Z-Estimators and Auxiliary Information under Weak Dependence

Federico Crudu[*][†]

University of Groningen and CRENoS

October 2009

Abstract

In this paper we introduce a weighted Z-estimator for moment condition models in the presence of auxiliary information on the unknown distribution of the data under the assumption of weak dependence. The resulting weighted estimator is shown to be consistent and asymptotically normal. Its small sample properties are checked via Monte Carlo experiments.

JEL Classification: C12, C14, C22.

Keywords: Z-estimators, M-estimators, GMM, Generalized Empirical Likelihood, blocking techniques, α-mixing.

[*]Corresponding address: Department of Economics and Econometrics, University of Groningen, Nettelbosje 2, 9747 AE, Groningen, The Netherlands. Email: f.crudu@rug.nl.

[†]This paper is based on Chapter 4 of my PhD thesis at the University of York. I am grateful to Francesco Bravo and Fabrizio Iacone for helpful comments on the previous drafts of the paper. I also wish to thank Patrick Marsh, Jan Podivinsky. All remaining errors are my own.

1 Introduction

In a fundamental paper Huber (1964) introduced a new class of estimators, called M-estimators, that represented a generalization of the classical maximum likelihood theory of estimation (M-estimator is in fact an abbreviation for Maximum-Likelihood-like estimator, e.g. Amemiya, 1985). M-estimators are defined as the result of an optimization problem (maximization or minimization is irrelevant for the final outcome) for a given criterion function and the corresponding parameter space. In some cases, the estimator is the root of the first derivative of the criterion function. Van der Vaart (2007, p. 41) recalls this difference and distinguishes between M-estimators and Z-estimators (Z stands for zero). Note that this difference already exists in Huber (1964), where an M-estimator of ρ-type is computed by means of optimization, while an M-estimator of ψ-type corresponds to the Z-estimator of van der Vaart.

Zhang (1994) extends the standard M-estimation setting to the possibility of including auxiliary information in the form of weights, which are estimated by empirical likelihood (EL). The work of Zhang is motivated by the fact that in a number of cases that arise in practice, a researcher may have prior knowledge of some features of the distribution (the population mean, or the median, for example, or other features related to the shape of the distribution), without necessarily knowing the actual distribution function underlying the data. In a more recent paper Bravo (2008) devises a two step procedure for M-estimation with auxiliary information and extends the results of Zhang (1994) to the more general class of generalized empirical likelihood (GEL) estimators.

A series of papers by Imbens and coauthors investigates the use of auxiliary information in the case of microeconometric models (see Imbens and Hellerstein, 1999, Imbens and Lancaster, 1994, Imbens, 1992). Imbens and Hellerstein (1999) for example estimate a wage regression by means of weighted least squares. The set of weights they use is based on

Census data and estimated via empirical likelihood. The EL weights shift the distribution of the primary sample towards the distribution of the Census data. When the population values of the Census distribution are of greater interest such effect is desirable (Imbens and Hellerstein, 1999). Imbens and Lancaster (1994) use macro data as auxiliary information in the context of a GMM estimator. A paper by Qian and Schmidt (1999) introduces two alternatives to the standard GMM setting, where auxiliary information is defined in terms of moment functions, independent of unknown parameters. The two estimators are asymptotically and, for an appropriate choice of the weighting matrix, numerically equivalent. In the statistical literature, similar results are for example related to the work of Kuk and Mak (1989) in the context of median estimation or to Chen and Qin (1993), who also exploit EL probabilities to carry the auxiliary information. As in the case of Qian and Schmidt (1999) our estimator may apply to rational expectation models, where the forecast error is correlated with another observable variable, which embeds the auxiliary information.

In what follows we describe a weighted Z-estimator for weakly dependent data and show consistency and asymptotic normality. The small sample properties of the estimator are investigated by means of Monte Carlo experiments on a location parameter model, where we compare different weighting schemes and estimators. Our estimator is justified by the fact that in applied research it is often possible to retrieve some auxiliary information about the otherwise unknown distribution of the data and incorporate it in the estimation algorithm in order to obtain a smaller asymptotic variance. The aforementioned weights are estimated by means of GEL from the available auxiliary information expressed in terms of moment functions (that are also independent of parameters). Since we assume weak dependence, the whole estimation procedure is framed into a blockwise structure to appropriately capture the time series features of the data (see, among others, Künsch, 1989, Kitamura, 1997, Bravo, 2005).[1] In theory, our estimator may be computationally very demanding since it requires both the estimation of the weights and the solution of the optimization program associated with the Z-estimator. However, as shown by Bravo (2008) it is possible to keep the two parts separated. The asymptotic results are the same as if we solve the two problems simultaneously.

The rest of the paper is organized as follows. In Sections 2 and 3 we outline the estimator and the main asymptotic results. In Section 4 we describe the finite sample properties of three specifications of our Z-estimator and two competing estimators. Section 5 contains some concluding remarks. Proofs and tables are relegated to the appendix.

