i(y_n) + E|\psi(y_n)| \|g\| \left\|\Sigma^{-1}\right\| \|g\|\right|,$$

it follows by a CLT for triangular arrays that $\lim \Pr\left\{n^{-1/2}\sum Z_{in} \le y\right\} = F(y)$. ■

Proof of Theorem 6. By (26) and mean value expansion

$$0 = \sum \widehat{w}_i\psi_i\left(\widehat{\theta}_w\right) = \sum \psi_i(\theta_0)/n + \sum \psi_i'(\theta^*)\left(\widehat{\theta}_w - \theta_0\right)/n + \widehat{\lambda}^\tau \sum \psi_i(\theta_0) g_i/n +$$
$$\widehat{\lambda}^\tau \sum \psi_i'(\theta^*) g_i\left(\widehat{\theta}_w - \theta_0\right)/n + O_{a.s.}\left(n^{-1}(\log\log n)\right), \qquad (28)$$

where $\|\theta^* - \theta_0\| \le \left\|\widehat{\theta}_w - \theta_0\right\|$ from which

$$\widehat{\theta}_w - \theta_0 = \left[\sum \psi_i'(\theta^*)/n + \sum \psi_i'(\theta^*)\widehat{\lambda}^\tau g_i/n\right]^{-1}$$
$$\left(\sum \psi_i(\theta_0)/n + \sum \psi_i(\theta_0)\widehat{\lambda}^\tau g_i/n + O_{a.s.}\left(n^{-1}(\log\log n)\right)\right).$$

Since $\theta^* \xrightarrow{a.s} \theta_0$, using (3), USLLN, LIL and CMT one gets

$$\left\|n^{1/2}\widehat{\lambda}^\tau \sum \psi_i(\theta_0) g_i/n + E\left[\psi(\theta_0) g^\tau\right]\Sigma^{-1}\overline{g}\right\| = o_{a.s}(1),$$
$$\left\|\widehat{\lambda}^\tau \sum \psi_i'(\theta^*) g_i/n\right\| \le \left\|\widehat{\lambda}\right\| \sup_{\theta\in N_0}\left\|\sum \psi_i'(\theta) g_i/n - E\psi'(\theta) g\right\| +$$
$$\left\|\widehat{\lambda}\right\|\|E\psi'(\theta) g\| = O_{a.s.}\left(n^{-1/2}(\log\log n)^{1/2}\right),$$
$$\left\|\sum \psi_i'(\theta^*)/n - E\left[\psi'(\theta_0)\right]\right\| = o_{a.s}(1).$$

The conclusion follows by CLT and CMT. ■

Proof of Theorem 7. By consistency of $\widehat{\theta}_w$ and WLLN $\widehat{\Sigma}_{\widehat{w}} \xrightarrow{p} \Sigma$,

$$\left\|\sum \widehat{w}_i\psi_i\left(\widehat{\theta}_w\right)\psi_i\left(\widehat{\theta}_w\right)^\tau - V\right\| \le \sup_{\theta\in N_0}\left\|\sum\left(\psi_i(\theta)\psi_i(\theta)^\tau/n - E\left[\psi(\theta)\psi(\theta)^\tau\right]\right)\right\| + o_p(1)$$

so that $\widehat{V}_{\widehat{w}} \xrightarrow{p} V$. Note that by the stochastic equicontinuity (I) and (26) for $l = 1, ..., k$

$$\left\|\Psi_{\widehat{w}_i}\left(\widehat{\theta}_w + b_n e_l\right) - \Psi_{\widehat{w}_i}\left(\widehat{\theta}_w\right) - \left(\Psi\left(\widehat{\theta}_w + b_n e_l\right) - \Psi\left(\widehat{\theta}_w\right)\right)\right\| = o_p\left(n^{-1/2}b_n^{-1}\right),$$

whereas by the local differentiability (II) and triangle inequality

$$\left\|\Psi\left(\widehat{\theta}_w + b_n e_l\right)/b_n - \Gamma e_l\right\| \le \left\|\Gamma\left(\widehat{\theta}_w - \theta_0\right)/b_n\right\| + o_p\left(n^{-1/2}b_n^{-1}\right),$$

so that again by triangle inequality $\left(\widehat{\Gamma}_{\widehat{w}}\right)_l \xrightarrow{p} (\Gamma)_l$ $l = 1, ..., k$. Finally by the consistency of $\widehat{\theta}_w$, ULLN and the triangle inequality

$$\left\|\sum \widehat{w}_i\left(\psi_i\left(\widehat{\theta}_w\right) - \psi_i(\theta_0)\right)g_i^\tau + \sum \widehat{w}_i\psi_i(\theta_0) g_i^\tau - E\left[\psi_i(\theta_0) g_i^\tau\right]\right\|$$
$$\le \sum\left\|\left(\psi_i\left(\widehat{\theta}_w\right) - \psi(x_i,\theta_0)\right)g_i^\tau\right\|/n + o_p(1) \le$$
$$\sum \sup_{\|\theta-\theta_0\|\le\delta_n}\|(\psi_i(\theta) - \psi_i(\theta_0)) g_i^\tau\|/n + o_p(1) = o_p(1)$$

so that the conclusion follows by CMT. ■

Proof of Theorem 8. Recall that the one-step weighted M-estimator for θ_0 is

$$\widehat{\theta}_w^1 = \widehat{\theta}_w - \left[\sum \widehat{w}_i\psi_i'\left(\widehat{\theta}_w\right)\right]^{-1}\sum \widehat{w}_i\psi_i\left(\widehat{\theta}_w\right). \quad (29)$$

By (26) and a mean value expansion it can be shown that (29) can be written as $n^{1/2}\left(\widehat{\theta}^{1}_{w}-\theta_0\right)=n^{1/2}A_{4n}^{-1}\sum_{j=1}^{3}A_{jn}$, where

$$\begin{aligned} A_{1n} &= -\sum\left[\psi_i(\theta_0)+\psi_i(\theta_0)\widehat{\lambda}^{\tau}g_i+o_{a.s.}(1)\right]/n, \\ A_{2n} &= -\sum\left[\psi_i\left(\widehat{\theta}_w\right)-\psi_i(\theta_0)-\psi_i'\left(\overline{\theta}\right)\left(\widehat{\theta}_w-\theta_0\right)\right]/n, \\ A_{3n} &= -\widehat{\lambda}^{\tau}\sum\left[\psi_i\left(\widehat{\theta}_w\right)-\psi_i(\theta_0)-\psi_i'\left(\overline{\theta}\right)\left(\widehat{\theta}_w-\theta_0\right)\right]g_i/n, \\ A_{4n} &= \sum\left[\psi_i'\left(\overline{\theta}\right)\left(1+\widehat{\lambda}^{\tau}g_i+o_{a.s.}(1)\right)\right]/n. \end{aligned}$$

Let $F_{\psi g}=E\left[\psi(\theta_0)\psi(\theta_0)^{\tau}\right]-E\left[\psi(\theta_0)g^{\tau}\right]\Sigma^{-1}E\left[\psi(\theta_0)g^{\tau}\right]^{\tau}$; by CLT, CMT and LLN it follows that $n^{1/2}A_{1n}\overset{d}{\to}N(0,F_{\psi g})$, $n^{1/2}A_{2n}=E\left[\psi'(\theta_0)\right]n^{1/2}\left(\widehat{\theta}_w-\theta_0\right)+o_{a.s}(1)$,

$$\begin{aligned} &\left\|n^{1/2}A_{3n}\right\|\le\left\|\widehat{\lambda}\right\|\left(o_{a.s.}(1)+\left\|E\left[\psi(\theta_0)g\right]\right\|\left\|n^{1/2}\left(\left(\widehat{\theta}_w-\theta_0\right)\right)\right\|\right)= \\ &O_{a.s.}\left(n^{-1/2}(\log\log n)^{1/2}\right)\left(o_{a.s.}(1)+O_p(1)\right)=o_p(1), \end{aligned}$$

and $\left\|A_{4n}-E\psi'(\theta_0)\right\|=o_{a.s}(1)$, whence the results follows by CMT.

Proof of Theorem 9. The arguments of the proof of Theorem 1 apply viz. a. viz. to $g_i^{\delta}:=g_i-\delta/n^{1/2}$, so that it is easy to see that $0=\overline{g}^{\delta}+\sum g_ig_i^{\tau}\widehat{\lambda}/n+O_{a.s.}\left(n^{-1}(\log\log n)^{1/2}\right)$. Thus the first conclusion follows by CLT and CMT. As for the second conclusion note that

$$\begin{aligned} &n^{1/2}\left(\widehat{F}_w(x)-F(x)\right)=n^{1/2}\left(\widehat{F}_n(x)-F(x)\right)-n^{1/2}\sum I\{x_i\le x\}\widehat{\lambda}^{\tau}g_i^{\delta}/n+ \\ &O_{a.s.}\left(n^{-3/2}(\log\log n)^{1/2}\right)=n^{1/2}\left(\widehat{F}_n(x)-F(x)\right)-n^{1/2}E\left[\widehat{\lambda}^{\tau}g_iI\{x_i\le x\}\right]+o_{a.s.}(1), \end{aligned}$$

and the result follows again by CLT and CMT. ■

Proof of Theorem 10. Let $g_i^{\delta}:=g_i-\delta/n^{1/2}$. Note that $\max_i\left|\widehat{\lambda}^{\tau}g_i^{\delta}\right|=o_{a.s.}(1)$ so that the proofs of Theorem 2 and 3 are still valid, hence the conclusion. ■

Proof of Theorem 11. Note that

$$n^{1/2}H_w(\theta)=\Gamma n^{1/2}(\theta-\theta_0)+n^{1/2}\Psi_n(\theta_0)\left(1+\widehat{\lambda}^{\tau}g_i^{\delta}\right)$$

so that by CLT and CMT

$$n^{1/2}\left(\widehat{\theta}-\theta_0\right)\overset{d}{\to}N\left(\Gamma^{-1}E\left[\psi(x,\theta_0)g(x)^{\tau}\right]\Sigma^{-1}\delta,\Sigma_{\Gamma g}\right)$$

where $\Sigma_{\Gamma g}$ is as defined in Theorem 4. Furthermore similarly to (27)

$$\begin{aligned} &\left\|n^{1/2}\widehat{\lambda}\right\|\left\|\sum\left[\psi_i\left(\overline{\theta}\right)-\psi_i(\theta_0)\right]g_i^{\delta}/n\right\|\le\left\|n^{1/2}\widehat{\lambda}\right\|\left\|\sum\left[\psi_i\left(\overline{\theta}\right)-\psi_i(\theta_0)\right]g_i/n\right\|\times \\ &\left(1+\|\delta\|/n^{1/2}\right)=o_p(1). \end{aligned}$$

Thus the first conclusion follows as in the proof of Theorem 4. The second conclusion follows as in Theorem 5 using

$$\sum z_{in}/n + E\left[\psi\left(\theta_0\right) g^\tau\right] \Sigma^{-1}\delta + o_p\left(1\right).$$

The third and last conclusion follows using an expansion analogous to that in (28), namely

$$0 = \sum \widehat{w}_i \psi_i\left(\widehat{\theta}_w\right) = \sum \psi_i\left(\theta_0\right)/n + \sum \psi_i'\left(\theta^*\right)\left(\widehat{\theta}_w - \theta_0\right)/n + \widehat{\lambda}^\tau \sum \psi_i\left(\theta_0\right) g_i^\delta/n +$$
$$\widehat{\lambda}^\tau \sum \psi_i'\left(\theta^*\right) g_i^\delta \left(\widehat{\theta}_w - \theta_0\right)/n + O_{a.s.}\left(n^{-1}\left(\log\log n\right)\right) =$$
$$\sum \psi_i\left(\theta_0\right)/n + \sum \psi_i'\left(\theta^*\right)\left(\widehat{\theta}_w - \theta_0\right)/n + \left[\left(\sum g_i g_i^\tau/n\right)^{-1} \overline{g}^\delta\right]^\tau \sum \psi_i\left(\theta_0\right) g_i/n +$$
$$\widehat{\lambda}^\tau \sum \psi_i'\left(\theta^*\right) g_i \left(\widehat{\theta}_w - \theta_0\right)/n + O_{a.s.}\left(n^{-1}\left(\log\log n\right)^{1/2}\right),$$

and the rest of the proof is identical to that of Theorem 6. ■ ■

Proof of Theorem 12. We use tensor notation and indicate arrays by their elements as for example in McCullagh (1987). Thus, for any index say j, a_j is a vector, a_{jk} is a matrix, etc. We also follow the summation convention, that is for any two repeated indices, their sum is understood. For $1 \leq a, b, c, ... \leq q$ and $1 \leq \alpha, \beta, ... \leq k$ let

$$\begin{aligned}
A^{abc...} &= \sum\left(g_i^a g_i^b g_i^c ... - \alpha^{abc...}\right)/n, \quad \alpha^{abc...} = E\left[g_i^a g_i^b g_i^c .\right], \\
B^{\alpha\alpha_1...\alpha_k} &= \sum\left(\partial^k \psi_i^\alpha\left(\theta_0\right)/\partial\theta^{\alpha_1}...\partial\theta^{\alpha_k} - \beta^{\alpha\alpha_1....\alpha_k}\right)/n, \\
\beta^{\alpha\alpha_1....\alpha_k} &= E\left[\partial^k \psi_i^\alpha\left(\theta_0\right)/\partial\theta^{\alpha_1}...\partial\theta^{\alpha_k}\right], \\
C^{\alpha\alpha_1...\alpha_k abc...} &= \sum\left[\left(\partial^k \psi_i^\alpha\left(\theta_0\right)/\partial\theta^{\alpha_1}...\partial\theta^{\alpha_k}\right) g_i^a g_i^b ... - \gamma^{\alpha\alpha_1....\alpha_k abc...}\right]/n, \\
\gamma^{\alpha\alpha_1....\alpha_k abc...} &= E\left[\left(\partial^k \psi_i^\alpha\left(\theta_0\right)/\partial\theta^{\alpha_1}...\partial\theta^{\alpha_k}\right) g_i^a g_i^b ...\right],
\end{aligned}$$

that is $A^{abc...}$, $B^{\alpha\alpha_1...\alpha_k}$ and $C^{\alpha\alpha_1...\alpha_k abc...}$ represent $O_p\left(n^{-1/2}\right)$ random arrays of, respectively. higher order moments of the standardised auxiliary information, higher order derivatives of the estimating functions, and of covariances between higher order derivatives of the estimating functions and the higher order arrays of moments of the standardised auxiliary information.

First we obtain a third-order stochastic expansion for $\widehat{\lambda}^a$ that solves $0 = \sum \widehat{w}_i g_i^a$. Recall that $\widehat{w}_i = \rho_1\left(\widehat{\lambda}^a g_i^a\right)/\sum \rho_1\left(\widehat{\lambda}^a g_i^a\right)$ thus using a third order Taylor expansion of the numerator and of the denominator and some algebra we obtain

$$0 = \sum\left[1 + \widehat{\lambda}^b g_i^b - \rho_3\left(\widehat{\lambda}^b g_i^b\right)^2/2 - \rho_3\left(\widehat{\lambda}^b g_i^b\right)^3/3!\right] g_i^a/n + O_p\left(n^{-2}\right),$$

which can then be inverted to give

$$\widehat{\lambda}^a = -A^a + A^{ab}A^b + \rho_3\alpha^{abc}A^bA^c/2 - A^{ab}A^{bc}A^c - \rho_3A^{abc}A^bA^c/2 + \rho_3\alpha^{abc}A^bA^{cd}A^d - \rho_4\alpha^{abcd}A^bA^cA^d/3! - \rho_3^2\alpha^{abc}\alpha^{cef}A^bA^eA^f/2 + O_p\left(n^{-2}\right). \quad (30)$$

Next using (30) we obtain that $\widehat{w}_i$ has the following stochastic expansion

$$\widehat{w}_i = 1 - g_i^aA^a + g_i^aA^{ab}A^b + \rho_3g_i^a\alpha^{abc}A^bA^c/2 - g_i^aA^{ab}A^{bc}A^c - \rho_3g_i^aA^{abc}A^bA^c/2 + \rho_3g_i^a\alpha^{abc}A^bA^{cd}A^d - \rho_4\alpha^{abcd}g_i^aA^bA^cA^d/3! - \rho_3^2g_i^a\alpha^{abc}\alpha^{cef}A^bA^eA^f/2 - \rho_3g_i^ag_i^bA^aA^b/2 + \rho_3g_i^ag_i^bA^a\left(A^{bc}A^c + \rho_3\alpha^{bcd}A^cA^d/2\right) - \rho_3g_i^ag_i^bA^{ac}A^{bd}A^cA^d/2 - \rho_3^2g_i^ag_i^bA^{ac}\alpha^{bde}A^cA^dA^e/2 - \rho_3^3g_i^ag_i^b\alpha^{acd}\alpha^{bef}A^cA^dA^eA^f/8 + \rho_4g_i^ag_i^bg_i^cA^aA^bA^c/3! + O_p\left(n^{-2}\right). \quad (31)$$

Finally we obtain a third order stochastic expansion for $\widehat{\theta}$ that solves $0 = \sum \widehat{w}_i\psi_i^\alpha\left(\widehat{\theta}\right)$. By a third order Taylor expansion about θ_0

$$0 = \sum \widehat{w}_i\Big[\psi_i^\alpha + \partial\psi_i^\alpha\left(\theta_0\right)/\partial\theta^\beta\left(\widehat{\theta} - \theta_0\right)^\beta + \partial^2\psi_i^\alpha\left(\theta_0\right)/\partial\theta^\beta\partial\theta^\gamma\left(\widehat{\theta} - \theta_0\right)^\beta\left(\widehat{\theta} - \theta_0\right)^\gamma/2 + \partial^3\psi_i^\alpha\left(\theta_0\right)/\partial\theta^\beta\partial\theta^\gamma\partial\theta^\delta\left(\widehat{\theta} - \theta_0\right)^\beta\left(\widehat{\theta} - \theta_0\right)^\gamma\left(\widehat{\theta} - \theta_0\right)^\delta/3!\Big] + O_p\left(n^{-2}\right),$$

where for notational simplicity $\psi_i^\alpha(\theta_0) = \psi_i^\alpha$. By (30) and (31) we get

$$0 = \psi^\alpha + \sum \psi_i^\alpha g_i^b \left(-A^b + A^{bc}A^c + \rho_3\alpha^{bcd}A^cA^d/2 - A^{bc}A^{cd}A^d - \rho_3A^{bcd}A^cA^d/2 + \right.$$
$$\rho_3A^{bcd}A^cA^d/2 + \rho_3\alpha^{bcd}A^cA^{de}A^e - \rho_4\alpha^{bcde}A^cA^dA^e/3! - \left.\rho_3^2\alpha^{bcd}\alpha^{def}A^cA^eA^f/2\right)/n -$$
$$\rho_3\sum \psi_i^\alpha g_i^b g_i^c \left[-A^b + \left(A^{bd}A^d + \rho_3\alpha^{bde}A^dA^e/2\right)\right]\left[-A^c + \left(A^{ce}A^e + \rho_3\alpha^{cef}A^eA^f/2\right)\right]/n +$$
$$\rho_4\sum \psi_i^\alpha g_i^b g_i^c A^bA^cA^d/(3!n) + B^{\alpha\beta}\left(\widehat{\theta} - \theta_0\right)^\beta + \sum B_i^{\alpha\beta}\left(\widehat{\theta} - \theta_0\right)^\beta g_i^a \left(-A^a + A^{ab}A^b + \right.$$
$$\rho_3\alpha^{abc}A^bA^c/2 - A^{ab}A^{bc}A^c - \rho_3A^{ab}\alpha^{bcd}A^cA^d/2 + \rho_3\alpha^{abc}A^bA^{cd}A^d -$$
$$\left.\rho_3^2\alpha^{abc}\alpha^{cde}A^bA^dA^e/2 - \rho_4\alpha^{abcd}A^bA^cA^d/3!\right)/n + \rho_3\sum B_i^{\alpha\beta}\left(\widehat{\theta} - \theta_0\right)^\beta g_i^a g_i^b A^aA^b/(2n) +$$
$$\beta^{\alpha\beta}\left(\widehat{\theta} - \theta_0\right)^\beta + \sum \beta^{\alpha\beta}\left(\widehat{\theta} - \theta_0\right)^\beta g_i^a \left(-A^a + A^{ab}A^b + \rho_3\alpha^{abc}A^bA^c/2 - \right.$$
$$-A^{ab}A^{bc}A^c - \rho_3A^{ab}\alpha^{bcd}A^cA^d/2 + \rho_3\alpha^{abc}A^bA^{cd}A^d - \rho_3^2\alpha^{abc}\alpha^{cde}A^bA^dA^e/2 -$$
$$\left.\rho_4\alpha^{abcd}A^bA^cA^d/3!\right)/n - \rho_3\sum \beta^{\alpha\beta}\left(\widehat{\theta} - \theta_0\right)^\beta g_i^a g_i^b \left(-A^a + A^{ac}A^c + \rho_3\alpha^{acd}A^cA^d/2\right)$$
$$\left(-A^b + A^{bd}A^d + \rho_3\alpha^{bde}A^dA^e/2\right)/(2n) + \rho_4\sum \beta^{\alpha\beta}\left(\widehat{\theta} - \theta_0\right)^\beta g_i^a g_i^b g_i^c A^aA^bA^c/(3!n) +$$
$$B^{\alpha\beta\gamma}\left(\widehat{\theta} - \theta_0\right)^\beta\left(\widehat{\theta} - \theta_0\right)^\gamma/2 + \beta^{\alpha\beta\gamma}\left(\widehat{\theta} - \theta_0\right)^\beta\left(\widehat{\theta} - \theta_0\right)^\gamma/2 -$$
$$\sum \beta^{\alpha\beta\gamma}\left(\widehat{\theta} - \theta_0\right)^\beta\left(\widehat{\theta} - \theta_0\right)^\gamma g_i^a A^a/(2n) + \beta^{\alpha\beta\gamma\delta}\left(\widehat{\theta} - \theta_0\right)^\beta\left(\widehat{\theta} - \theta_0\right)^\gamma \times$$
$$\left(\widehat{\theta} - \theta_0\right)^\delta/3! + O_p\left(n^{-2}\right),$$

where $\psi^\alpha = \sum_i \psi_i^\alpha/n$ and similarly for $B^{\alpha\beta\cdots}$. Inverting this expansion we get

$$\left(\widehat{\theta} - \theta_0\right)^\alpha = -\beta_{\alpha\beta}\psi^\beta + \beta_{\alpha\beta}\gamma^{\beta a}A^a + Q_2 + Q_3 + O_p\left(n^{-2}\right),$$

where $\beta_{\alpha\beta}$ is the matrix inverse of $\beta^{\alpha\beta}$,

$$Q_2 = \beta_{\alpha\beta}\left[C^{\beta a}A^a - \gamma^{\beta a}A^{ab}A^b - \rho_3\gamma^{\beta a}\alpha^{abc}A^bA^c/2 + \rho_3\gamma^{\beta ab}A^aA^b/2 + \right. \quad (32)$$
$$B^{\beta\gamma}\beta_{\gamma\delta}\psi^\delta - B^{\beta\gamma}\beta_{\gamma\delta}\gamma^{\delta a}A^a - \gamma^{\beta\gamma a}\beta_{\gamma\delta}\psi^\delta A^a + \gamma^{\beta\gamma a}\beta_{\gamma\delta}\gamma^{\delta b}A^aA^b -$$
$$\left.\beta^{\beta\gamma\delta}\left(\beta_{\gamma\varepsilon}\beta_{\delta\zeta}\psi^\varepsilon\psi^\zeta - 2\beta_{\gamma\varepsilon}\beta_{\delta\zeta}\psi^\varepsilon\gamma^{\zeta a}A^a + \beta_{\gamma\varepsilon}\beta_{\delta\zeta}\gamma^{\varepsilon a}\gamma^{\zeta b}A^aA^b\right)/2\right],$$

and

$$Q_3 = \beta_{\alpha\beta}\left[-B^{\beta\gamma}\beta_{\gamma\delta}\left(C^{\delta a}A^a + \gamma^{\delta a}A^{ab}A^b + \rho_3\gamma^{\delta a}\alpha^{abc}A^bA^c/2 - \rho_3\gamma^{\delta ab}A^aA^b/2\right) + \right. \quad (33)$$
$$B^{\beta\gamma}\beta_{\gamma\delta}\left(-B^{\delta\varepsilon}\beta_{\varepsilon\zeta}\psi^\zeta + B^{\delta\varepsilon}\beta_{\varepsilon\zeta}\gamma^{\zeta a}A^a + \gamma^{\delta\varepsilon a}\beta_{\delta\varepsilon}\psi^\delta A^a - \gamma^{\delta\varepsilon a}\beta_{\varepsilon\zeta}\gamma^{\zeta b}A^aA^b\right) +$$
$$B^{\beta\gamma}\beta_{\gamma\delta}\beta^{\delta\varepsilon\zeta}\left(\beta_{\varepsilon\eta}\beta_{\zeta\vartheta}\psi^\eta\psi^\theta/2 + \beta_{\varepsilon\eta}\beta_{\zeta\kappa}\gamma^{\eta a}\gamma^{\kappa b}A^aA^b/2 - \beta_{\varepsilon\eta}\beta_{\zeta\vartheta}\psi^\eta\gamma^{\eta a}A^a\right) +$$
$$-C^{\beta a}\left(A^{ab}A^b - \rho_3\alpha^{abc}A^bA^c/2\right) + \gamma^{\beta a}A^{ab}\left(A^{bc}A^c + \rho_3\alpha^{bcd}A^cA^d/2\right) -$$
$$\rho_3\gamma^{\beta a}\left(A^{abc}A^bA^c/2 + \alpha^{abc}A^bA^{cd}A^d - \rho_3\alpha^{abc}\alpha^{cde}A^cA^dA^e/2\right) -$$
$$\rho_4\gamma^{\beta a}\alpha^{abcd}A^bA^cA^d/3! + C^{\beta\gamma a}\beta_{\gamma\delta}\left(-\psi^\delta A^a + \gamma^{\delta b}A^aA^b\right) -$$
$$\gamma^{\beta\gamma a}\beta_{\gamma\delta}\left(-\psi^\delta A^{ab}A^b + \gamma^{\delta b}A^{ac}A^bA^c\right) - \rho_3\gamma^{\beta\gamma a}\beta_{\gamma\delta}\alpha^{abc}\left(-\psi^\delta A^bA^c + \gamma^{\delta d}A^bA^cA^d\right)/2 -$$
$$B^{\beta\gamma\delta}\left(-\beta_{\gamma\varepsilon}\psi^\varepsilon + \beta_{\gamma\varepsilon}\gamma^{\varepsilon a}A^a\right)\left(-\beta_{\delta\eta}\psi^\eta + \beta_{\delta\eta}\gamma^{\eta\beta}A^\beta\right)/2 -$$
$$\left.\beta^{\beta\gamma\delta\varepsilon}\left(-\beta_{\gamma\eta}\psi^\eta + \beta_{\gamma\eta}\gamma^{\eta a}A^a\right)\left(-\beta_{\delta\kappa}\psi^\kappa + \beta_{\delta\kappa}\gamma^{\kappa b}A^b\right)\left(-\beta_{\varepsilon\xi}\psi^\xi + \beta_{\varepsilon\xi}\gamma^{\zeta c}A^c\right)/3!\right].$$

■

Proof of Corollary 13. The result follows by direct calculations in (32) using using

$$\begin{aligned} E\left(A^{ab\cdots}A^{a_1b_1\cdots}\right) &= \alpha^{ab\ldots a_1b_1\cdots}/n,\ E\left(A^{ab\cdots}A^{a_1b_1\cdots}A^{a_2b_2\cdots}\right) = \alpha^{ab\ldots a_1b_1\ldots a_2b_2\cdots}/n^2, \\ E\left(A^{ab\cdots}A^{a_1b_1\cdots}A^{a_2b_2\cdots}A^{a_3b_3\cdots}\right) &= \alpha^{ab\ldots a_1b_1\ldots a_2b_2\ldots a_3b_3\cdots}/n^3 + [3]\left(n-1\right)\alpha^{ab\ldots a_1b_1\cdots}\alpha^{a_2b_2\ldots a_3b_3\cdots}/n^2 \end{aligned}$$

where $[3] = \alpha^{ab\ldots a_1b_1\cdots}\alpha^{a_2b_2\ldots a_3b_3\cdots} + \alpha^{ab\ldots a_2b_2\cdots}\alpha^{a_1b_1\ldots a_3b_3\cdots} + \alpha^{ab\ldots a_3b_3\cdots}\alpha^{a_1b_1\ldots a_2b_2\cdots}$, and simple algebra. ■

Proof of Corollary 14. The result follows by ULLN and CMT as in the proof of Theorem 7. ■

Proof of Proposition 15. We verify the conditions of Theorems 2 and 4. Note that (I)-(III) and $E\sup_{\theta_p\in\Theta}\|z_{2i}sign_p\{\varepsilon_i\}\| \leq (1+p)E\|z_2\| < \infty$ imply by Theorem 2 that $\widehat{\theta}_{pw} \xrightarrow{a.s.} \theta_{p0}$. Note also that by the results of Andrews (1994)

$$\sum z_{2i}sign_p\left\{\varepsilon_i - z_{1i}^\tau\left(\theta_p - \theta_{p0}\right)/n - E\left[p - F_\varepsilon\left(z_{1i}^\tau\left(\theta_p - \theta_{p0}\right)|z\right)\right]\right\}$$

is stochastically equicontinuous; furthermore by CLT

$$\sum z_{2i}sign_p\{\varepsilon_i\}/n^{1/2} \xrightarrow{d} N\left(0, p\left(1-p\right)E\left(z_2z_2'\right)\right)$$

so that by the differentiability condition (IV) it can be shown that $n^{1/2}\left(\widehat{\theta}_{pw} - \theta_{p0}\right) = O_p(1)$. Then (V)-(VII) imply the rest of the conditions of Theorem 4 hence the conclusion. To prove the consistency of $\widehat{\Sigma}_{\zeta g}$ note that

$$\begin{aligned}&\left\|\sum \widehat{w}_i I\left\{|\widehat{\varepsilon}_i| \leq b_n/2\right\} z_{1i} z_{2i}^{\tau}/b_n - E\left[f_{\varepsilon}\left(0|z\right) z_1 z_2^{\tau}\right]\right\| \leq \\ &\left\|\sum \left(I\left\{|\widehat{\varepsilon}_i| \leq b_n/2\right\} - I\left\{|\varepsilon_i| \leq b_n/2\right\}\right) z_{1i} z_{2i}^{\tau}/\left(n b_n\right)\right\| + \\ &\left\|\sum \left(I\left\{|\varepsilon_i| \leq b_n/2\right\} - E\left[f_{\varepsilon}\left(0|z\right) z_1 z_2^{\tau}\right]\right)\right\| + o_p\left(1\right) = A_{1n} + A_{2n}.\end{aligned}$$

By WLLN it is easy to see that $\sum I\left\{|\varepsilon_i| \leq b_n/2\right\}/\left(n b_n\right) \xrightarrow{p} E\left[f_{\varepsilon}\left(\lambda|z\right) z_1 z_2^{\tau}\right]$ where $|\lambda| \leq b_n = o\left(1\right)$ and hence by dominated convergence $E\left[f_{\varepsilon}\left(\lambda|z\right) z_1 z_2^{\tau}\right] \to E\left[f_{\varepsilon}\left(0|z\right) z_1 z_2^{\tau}\right]$; thus $A_{2n} = o_p\left(1\right)$ by triangle inequality. To show that $A_{1n} = o_p\left(1\right)$ note that

$$A_{1n} \leq \left\|\sum \left(I\left\{|\varepsilon_i + c_n| \leq b_n/2\right\} + I\left\{|\varepsilon_i - c_n| \leq b_n/2\right\}\right) z_{1i} z_{2i}^{\tau}/\left(n b_n\right)\right\|$$

where $c_n = \|z_{1i}\| \left\|\widehat{\theta}_{pw} - \theta_{p0}\right\|$. Let

$$A_{11n} = \left\|\sum I\left\{|\varepsilon_i + c_n| \leq b_n/2\right\} z_{1i} z_{2i}^{\tau}/\left(n b_n\right)\right\|,$$

and $B_n = \left\{\left\|\widehat{\theta}_{pw} - \theta_{p0}\right\| \leq h c_n\right\}$ for a constant $h > 0$ so that $\Pr\left(B_n^c\right) \to 0$ because $n^{1/2}\left(\widehat{\theta}_{pw} - \theta_0\right) = O_p\left(1\right)$. Then for any $\eta > 0$ using Markov inequality, III' and IX

$$\begin{aligned}&\Pr\left(A_{11n} > \eta\right) \leq \Pr\left(A_{11n} \cap B_n > \eta\right) + \Pr\left(B^c\right) = \\ &E \sum \left\|\int_{-(\|z_i\|+1)hc_n + c_n}^{(\|z_i\|+1)hc_n + c_n} f_{\varepsilon}\left(\lambda|z_i\right) d\lambda z_{1i} z_{2i}^{\tau}/\left(\eta n b_n\right)\right\| \leq \\ &E \sum \left\|\int_{-(\|z_i\|+1)hc_n + c_n}^{(\|z_i\|+1)hc_n + c_n} \overline{f} d\lambda z_{1i} z_{2i}^{\tau}/\left(\eta n b_n\right)\right\| = \left(h/\eta\right) E\left[\left(\|z\| + 1\right)\|z_1 z_2\|\right] \to 0\end{aligned}$$

for $h \to 0$. By similar arguments

$$A_{12n} = \left\|\sum I\left\{|\varepsilon_i - c_n| \leq b_n/2\right\} z_{1i} z_{2i}^{\tau}/\left(n b_n\right)\right\| = o_p\left(1\right)$$

implying that $A_{1n} = o_p\left(1\right)$.

Proof of Proposition 16. We verify the conditions of Theorems 3 and 4. Note that $\psi\left(\theta\right) = \left[\left(x_i - \theta\right) \wedge k\right] \vee \left(-k\right)$ is monotonic and (I) implies that θ_0 is unique; thus by Theorem 3 $\widehat{\theta}_w \xrightarrow{a.s.} \theta_0$. Also by the results of Andrews (1994)

$$\sum \left[\left(x_i - \theta\right) \wedge k\right] \vee \left(-k\right)/n - E\left[\left[\left(x_i - \theta\right) \wedge k\right] \vee \left(-k\right)\right]$$

is stochastically equicontinuous; furthermore by CLT $\sum\left[(x_i-\theta_0)\wedge k\right]\vee(-k)/n^{1/2}\xrightarrow{d}$ $N\left(0,V_k\left(\theta_0\right)\right)$ where $V_k\left(\theta_0\right)=\int_{\theta_0-k}^{\theta_0+k}x^2dF\left(x\right)+k^2\left(1+F\left(\theta_0-k\right)-F\left(\theta_0+k\right)\right)$. Clearly the differentiability condition I is satisfied hence $n^{1/2}\left(\widehat{\theta}_w-\theta_0\right)=O_p\left(1\right)$. Also

$$E\sup_{\theta\in N_0}\left|\left[(x_i-\theta)\wedge k\right]\vee(-k)\right|^2\leq\left(\delta+E\left|x\right|\right)^2\vee k^2<\infty,$$

and similarly for $E\sup_{\theta\in N_0}\left\|\left(\left[(x_i-\theta)\wedge k\right]\vee(-k)\right)g\left(x\right)\right\|<\infty$. Thus the conditions of Theorem 4 are met hence the conclusion. The strong consistency of $\widehat{\sigma}_{\theta g}^2$ follows by noting that by strong consistency of $\widehat{\theta}_w$ and the continuity of F imply $\sum\widehat{w}_iI\left\{x_i\leq\widehat{\theta}_w-k\right\}\xrightarrow{a.s}F\left(\theta_0-k\right)$. A similar argument applies to the other terms thus the conclusion follows by CMT. ■

Proof of Proposition 17. We verify the conditions of Theorems 2 and 6. Let $W\left(z,\theta\right)=\phi\left(z^\tau\theta\right)/F\left(z^\tau\theta\right)F\left(-z^\tau\theta\right)$,

$$E\left[\psi\left(x,\theta\right)\right]=E\left\{W\left(z,\theta\right)\left[F\left(z^\tau\theta_0\right)-F\left(z^\tau\theta\right)\right]z\right\}$$

which is clearly 0 at θ_0. Also note that as long as $\Pr\left\{z^\tau\left(\theta-\theta_0\right)\neq0\right\}>0$ the monotonicity of $F\left(\cdot\right)$ implies that θ_0 is unique. Thus by compactness of Θ and continuity of $E\left[\psi\left(z,\theta\right)\right]$ the identification condition $\inf_{\|\theta-\theta_0\|>\delta}\left\|E\left[\psi\left(z,\theta\right)\right]\right\|>0$ is satisfied, hence $\widehat{\theta}_w\xrightarrow{a.s}\theta_0$. Also $E\left[\psi\left(z,\theta_0\right)'\right]=E\left[\lambda\left(z^\tau\theta_0\right)\lambda\left(-z^\tau\theta_0\right)zz^\tau\right]$ exists and nonsingular by $\lambda\left(\cdot\right)$ bounded away from zero on any open interval and $E\left(zz^\tau\right)$ nonsingular, and $E\sup_{N_0}\left[\left\|\psi'\left(x,\theta\right)\right\|\left\|g\left(x\right)\right\|\right]=E\sup_{N_0}\left[\left\|\left(\lambda_v\left(v\right)y+\lambda_v\left(-v\right)\left(1-y\right)\right)zz^\tau\right\|\left\|g\left(x\right)\right\|\right]<2C\left(1+E\left\|z\right\|^2\left\|g\left(x\right)\right\|\right)<\infty$ by $\lambda_v\left(\cdot\right)$ uniformly bounded. Finally by CLT

$$\sum W\left(z_i,\theta_0\right)\left[y_i-F\left(z_i^\tau\theta_0\right)\right]z_i/n^{1/2}\xrightarrow{d}N\left(0,E\left[\lambda\left(z^\tau\theta_0\right)\lambda\left(-z^\tau\theta_0\right)zz^\tau\right]\right).$$

Thus all the conditions of Theorem 6 are met hence the result. Finally the strong consistency of the variance estimator follows by noting that by consistency of $\widehat{\theta}_w$, ULLN and the triangle inequality

$$\begin{aligned}&\left\|\sum\lambda\left(z_i^\tau\widehat{\theta}_w\right)\lambda\left(-z_i^\tau\widehat{\theta}_w\right)z_iz_i^\tau-E\left[\lambda\left(z^\tau\theta_0\right)\lambda\left(-z^\tau\theta_0\right)zz^\tau\right]\right\|\leq\\&\sum\sup_{\|\theta-\theta_0\|\leq\delta_n}\left\|\left(\lambda\left(z_i^\tau\widehat{\theta}_w\right)\lambda\left(-z_i^\tau\widehat{\theta}_w\right)-\lambda\left(z_i^\tau\theta_0\right)\lambda\left(-z_i^\tau\theta_0\right)\right)z_iz_i^\tau\right\|/n+o_{a.s}\left(1\right)+\\&\left\|\sum\lambda\left(z_i^\tau\theta_0\right)\lambda\left(-z_i^\tau\theta_0\right)z_iz_i^\tau/n-E\left[\lambda\left(z^\tau\theta_0\right)\lambda\left(-z^\tau\theta_0\right)zz^\tau\right]\right\|=o_{a.s}\left(1\right),\end{aligned}$$

where $\delta_n\to0$ such that $\|\theta-\theta_0\|\leq\delta_n$ $a.s.$ A similar argument can be used to show the strong consistency of the other terms appearing in the estimator so the conclusion follows by CMT. ■ ■

6 Figures and tables

Table 1a Finite sample bias B, variances V, $\widehat{V}$ and efficiency E of $\widehat{\theta}$, $\widehat{\theta}^{GMM}$ and $\widehat{\theta}_w^{GEL}$ in instrumental variable median regression model for $n = 50$ and

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1^{GMM}$	$\widehat{\theta}_2^{GMM}$	$\widehat{\theta}_{1w}^{EL}$	$\widehat{\theta}_{2w}^{EL}$	$\widehat{\theta}_{1w}^{EU}$	$\widehat{\theta}_{2w}^{EU}$	$\widehat{\theta}_{1w}^{NT}$	$\widehat{\theta}_{2w}^{NT}$
$N(0,1)$ errors										
$g(x)^{cs}$										
B	-.052	.006	.031	$\underline{.004}$	-.022	$\underline{.004}$	.028	.006	-.027	.006
V	.235	.166	.218	.157	$\underline{.196}$	$\underline{.151}$	.211	.158	.200	.153
$\widehat{V}$	.233	.168	.217	.160	.199	.152	.209	.158	.203	.150
E	1.00	1.00	1.08	1.06	$\overline{1.19}$	$\overline{1.10}$	1.11	1.05	1.17	1.08
$g_1(x)^{ms}$										
B	-.052	$\underline{.006}$	.034	.010	$\underline{-.031}$	.013	.035	.014	-.033	.011
V	.235	.166	.226	.166	.223	.159	.226	.158	$\underline{.221}$	$\underline{.156}$
$\widehat{V}$	.233	.168	.224	.169	.224	.162	.228	.155	.224	.158
E	1.00	1.00	1.04	1.00	1.05	1.04	1.04	1.05	$\overline{1.06}$	$\overline{1.06}$
$g_2(x)^{ms}$										
B	-.052	$\underline{.006}$	.038	.020	-.039	.018	.041	.017	$\underline{-.037}$	.016
V	.235	.166	.232	.170	.230	.165	.233	.164	$\underline{.228}$	$\underline{.159}$
$\widehat{V}$	.233	.168	.235	.172	.230	.168	.235	.168	.231	.161
E	1.00	1.00	1.01	.976	1.02	1.01	1.01	1.01	$\overline{1.03}$	$\overline{1.04}$
$t(4)$ errors										
$g(x)^{cs}$										
B	-.042	.006	.023	.007	$\underline{-.020}$	$\underline{.005}$	.023	.006	-.021	$\underline{.005}$
V	.242	.203	.229	.178	$\underline{.220}$	$\underline{.167}$	.227	.173	$\underline{.220}$	.168
$\widehat{V}$	.246	.206	.228	.180	.222	.165	.225	.183	.222	.170
E	1.00	1.00	1.06	1.14	$\overline{1.10}$	$\overline{1.21}$	1.06.	1.17	$\overline{1.10}$	$\overline{1.21}$
$g_1(x)^{ms}$										
B	-.042	$\underline{.006}$	.033	.012	-.034	.012	.036	.013	$\underline{-.033}$	.010
V	.242	.203	.234	.189	.236	.190	.238	$\underline{.181}$	$\underline{.231}$	.184
$\widehat{V}$	.246	.206	.235	.193	.239	.193	.240	.193	.233	.188
E	1.00	1.00	1.03	1.07	1.02	1.07	1.01	$\overline{1.12}$	$\overline{1.05}$	1.10
$g_2(x)^{ms}$										
B	-.042	$\underline{.006}$	.039	.028	-.038	.020	.038	.021	-.036	.018
V	.242	.203	.239	.199	.241	.198	.240	$\underline{.190}$	$\underline{.237}$	.193
$\widehat{V}$	.246	.206	.247	.205	.246	.202	.245	.203	.238	.196
E	1.00	1.00	1.01	1.02	1.00	1.02	1.00	$\overline{1.06}$	$\overline{1.02}$	1.05

GMM Efficient GMM, EL Empirical likelihood, EU Euclidean likelihood, NT Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 1a. Continued

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1^{GMM}$	$\widehat{\theta}_2^{GMM}$	$\widehat{\theta}_{1w}^{EL}$	$\widehat{\theta}_{2w}^{EL}$	$\widehat{\theta}_{1w}^{EU}$	$\widehat{\theta}_{2w}^{EU}$	$\widehat{\theta}_{1w}^{NT}$	$\widehat{\theta}_{2w}^{NT}$
$\chi^2(4)-4$ errors										
$g(x)^{cs}$										
B	-.081	-.032	-.067	-.029	-.061	-.028	-.063	-.027	-.064	-.029
V	.419	.188	.390	.169	.385	.158	.389	.164	.385	.161
$\widehat{V}$	.428	.197	.398	.177	.389	.163	.393	.167	.394	.164
E	1.00	1.00	1.07	1.11	1.08	1.19	1.07	1.15	1.08	1.17
$g_1(x)^{ms}$										
B	-.081	-.032	-.077	-.036	-.072	-.037	-.073	-.033	-.073	-.035
V	.419	.188	.398	.180	.394	.169	.398	.173	.394	.168
$\widehat{V}$	.428	.197	.405	.184	.406	.175	.402	.177	.401	.174
E	1.00	1.00	1.05	1.04	1.06	1.11	1.05	1.09	1.06	1.12
$g_2(x)^{ms}$										
B	-.081	-.032	-.083	-.041	-.084	-.042	-.079	-.038	-.078	-.040
V	.419	.188	.415	.185	.411	.181	.408	.178	.408	.175
$\widehat{V}$	.428	.197	.421	.193	.416	.184	.418	.184	.414	.179
E	1.00	1.00	1.01	1.01	1.02	1.04	1.02	1.05	1.03	1.05
$\frac{\chi^2(4)-4}{\sqrt{8}}$ errors										
$g(x)^{cs}$										
B	-.057	-.023	-.049	-.019	-.041	-.018	-.044	-.021	-.043	-.018
V	.275	.160	.249	.150	.238	.143	.243	.151	.241	.144
$\widehat{V}$	.280	.168	.252	.157	.241	.147	.246	.155	.257	.145
E	1.00	1.00	1.10	1.07	1.15	1.12	1.13	1.06	1.14	1.11
$g_1(x)^{ms}$										
B	-.057	-.023	-.062	-.022	-.049	-.021	-.053	-.024	-.047	-.019
V	.275	.160	.261	.154	.256	.150	.251	.155	.251	.145
$\widehat{V}$	.280	.168	.265	.159	.259	.156	.260	.158	.254	.152
E	1.00	1.00	1.05	1.04	1.07	1.07	1.09	1.03	1.09	1.10
$g_2(x)^{ms}$										
B	-.057	-.023	-.069	-.028	-.054	-.024	-.058	-.027	-.052	-.020
V	.275	.160	.268.	.156	.264	.158	.266	.157	.261	.152
$\widehat{V}$	.280	.168	.271	.165	.268	.162	.271	.164	.264	.156
E	1.00	1.00	1.03	1.01	1.04	1.01	1.03	1.02	1.05	1.05

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correct and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 1b. Finite sample bias B, variances V, $\widehat{V}$ and efficiency E of $\widehat{\theta}$, $\widehat{\theta}^{GMM}$ and $\widehat{\theta}_w^{GEL}$ in instrumental variable median regression model for $n = 100$ and

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1^{GMM}$	$\widehat{\theta}_2^{GMM}$	$\widehat{\theta}_{1w}^{EL}$	$\widehat{\theta}_{2w}^{EL}$	$\widehat{\theta}_{1w}^{EU}$	$\widehat{\theta}_{2w}^{EU}$	$\widehat{\theta}_{1w}^{NT}$	$\widehat{\theta}_{2w}^{NT}$
$N(0,1)$ errors										
$g(x)^{cs}$										
B	-.037	.005	.021	.006	-$\underline{.015}$	$\underline{.003}$	.020	.004	-.018	.004
V	.194	.136	.175	.126	$\underline{.152}$	$\underline{.122}$	.168	.128	.159	$\underline{.122}$
$\widehat{V}$	.193	.135	.174	.129	.154	.124	.166	.127	.161	.123
E	1.00	1.00	1.10	1.07	$\overline{1.28}$	$\overline{1.11}$	1.15	1.06	1.22	$\overline{1.11}$
$g_1(x)^{ms}$										
B	-.037	$\underline{.005}$	.023	.008	$\underline{-.021}$	.007	.023	.009	$\underline{-.021}$	.006
V	.194	.136	.179	.128	$\underline{.164}$	.127	.172	.131	.166	$\underline{.126}$
$\widehat{V}$	.193	.135	.176	.127	.166	.124	.174	.129	.167	.126
E	1.00	1.00	1.08	1.06	$\overline{1.18}$	1.08	1.13	1.04	1.17	$\overline{1.07}$
$g_2(x)^{ms}$										
B	-.037	.005	.026	.010	-.032	.009	.027	.011	$\underline{-.024}$	.008
V	.194	.136	.183	.132	.183	.133	.180	.136	$\underline{.178}$	$\underline{.132}$
$\widehat{V}$	.193	.135	.185	.134	.187	.130	.185	.132	.181	.130
E	1.00	1.00	1.06	1.03	1.06	1.02	1.08	1	$\overline{1.09}$	$\overline{1.05}$
$t(4)$ errors										
$g(x)^{cs}$										
B	-.034	.005	.019	.006	$\underline{-.016}$	$\underline{.004}$	.018	.005	-.017	$\underline{.004}$
V	.182	.153	.170	.132	$\underline{.162}$	$\underline{.130}$	.167	.134	$\underline{.162}$	.132
$\widehat{V}$	.181	.152	.168	.134	.164	.128	.156	.133	.162	.130
E	1.00	1.00	1.07	1.16	$\overline{1.12}$	$\overline{1.17}$	1.09	1.14	$\overline{1.12}$	1.16
$g_1(x)^{ms}$										
B	-.034	$\underline{.005}$	$\underline{.025}$	.008	-.028	.009	.027	.010	$\underline{-.025}$	.006
V	.182	.153	.174	.141	$\underline{.170}$	.137	.177	.139	$\underline{.170}$	$\underline{.132}$
$\widehat{V}$	.181	.152	.177	.143	.172	.139	.175	.142	.175	.134
E	1.00	1.00	1.05	1.08	$\overline{1.07}$	1.12	1.03	1.10	$\overline{1.07}$	$\overline{1.16}$
$g_2(x)^{ms}$										
B	-.034	$\underline{.005}$	.033	.011	-.031	.012	.032	.012	$\underline{-.029}$	.010
V	.182	.153	.177	.141	.178	.141	.180	.142	$\underline{.175}$	$\underline{.135}$
$\widehat{V}$	.181	.152	.179	.143	.180	.144	.183	.145	.177	.138
E	1.00	1.00	1.03	1.08	1.02	1.08	1.01	1.07	$\overline{1.04}$	$\overline{1.13}$

GMM Efficient GMM, EL Empirical likelihood, EU Euclidean likelihood, NT Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 1b. Continued

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1^{GMM}$	$\widehat{\theta}_2^{GMM}$	$\widehat{\theta}_{1w}^{EL}$	$\widehat{\theta}_{2w}^{EL}$	$\widehat{\theta}_{1w}^{EU}$	$\widehat{\theta}_{2w}^{EU}$	$\widehat{\theta}_{1w}^{NT}$	$\widehat{\theta}_{2w}^{NT}$
$\chi^2(4)-4$ errors										
$g(x)^{cs}$										
B	-.061	-.024	-.050	-.022	$\underline{-.046}$	$\underline{-.021}$	-.047	-.020	-.048	-.022
V	.291	.131	.269	.115	$\underline{.263}$	$\underline{.108}$	.271	.113	.268	.110
$\widehat{V}$	.304	.145	.265	.120	.260	.110	.270	.116	.265	.112
E	1.00	1.00	1.08	1.14	$\overline{1.11}$	$\overline{1.21}$	1.08	1.16	1.08	1.19
$g_1(x)^{ms}$										
B	-.061	$\underline{-.024}$	-.056	.026	$\underline{-.052}$	-.027	-.053	-.025	-.053	$\underline{-.024}$
V	.291	.131	.271	.122	.282	.118	.284	.120	$\underline{.280}$	$\underline{.117}$
$\widehat{V}$	.304	.145	.283	.138	.279	.132	.281	.131	.277	.130
E	1.00	1.00	1.07	1.07	1.03	1.11	1.02	1.09	$\overline{1.04}$	$\overline{1.12}$
$g_2(x)^{ms}$										
B	-.061	$\underline{-.024}$	-.060	-.029	-.058	-.029	-.056	-.028	$\underline{-.055}$	-.027
V	.291	.131	.287	.125	.290	.128	.290	.126	$\underline{.286}$	$\underline{.123}$
$\widehat{V}$	.304	.145	.301	.140	.290	.124	.292	.129	.288	.133
E	1.00	1.00	1.01	1.05	1.00	1.02	1.00	1.04	$\overline{1.02}$	$\overline{1.06}$
$\frac{\chi^2(4)-4}{\sqrt{8}}$ errors										
$g(x)^{cs}$										
B	-.045	-.018	-.043	-.015	$\underline{-.032}$	-.014	-.034	-.016	$\underline{-.032}$	$\underline{-.013}$
V	.204	.118	.182	.110	$\underline{.174}$	$\underline{.104}$	$\underline{.174}$	.109	.176	.106
$\widehat{V}$	.206	.120	.185	.111	.178	.099	.182	.104	.180	.102
E	1.00	1.00	1.12	1.07	$\overline{1.17}$	$\overline{1.13}$	$\overline{1.17}$	1.08	1.16	1.11
$g_1(x)^{ms}$										
B	-.045	-.018	-.046	-.017	-.037	-.016	-.040	-.018	$\underline{-.036}$	$\underline{-.014}$
V	.204	.118	.191	.115	.186	.112	.186	.114	$\underline{.184}$	$\underline{.110}$
$\widehat{V}$	.206	.120	.193	.116	.189	.115	.191	.116	.188	.115
E	1.00	1.00	1.07	1.03	1.10	1.05	1.10	1.03	$\overline{1.10}$	$\overline{1.07}$
$g_2(x)^{ms}$										
B	-.045	-.018	-.049	-.020	-.043	-.019	-.042	-.020	$\underline{-.041}$	$\underline{-.015}$
V	.204	.118	.195	.118	.191	.116	.192	.116	$\underline{.187}$	$\underline{.114}$
$\widehat{V}$	.206	.120	.197	.121	.195	.122	.195	.121	.192	.121
E	1.00	1.00	1.04	1.00	1.07	1.01	1.06	1.02	$\overline{1.09}$	$\overline{1.03}$

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 2. Finite sample bias B, variances V, $\widehat{V}$ and efficiency E of $\widehat{\theta}$, $\widehat{\theta}^{GMM}$, and $\widehat{\theta}_w^{GEL}$ in robust location estimation with

p		$\widehat{\theta}$	$\widehat{\theta}^{GMM}$	$\widehat{\theta}_w^{EL}$	$\widehat{\theta}_w^{EU}$	$\widehat{\theta}_w^{NT}$
				$N(0,1)$		
.25	B	.002	.002	.001	.002	.001
	V	.015	.011	.008	.0104	.0083
	$\widehat{V}$	.016	.012	.008	.0103	.0846
	E	1.00	1. 36	1.83	1.480	1.855
.40	B	-.0073	-.0033	.0013	-.0248	.0014
	V	.0150	.0100	.0061	.0092	.0062
	$\widehat{V}$	.0144	.0104	.0059	.0112	.0060
	E	1.00	1.500	2.459	1.630	2.419
.60	B	.0041	.0039	-.0022	.0038	-.0021
	V	.0171	.0078	.0046	.0080	.0059
	$\widehat{V}$	.0182	.0825	.0059	.0102	.0050
	E	1.00	2.192	3. 717	2.137	3. 423
.75	B	.0047	.0040	-.0040	.0038	-.0042
	V	.0187	.0107	.0075	.0103	.0082
	$\widehat{V}$	.0192	.0102	.0078	.0101	.0089
	E	1.00	1. 747	2.493	1.815	2.280
				$t(4)$		
.25	B	-.0063	-.0059	.0031	-.0060	.0031
	V	.0220	.0149	.0119	.0153	.0112
	$\widehat{V}$	.0229	.0153	.0120	.0138	.0114
	E	1.00	1. 476	1.848	1.437	1.964
.40	B	-.0012	-.0011	-.0013	-.0011	.0013
	V	.0236	.0132	.0079	.0143	.0087
	$\widehat{V}$	.0227	.0139	.0075	.0129	.0081
	E	1.00	1. 787	2.987	1.642	2.712

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting
An underline (overline) indicates smallest (largest) value in the corresponding row.

Table 2. Continued

p		$\widehat{\theta}$	$\widehat{\theta}^{GMM}$	$\widehat{\theta}_w^{EL}$	$\widehat{\theta}_w^{EU}$	$\widehat{\theta}_w^{NT}$
				$t\,(4)$		
.60	B	-.0028	-.0025	-.0020	-.0031	-.0029
	V	.0244	.0090	.0085	.0101	.0090
	$\widehat{V}$	.0239	.0119	.0090	.0103	.0095
	E	1.00	2.711	2. 870	2. 415	2. 711
.75	B	-.0060	-.0063	-.0058	-.0052	-.0051
	V	.0259	.0154	.0133	.0163	.0131
	$\widehat{V}$	.0269	.0169	.0116	.0147	.0127
	E	1.00	1. 681	1.947	1.589	1.977
				$\chi^2\,(4) - 4$		
.25	B	-.2414	-.2512	-.2025	-.2261	-.2325
	V	.2587	.2001	.1885	.2088	.1844
	$\widehat{V}$	.2232	.2002	.2033	.1927	.1995
	E	1.00	1. 292	1.372	1. 2930	1.402
.40	B	-.2271	-.2243	-.2073	-.2152	-.2131
	V	.2323	.1623	.1196	.1734	.1205
	$\widehat{V}$	.2168	.1598	.1123	.1655	.1099
	E	1.00	1. 431	1.939	1.341	1.925
.60	B	-.2162	-.2100	-.1996	-.209	-.1956
	V	.2420	.1420	.1100	.1571	.1006
	$\widehat{V}$	.233	.1533	.0941	.1553	.0904
	E	1.00	1.704	2. 437	1. 540	2. 657
.75	B	-.2105	-.2054	-.2289	-.2107	-.2128
	V	.2599	.1599	.1093	.1609	.1150
	$\widehat{V}$	.2407	.1407	.0965	.1779	.1043
	E	1.00	1. 625	2.377	1.615	2.260

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting
An underline (overline) indicates smallest (largest) value in the corresponding row.

Table 2. Continued

p		$\widehat{\theta}$	$\widehat{\theta}^{GMM}$	$\widehat{\theta}_w^{EL}$	$\widehat{\theta}_w^{EU}$	$\widehat{\theta}_w^{NT}$
$\frac{\chi^2(4)-4}{\sqrt{8}}$						
.25	B	-.2196	-.2009	-.$\underline{1823}$	-.196 7	-.1836
	V	.0310	.0260	.$\underline{0245}$	.0271	.0248
	$\widehat{V}$	.0334	.0240	.0284	.0297	.$\underline{0243}$
	E	1.00	1. 192	1.265	1.148	1. 25
.40	B	-.1816	-.1861	-.$\underline{1554}$	-.1571	-.1622
	V	.0278	.02 27	.$\underline{0148}$	.0222	.0156
	$\widehat{V}$	.0238	.0207	.0135	.0216	.$\underline{0124}$
	E	1.00	1. 227	$\overline{1.883}$	1. 255	1. 786
.60	B	-.1751	-.1743	-.1576	-.1692	-.$\underline{1525}$
	V	.0292	.0205	.0144	.0210	.$\underline{0131}$
	$\widehat{V}$	.0302	.0193	.0119	.0184	.$\underline{0112}$
	E	1.00	1. 424	2. 027	1. 390	$\overline{2.229}$
.75	B	-.1915	-.1828	-.2060	-.1791	-.$\underline{1745}$
	V	.0317	.0219	.$\underline{0132}$	.0204	.0141
	$\widehat{V}$	.2407	.1407	.$\underline{0965}$	.1779	.1043
	E	1.00	1.447	$\overline{2.401}$	1.553	2.248

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting
An underline (overline) indicates smallest (largest) value in the corresponding row.

Table 3a. Finite sample bias B, variances V, $\widehat{V}$ and efficiency E of $\widehat{\theta}$, $\widehat{\theta}^{GMM}$ and $\widehat{\theta}_w^{GEL}$ in binary dependent variable regression model for $n = 50$ and

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1^{GMM}$	$\widehat{\theta}_2^{GMM}$	$\widehat{\theta}_{1w}^{EL}$	$\widehat{\theta}_{2w}^{EL}$	$\widehat{\theta}_{1w}^{EU}$	$\widehat{\theta}_{2w}^{EU}$	$\widehat{\theta}_{1w}^{NT}$	$\widehat{\theta}_{2w}^{NT}$
$N(0,1)$										
$g(x)^{cs}$										
B	.057	.041	.052	.038	$\underline{.049}$	$\underline{.033}$	.054	.036	.050	.035
V	.086	.090	.070	.068	$\underline{.063}$	.062	.069	.065	.064	$\underline{.061}$
$\widehat{V}$	.088	.092	.073	.070	.065	.064	.073	.068	.066	.067
E	1.00	1.00	1.22	1.32	$\overline{1.36}$	1.47	1.25	1.38	1.34	$\overline{1.47}$
$g_1(x)^{ms}$										
B	.057	.041	.058	.042	.056	$\underline{.038}$	.058	.039	$\underline{.053}$	.039
V	.086	.090	.079	.074	.076	.078	$\underline{.071}$	.077	$\underline{.071}$	$\underline{.074}$
$\widehat{V}$	.088	.092	.083	.078	.081	.082	.077	.081	.075	.077
E	1.00	1.00	1.09	1.27	1.13	1.15	$\overline{1.21}$	1.17	$\overline{1.21}$	$\overline{1.21}$
$g_2(x)^{ms}$										
B	.057	.041	.063	.047	.063	.044	.066	.044	.060	.043
V	.086	.090	.084	.079	.081	.078	.080	.081	$\underline{.078}$	$\underline{.078}$
$\widehat{V}$	.088	.092	.087	.082	.084	.082	.083	.088	.082	.083
E	1.00	1.00	1.02	1.14	1.06	1.15	1.07	1.11	$\overline{1.10}$	$\overline{1.15}$
$t(4)$										
$g(x)^{cs}$										
B	-.141	-.108	-.136	-.091	$\underline{-.120}$	$\underline{-.071}$	-.126	-.076	-.123	-.072
V	.055	.104	.046	.084	$\underline{.042}$	$\underline{.081}$	.043	.085	$\underline{.042}$	.080
$\widehat{V}$	.061	.106	.048	.089	.018	.028	.021	.030	.020	.029
E	1.00	1.00	1.19	1.24	$\overline{1.31}$	$\overline{1.28}$	1.28	1.22	$\overline{1.31}$	1.30
$g_1(x)^{ms}$										
B	-.141	-.108	-.142	-.095	-.138	-.085	-.151	$\underline{-.082}$	$\underline{-.135}$	$\underline{-.082}$
V	.055	.104	.051	.091	.048	.088	.048	.090	$\underline{.046}$	$\underline{.087}$
$\widehat{V}$	.061	.106	.056	.097	.052	.092	.053	.093	.049	.091
E	1.00	1.00	1.08	1.14	1.14	1.18	1.14	1.15	$\overline{1.19}$	$\overline{1.19}$
$g_2(x)^{ms}$										
B	$\underline{-.141}$	-.108	-.148	-.101	-.145	-.093	-.152	-.093	$\underline{-.141}$	$\underline{-.088}$
V	.055	.104	.055	.096	.054	.093	.053	.094	$\underline{.051}$	$\underline{.091}$
$\widehat{V}$	.061	.106	.060	.101	.056	.096	.056	.097	.059	.096
E	1.00	1.00	1.00	1.08	1.07	1.11	1.04	1.10	$\overline{1.08}$	$\overline{1.14}$

GMM Efficient GMM, EL Empirical likelihood, EU Euclidean likelihood, NT Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 3a. Continued

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1^{GMM}$	$\widehat{\theta}_2^{GMM}$	$\widehat{\theta}_{1w}^{EL}$	$\widehat{\theta}_{2w}^{EL}$	$\widehat{\theta}_{1w}^{EU}$	$\widehat{\theta}_{2w}^{EU}$	$\widehat{\theta}_{1w}^{NT}$	$\widehat{\theta}_{2w}^{NT}$
$\chi^2(4)-4$										
$g(x)^{cs}$										
B	-.895	-.305	-.825	-.270	-.785	-.236	-.807	-.262	-.797	-.248
V	.222	.251	.182	.206	.161	.187	.176	.195	.165	.190
$\widehat{V}$	.235	.264	.196	.212	.174	.193	.189	.203	.177	.196
E	1.00	1.00	1. 22	1. 22	1. 38	1. 34	1. 26	1. 29	1. 34	1.32
$g_1(x)^{ms}$										
B	-.895	-.305	-.850	-.288	-.832	-.266	-.829	-.274	-.821	-.264
V	.222	.251	.201	.215	.186	.201	.193	.204	.184	.194
$\widehat{V}$	.235	.264	.214	.221	.118	.140	.129	.139	.119	.135
E	1.00	1.00	1. 10	1.17	1. 19	1.25	1. 15	1. 23	1. 21	1. 29
$g_2(x)^{ms}$										
B	-.895	-.305	-.878	-.298	-.856	-.296	-.845	-.288	-.832	-.285
V	.222	.251	.220	.240	.210	.237	.217	.235	.192	.215
$\widehat{V}$	.235	.264	.153	.161	.143	.153	.148	.160	.139	.152
E	1.00	1.00	1. 01	1. 04	1. 06	1. 06	1. 02	1. 07	1. 14	1. 17
$\frac{\chi^2(4)-4}{\sqrt{8}}$										
$g(x)^{cs}$										
B	.531	.318	.509	.289	.497	.278	.505	.298	.491	.291
V	.042	.053	.030	.038	.028	.032	.030	.039	.028	.031
$\widehat{V}$	.045	.057	.034	.042	.030	.034	.035	.041	.031	.034
E	1.00	1.00	1. 40	1. 39	1. 50	1.65	1. 40	1. 35	1. 50	1. 71
$g_1(x)^{ms}$										
B	.531	.318	.513	.302	.502	.298	.507	.300	.498	.296
V	.042	.053	.034	.038	.034	.037	.033	.040	.033	.035
$\widehat{V}$	.045	.057	.039	.042	.038	.041	.038	.042	.035	.038
E	1.00	1.00	1. 23	1. 39	1. 23	1.43	1. 27	1. 32	1. 27	1.51
$g_2(x)^{ms}$										
B	.531	.318	.518	.306	.510	.309	.512	.306	.507	.303
V	.042	.053	.037	.040	.037	.041	.036	.039	.035	.037
$\widehat{V}$	.045	.057	.041	.044	.042	.045	.039	.043	.039	.042
E	1.00	1.00	1. 13	1. 32	1.13	1.29	1. 16	1. 23	1.20	1.26

GMM Efficient GMM, EL Empirical likelihood, EU Euclidean likelihood, NT Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 3b. Finite sample bias B, variances V, $\widehat{V}$ and efficiency E of $\widehat{\theta}$, $\widehat{\theta}^{GMM}$ and $\widehat{\theta}^{GEL}_w$ in binary dependent variable regression model for $n = 100$ and

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}^{GMM}_1$	$\widehat{\theta}^{GMM}_2$	$\widehat{\theta}^{EL}_{1w}$	$\widehat{\theta}^{EL}_{2w}$	$\widehat{\theta}^{EU}_{1w}$	$\widehat{\theta}^{EU}_{2w}$	$\widehat{\theta}^{NT}_{1w}$	$\widehat{\theta}^{NT}_{2w}$
					$N(0,1)$					
					$g(x)^{cs}$					
B	.030	.021	.027	.020	.025	.017	.028	.019	.024	.019
V	.028	.031	.024	.023	.020	.021	.023	.024	.021	.022
$\widehat{V}$	.026	.030	.023	.023	.019	.021	.023	.022	.022	.023
E	1.00	1.00	1.20	1.31	1.39	1.44	1.22	1.29	1.32	1.40
					$g_1(x)^{ms}$					
B	.030	.021	.030	.022	.029	.020	.031	.023	.028	.021
V	.028	.031	.027	.025	.024	.025	.026	.027	.024	.024
$\widehat{V}$	.026	.030	.025	.025	.023	.025	.026	.024	.025	.026
E	1.00	1.00	1.04	1.24	1.17	1.24	1.08	1.15	1.17	1.29
					$g_2(x)^{ms}$					
B	.030	.021	.032	.024	.032	.022	.034	.023	.031	.023
V	.028	.031	.029	.027	.028	.027	.028	.029	.027	.026
$\widehat{V}$	.026	.030	.027	.027	.027	.027	.027	.027	.028	.026
E	1.00	1.00	0.96	1.15	1.00	1.15	1.00	1.07	1.04	1.19
					$t(4)$					
					$g(x)^{cs}$					
B	-.125	-.089	-.112	-.077	-.098	-.075	-.109	-.078	-.106	-.075
V	.046	.056	.036	.042	.033	.036	.035	.040	.033	.038
$\widehat{V}$	.045	.058	.037	.044	.035	.040	.037	.042	.036	.040
E	1.00	1.00	1.28	1.33	1.39	1.55	1.31	1.40	1.39	1.47
					$g_1(x)^{ms}$					
B	-.125	-.089	-.118	-.083	-.115	-.082	-.116	-.083	-.111	-.080
V	.046	.056	.040	.046	.039	.043	.038	.043	.036	.041
$\widehat{V}$	.045	.058	.043	.049	.041	.045	.041	.045	.039	.043
E	1.00	1.00	1.15	1.22	1.18	1.30	1.21	1.30	1.27	1.36
					$g_2(x)^{ms}$					
B	-.125	-.089	-.124	-.089	-.123	-.090	-.126	-.085	-.121	-.085
V	.046	.056	.044	.050	.045	.048	.042	.046	.042	.047
$\widehat{V}$	.045	.058	.047	.052	.048	.051	.020	.031	.019	.030
E	1.00	1.00	1.04	1.12	1.02	1.17	1.09	1.22	1.09	1.19

GMM Efficient GMM, EL Empirical likelihood, EU Euclidean likelihood, NT Nonparametric tilting. $g(x)^{cs}$, $g(x)^{ms}_j$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 3b. Continued

	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1^{GMM}$	$\widehat{\theta}_2^{GMM}$	$\widehat{\theta}_{1w}^{EL}$	$\widehat{\theta}_{2w}^{EL}$	$\widehat{\theta}_{1w}^{EU}$	$\widehat{\theta}_{2w}^{EU}$	$\widehat{\theta}_{1w}^{NT}$	$\widehat{\theta}_{2w}^{NT}$
$\chi^2(4)-4$										
$g(x)^{cs}$										
B	-.851	-.277	-.823	-.241	-.815	-.229	-.820	-.232	-.827	-.227
V	.196	.229	.142	.151	.121	.136	.132	.140	.126	.140
$\widehat{V}$	.187	.203	.156	.174	.138	.150	.143	.156	.138	.106
E	1.00	1.00	1. 38	1. 51	1.61	1.68	1.48	1.63	1.55	1.63
$g_1(x)^{ms}$										
B	-.851	-.277	-.837	-.258	-.839	-.244	-.835	-.243	-.835	-.238
V	.196	.229	.156	.176	.139	.159	.144	.158	.135	.154
$\widehat{V}$	.187	.203	.172	.191	.153	.185	.159	.171	.148	.172
E	1.00	1.00	1. 25	1. 30	1. 41	1. 44	1. 36	1.44	1. 45	1. 48
$g_2(x)^{ms}$										
B	-.851	-.277	-.848	-.269	-.847	-.257	-.844	-.252	-.840	-.248
V	.196	.229	.185	.199	.171	.189	.167	.183	.167	.178
$\widehat{V}$	.187	.203	.198	.212	.188	.201	.184	.196	.179	.192
E	1.00	1.00	1. 05	1. 15	1. 14	1. 21	1. 17	1. 25	1. 17	1. 28
$\frac{\chi^2(4)-4}{\sqrt{8}}$										
$g(x)^{cs}$										
B	.377	.241	.369	.212	.363	.208	.366	.215	.366	.208
V	.037	.040	.028	.032	.024	.025	.026	.027	.024	.027
$\widehat{V}$	.039	.043	.031	.036	.027	.027	.028	.029	.028	.029
E	1.00	1.00	1. 32	1. 25	1. 54	1. 60	1. 42	1. 48	1. 54	1. 48
$g_1(x)^{ms}$										
B	.377	.241	.372	.229	.367	.221	.379	.222	.366	.217
V	.037	.040	.032	.035	.029	.031	.028	.032	.028	.030
$\widehat{V}$	.039	.043	.035	.038	.034	.035	.032	.035	.030	.033
E	1.00	1.00	1. 15	1. 14	1. 27	1. 29	1. 32	1. 25	1. 32	1. 33
$g_2(x)^{ms}$										
B	.377	.241	.376	.235	.378	.225	.383	.230	.373	.221
V	.037	.040	.037	.039	.035	.035	.032	.035	.032	.034
$\widehat{V}$	.039	.043	.040	.041	.039	.040	.037	.038	.035	.037
E	1.00	1.00	1.00	1. 02	1. 05	1. 14	1. 15	1. 14	1. 15	1. 17

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 4a. Finite sample bias B, variances V, $\widehat{V}$ and efficiency E of the bias corrected $\widehat{\theta}^{c}$, $\widehat{\theta}^{c,GMM}$ and $\widehat{\theta}^{c,GEL}_{w}$ in binary dependent variable regression model for $n = 50$ and

	$\widehat{\theta}_1^c$	$\widehat{\theta}_2^c$	$\widehat{\theta}_1^{c,GMM}$	$\widehat{\theta}_2^{c,GMM}$	$\widehat{\theta}_{1w}^{c,EL}$	$\widehat{\theta}_{2w}^{c,EL}$	$\widehat{\theta}_{1w}^{c,EU}$	$\widehat{\theta}_{2w}^{c,EU}$	$\widehat{\theta}_{1w}^{c,NT}$	$\widehat{\theta}_{2w}^{c,NT}$
					$N(0,1)$					
					$g(x)^{cs}$					
B	$\underline{.004}$	.010	.005	.012	$\underline{.004}$	$\underline{.007}$	.006	.009	.005	.010
V	.038	.050	.030	.030	$\underline{.028}$	$\underline{.026}$	.033	.032	.030	.029
$\widehat{V}$	.040	.053	.032	.034	.029	.030	.034	.034	.031	.031
E	1.00	1.00	1. 27	1. 67	$\overline{1.35}$	$\overline{1.92}$	1. 15	1. 56	1. 27	1. 72
					$g_1(x)^{ms}$					
B	$\underline{.004}$	$\underline{.010}$	.012	.022	.014	.021	.014	.024	.010	.020
V	.038	.050	.034	.032	$\underline{.031}$	.033	.032	.034	$\underline{.031}$	$\underline{.034}$
$\widehat{V}$	.040	.053	.036	.038	.034	.037	.036	.035	.036	.038
E	1.00	1.00	1. 12	1.56	$\overline{1.22}$	1.51	1. 19	1. 47	$\overline{1.22}$	$\overline{1.47}$
					$g_2(x)^{ms}$					
B	$\underline{.004}$	$\underline{.010}$	.018	.013	.019	.012	.019	.014	.016	.012
V	.038	.050	.040	.038	.037	.035	.038	.038	$\underline{.035}$	$\underline{.037}$
$\widehat{V}$	.040	.053	.043	.042	.040	.039	.041	.039	.038	.023
E	1.00	1.00	0.95	1. 31	1. 03	1 43	1.00	1.31	$\overline{1.08}$	$\overline{1.35}$
					$t(4)$					
					$g(x)^{cs}$					
B	-.019	-.010	-.024	-.006	$\underline{-.017}$	$\underline{-.006}$	-.025	-.013	-.021	-.012
V	.031	.040	.022	.035	$\underline{.020}$	$\underline{.031}$	.020	.034	.022	.033
$\widehat{V}$	.036	.046	.030	.037	.027	.033	.029	.036	.027	.036
E	1.00	1.00	1. 41	1. 14	$\overline{1.55}$	$\overline{1.29}$	1. 55	1. 18	1 41	1.21
					$g_1(x)^{ms}$					
B	-.019	$\underline{-.010}$	-.026	-.011	-.023	$\underline{-.010}$	-.026	-.012	-.021	$\underline{-.010}$
V	.031	.040	.024	.038	.026	.034	.025	.037	$\underline{.022}$	$\underline{.035}$
$\widehat{V}$	.036	.046	.029	.039	.030	.037	.028	.041	.026	.038
E	1.00	1.00	1. 29	1. 21	1.19	1.08	1. 24	1. 08	$\overline{1.40}$	$\overline{1.14}$
					$g_2(x)^{ms}$					
B	$\underline{-.019}$	$\underline{-.010}$	-.030	-.015	-.027	-.014	-.031	-.015	-.025	-.013
V	.031	.040	.027	.041	.029	.039	.030	.041	$\underline{.024}$	$\underline{.036}$
$\widehat{V}$	.036	.046	.031	.044	.032	.042	.033	.043	.027	.039
E	1.00	1.00	1. 15	1. 05	1. 07	1.02	1. 03	1.00	$\overline{1.29}$	$\overline{1.11}$

GMM Efficient GMM, EL Empirical likelihood, EU Euclidean likelihood, NT Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correct and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 4a. Continued

	$\widehat{\theta}_1^c$	$\widehat{\theta}_2^c$	$\widehat{\theta}_1^{c,GMM}$	$\widehat{\theta}_2^{c,GMM}$	$\widehat{\theta}_{1w}^{c,EL}$	$\widehat{\theta}_{2w}^{c,EL}$	$\widehat{\theta}_{1w}^{c,EU}$	$\widehat{\theta}_{2w}^{c,EU}$	$\widehat{\theta}_{1w}^{c,NT}$	$\widehat{\theta}_{2w}^{c,NT}$
$\chi^2(4)-4$										
$g(x)^{cs}$										
B	-.136	-.088	-.113	-.076	-.096	-.070	-.114	-.073	-.103	-.073
V	.175	.207	.109	.133	.091	.121	.099	.129	.096	.125
$\widehat{V}$	.185	.212	.119	.141	.102	.132	.106	.136	.103	.134
E	1.00	1.00	1.60	1.55	1.92	1.71	1.77	1.60	1.82	1.66
$g_1(x)^{ms}$										
B	-.136	-.088	-.138	-.082	-.130	-.083	-.136	-.79	-.126	-.078
V	.175	.207	.130	.161	.125	.148	.118	.140	.118	.137
$\widehat{V}$	.185	.212	.140	.169	.136	.155	.131	.153	.129	.150
E	1.00	1.00	1.34	1.28	1.40	1.40	1.48	1.49	1.48	1.51
$g_2(x)^{ms}$										
B	-.136	-.088	-.142	-.087	-.146	-.90	-.140	-.86	-.135	-.085
V	.175	.207	.158	.185	.157	.190	.154	.179	.149	.174
$\widehat{V}$	.185	.212	.179	.194	.165	.200	.164	.190	.158	.188
E	1.00	1.00	1.10	1.12	1.11	1.08	1.14	1.16	1.17	1.31
$\frac{\chi^2(4)-4}{\sqrt{8}}$										
$g(x)^{cs}$										
B	.113	.057	.101	.050	.096	.045	.099	.048	.099	.049
V	.042	.064	.036	.046	.029	.038	.032	.042	.031	.040
$\widehat{V}$	.045	.070	.041	.056	.038	.042	.036	.044	.035	.046
E	1.00	1.00	1.17	1.39	1.45	1.68	1.31	1.52	1.35	1.60
$g_1(x)^{ms}$										
B	.113	.057	.106	.055	.105	.055	.107	.052	.105	.053
V	.042	.064	.040	.052	.034	.047	.035	.048	.034	.044
$\widehat{V}$	.045	.070	.045	.056	.038	.050	.039	.051	.038	.047
E	1.00	1.00	1.05	1.23	1.23	1.36	1.20	1.33	1.23	1.45
$g_2(x)^{ms}$										
B	.113	.057	.111	.058	.113	.059	.110	.057	.109	.057
V	.042	.064	.044	.056	.039	.053	.037	.055	.037	.051
$\widehat{V}$	.045	.070	.048	.062	.042	.058	.043	.059	.043	.057
E	1.00	1.00	0.95	1.14	1.40	1.20	1.13	1.16	1.13	1.25

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 4b. Finite sample bias B, variances V, $\widehat{V}$ and efficiency E of bias corrected $\widehat{\theta}^{c}$, $\widehat{\theta}^{c,GMM}$ and $\widehat{\theta}^{c,GEL}_{w}$ in binary dependent variable regression model for $n = 100$ and

	$\widehat{\theta}^{c}_{1}$	$\widehat{\theta}^{c}_{2}$	$\widehat{\theta}^{c,GMM}_{1}$	$\widehat{\theta}^{c,GMM}_{2}$	$\widehat{\theta}^{c,EL}_{1w}$	$\widehat{\theta}^{c,EL}_{2w}$	$\widehat{\theta}^{c,EU}_{1w}$	$\widehat{\theta}^{c,EU}_{2w}$	$\widehat{\theta}^{c,NT}_{1w}$	$\widehat{\theta}^{c,NT}_{2w}$
$N(0,1)$										
$g(x)^{cs}$										
B	.003	.008	.004	.010	.003	.006	.005	.008	.004	.008
V	.020	.026	.018	.017	.015	.015	.020	.018	.016	.017
$\widehat{V}$	.023	.028	.020	.019	.018	.017	.021	.020	.020	.019
E	1.00	1.00	1.11	1.52	1.33	1.73	1.00	1.44	1.18	1.53
$g_1(x)^{ms}$										
B	.003	.008	.008	.015	.007	.010	.010	.012	.007	.009
V	.020	.026	.019	.021	.019	.018	.018	.020	.017	.019
$\widehat{V}$	.023	.028	.021	.023	.020	.021	.022	.023	.020	.022
E	1.00	1.00	1.05	1.53	1.05	1.44	1.11	1.37	1.18	1.37
$g_2(x)^{ms}$										
B	.003	.008	.012	.009	.012	.008	.013	.009	.011	.009
V	.020	.026	.024	.022	.021	.024	.022	.022	.019	.021
$\widehat{V}$	.023	.028	.027	.025	.024	.023	.027	.024	.026	.023
E	1.00	1.00	.833	1.18	.952	1.08	0.91	1.18	1.05	1.24
$t(4)$										
$g(x)^{cs}$										
B	-.016	-.013	-.010	-.009	-.008	-.007	-.011	-.008	-.010	-.008
V	.023	.027	.015	.021	.013	.020	.015	.022	.015	.022
$\widehat{V}$	.025	.029	.018	.024	.016	.024	.018	.025	.017	.023
E	1.00	1.00	1.53	1.28	1.76	1.35	1.53	1.23	1.53	1.23
$g_1(x)^{ms}$										
B	-.016	-.013	-.015	-.012	-.012	-.010	-.013	-.016	-.012	-.011
V	.023	.027	.018	.024	.019	.025	.017	.025	.017	.024
$\widehat{V}$	.025	.029	.020	.025	.018	.026	.020	.028	.016	.023
E	1.00	1.00	1.27	1.12	1.21	1.08	1.35	1.08	1.35	1.12
$g_2(x)^{ms}$										
B	-.016	-.013	-.018	-.016	-.017	-.014	-.015	-.017	-.016	-.014
V	.023	.027	.021	.025	.022	.027	.022	.028	.021	.026
$\widehat{V}$	.025	.029	.024	.028	.026	.031	.025	.030	.023	.029
E	1.00	1.00	1.09	1.08	1.04	1.00	1.04	.964	1.09	1.03

GMM Efficient GMM, *EL* Empirical likelihood, *EU* Euclidean likelihood, *NT* Nonparametric tilting. $g(x)^{cs}$, $g(x)^{ms}_{j}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

Table 3. Continued

	$\widehat{\theta}_1^c$	$\widehat{\theta}_2^c$	$\widehat{\theta}_1^{c,GMM}$	$\widehat{\theta}_2^{c,GMM}$	$\widehat{\theta}_{1w}^{c,EL}$	$\widehat{\theta}_{2w}^{c,EL}$	$\widehat{\theta}_{1w}^{c,EU}$	$\widehat{\theta}_{2w}^{c,EU}$	$\widehat{\theta}_{1w}^{c,NT}$	$\widehat{\theta}_{2w}^{c,NT}$
$\chi^2(4)-4$										
$g(x)^{cs}$										
B	-.102	-.065	-.095	-.058	$\underline{-.087}$	$\underline{-.050}$	-.090	-.053	-.088	-.054
V	.121	.143	.075	.085	$\underline{.063}$	$\underline{.075}$	.070	.078	.068	.079
$\widehat{V}$	.136	.172	.092	.098	.081	.088	.086	.091	.082	.087
E	1.00	1.00	1.61	1.68	$\overline{1.92}$	$\overline{1.90}$	1.73	1.83	1.78	1.81
$g_1(x)^{ms}$										
B	-.102	-.065	-.101	-.063	-.095	$\underline{-.056}$	-.098	-.058	$\underline{-.093}$	-.057
V	.121	.143	.105	.125	.104	.116	$\underline{.101}$	.118	$\underline{.101}$	$\underline{.114}$
$\widehat{V}$	.136	.172	.125	.133	.112	.129	.113	.130	.114	.125
E	1.00	1.00	1.10	1.14	1.16	1.23	$\overline{1.20}$	1.21	$\overline{1.20}$	$\overline{1.25}$
$g_2(x)^{ms}$										
B	-.102	-.065	-.112	-.075	-.103	-.060	-.104	-.064	$\underline{-.102}$	$\underline{-.062}$
V	.121	.143	.125	.132	.118	.129	.121	.132	$\underline{.111}$	$\underline{.126}$
$\widehat{V}$	.136	.172	.139	.177	.132	.146	.142	.151	.132	.148
E	1.00	1.00	.968	1.08	1.02	1.11	1.00	1.08	$\overline{1.09}$	$\overline{1.13}$
$\frac{\chi^2(4)-4}{\sqrt{8}}$										
$g(x)^{cs}$										
B	.092	.047	.084	.043	$\underline{.083}$	$\underline{.042}$	.084	.044	$\underline{.083}$	.044
V	.024	.030	.017	.021	$\underline{.016}$	$\underline{.018}$	.017	.020	.017	$\underline{.018}$
$\widehat{V}$	.029	.034	.020	.024	.021	.022	.022	.024	.020	.023
E	1.00	1.00	1.41	1.43	$\overline{1.50}$	$\overline{1.66}$	1.41	1.50	1.50	$\overline{1.66}$
$g_1(x)^{ms}$										
B	.092	.047	.089	.048	.087	.050	$\underline{.086}$	.049	$\underline{.086}$	$\underline{.046}$
V	.024	.030	.021	.021	$\underline{.018}$	.021	.020	.023	.019	$\underline{.020}$
$\widehat{V}$	.029	.034	.025	.026	.026	.027	.026	.027	.025	.026
E	1.00	1.00	1.14	1.50	$\overline{1.33}$	1.43	1.20	1.30	1.26	$\overline{1.50}$
$g_2(x)^{ms}$										
B	.092	$\underline{.047}$	.094	.053	.095	.052	.092	.079	$\underline{.091}$	.048
V	.024	.030	.023	.026	.022	.025	.025	.029	$\underline{.021}$	$\underline{.022}$
$\widehat{V}$	.029	.034	.027	.030	.029	.029	.030	.031	.028	.027
E	1.00	1.00	1.04	1.15	1.09	1.20	.960	1.03	$\overline{1.14}$	$\overline{1.36}$

GMM Efficient GMM, EL Empirical likelihood, EU Euclidean likelihood, NT Nonparametric tilting. $g(x)^{cs}$, $g(x)_j^{ms}$ (j=1,2) indicate, respectively, correctly and moderately misspecified auxiliary information.
For each entry an underline (overline) indicates smallest (largest) value in the corresponding row.

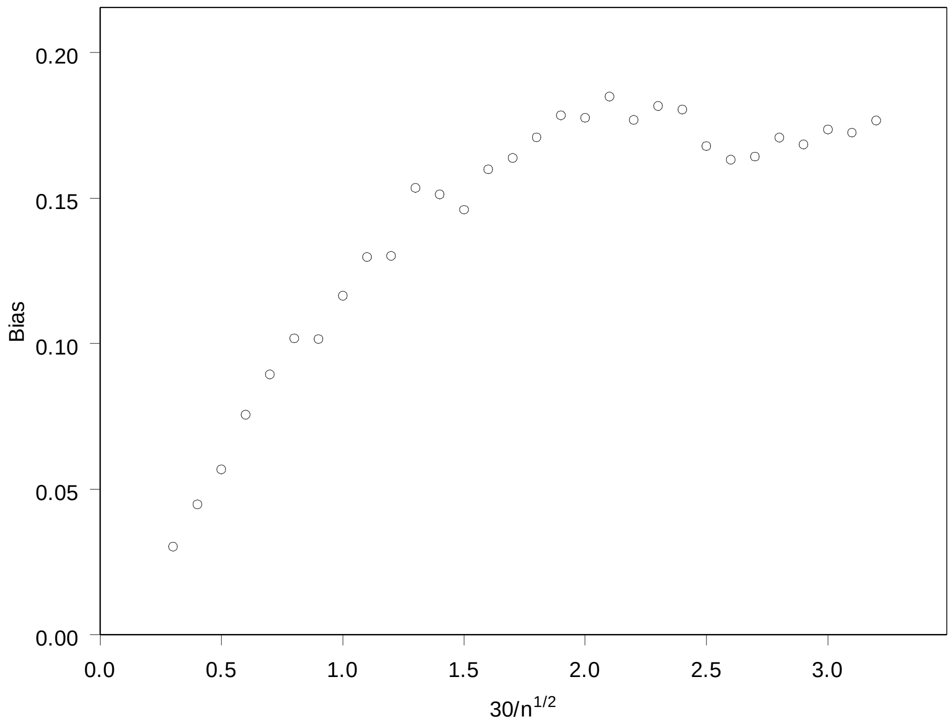

Figure 1: Finite sample bias of $\widehat{\theta}_2$ as an increasing function of the value of local misspecification.