2 Z-Estimation and Generalized Empirical Likelihood

Let $\{x_t\}$ be an $\mathbb{R}^{L_x}$-valued stationary process from an unknown distribution F, such that the following strong mixing conditions are satisfied

$$\alpha_x(k) \to 0, \ \ k \to \infty$$

where $\alpha_x(k) = \sup_{A,B} |\Pr(A \cap B) - \Pr(A)\Pr(B)|$, $A \in F^{0}_{-\infty}$, $B \in F^{\infty}_{k}$, and $F^{m''}_{m'} = \sigma(x_i : m' \leq i \leq m'')$. We also assume $\sum_{k=1}^{\infty} \alpha_x(k)^{1-\frac{1}{c}} < \infty$ for some constant $c > 1$. Consider now set of differentiable functions,

$$m(\beta) = E(m(x_t, \beta))$$

[1] A similar estimator is due to Smith (2005), who develops a kernel-based GEL estimator that incorporates the auxiliary information in the estimation algorithm.

such that $m:\mathbb{R}^{L_x}\times\mathbb{R}^{L_\beta}\rightarrow\mathbb{R}^{L_m}$, and $m(\beta_0)=0$. Moreover, $\beta_0\in int\{\mathcal{B}\}$ and $\mathcal{B}\subset\mathbb{R}^{L_\beta}$, and L_β is assumed to equal L_m. A Z-estimator for β_0, say $\hat{\beta}$, satisfies the relationship

$$\left\|\hat{m}\left(\hat{\beta}\right)\right\|=\inf_{\beta\in\mathcal{B}}\|\hat{m}(\beta)\|, \tag{1}$$

where $\hat{m}(\beta)=\frac{1}{n}\sum_{t=1}^{n}m_t(\beta)$ and $m_t(\beta)=m(x_t,\beta)$. Furthermore, we take into account the presence of weak dependence by means of the usual blockwise approach. Thus, the blockwise counterpart of (1) is

$$\left\|\hat{h}\left(\hat{\beta}\right)\right\|=\inf_{\beta\in\mathcal{B}}\left\|\hat{h}(\beta)\right\| \tag{2}$$

and similarly $\hat{h}(\beta)=\frac{1}{b}\sum_{i=1}^{b}h_i(\beta)$, $h_i(\beta)=h(z_i,\beta)$, and $h(z_i,\beta)=\frac{1}{M}\sum_{j=1}^{M}m\left(x_{(i-1)L+j},\beta\right)$, where $i=1,...,b$ and $b=\left[\frac{n-M}{L}\right]+1$

Let us assume now that there exists some auxiliary information about the unknown distribution of the data, shaped into a certain function $f:\mathbb{R}^{L_x}\rightarrow\mathbb{R}^{L_f}$ that we can define in terms of a moment condition model, independent of the unknown parameter

$$E(f_t)=0$$

for $f_t=f(x_t)$. As for equation (2), we can define its blockwise counterpart as

$$g(z_i)=\frac{1}{M}\sum_{j=1}^{M}f\left(x_{(i-1)L+j}\right). \tag{3}$$

At this stage, our problem is to find a suitable way to incorporate the auxiliary information described in (3). In order to do that we follow Bravo (2008, see also Zhang, 1995). This is, we estimate a set of probabilities by means of GEL, using the moment functions in (3). The resulting probabilities are used to weight our initial Z-estimator (2), in order to obtain a blockwise GEL (BGEL) weighted Z-estimator.

The subsequent BGEL function is

$$\hat{R}(\lambda) = \frac{1}{b}\sum_{i=1}^{b} \rho(\lambda' g_i),$$

where $g_i = g(z_i)$ $\rho(\nu)$ is the usual carrier function, concave in its domain, and normalized to be $\rho_1(0) = \rho_2(0) = -1$, where $\rho_j(\nu)$, $j = 1, 2$ is the jth derivative. The (unfeasible) BGEL probabilities are then defined as

$$\pi_i(\lambda) = \frac{\rho_1(\lambda' g_i)}{\sum_{j=1}^{b} \rho_1(\lambda' g_j)}$$

Therefore, an estimator for π_i crucially requires an estimator for λ. That is

$$\hat{\lambda} = \arg\max_{\lambda \in \Lambda_n} R(\lambda) \tag{4}$$

where Λ_n is open and contains zero.

The BGEL-weighted estimation functions are defined as

$$\hat{h}_\pi(\beta) = \sum_{i=1}^{b} \hat{\pi}_i h_i(\beta).$$

where $\hat{\pi}_i = \hat{\pi}_i\left(\hat{\lambda}\right)$ is the BGEL estimator for the probability density function as described above. Then the corresponding Z-estimator with auxiliary information, $\hat{\beta}_\pi$, implies

$$\left\|\hat{h}_\pi\left(\hat{\beta}_\pi\right)\right\| = \inf_{\beta \in \mathcal{B}} \left\|\hat{h}_\pi(\beta)\right\|.$$

In Section 3 it will be shown that the resulting estimator $\hat{\beta}_\pi$ is consistent and asymptotically

Normal, with asymptotic variance V_β:

$$V_\beta = \left(M(\beta_0)'\right)^{-1}\left(S(\beta_0) - B(\beta_0)\Sigma^{-1}B(\beta_0)'\right)\left(M(\beta_0)\right)^{-1},$$

where $M(\beta) = E(\partial m_t(\beta)/\partial\beta)$ and $B(\beta) = E\sum_{s=-\infty}^{\infty}\left(f_t m_{t-s}(\beta)'\right)$. From this expression it follows that the estimator we propose is asymptotically more efficient than an estimator that does not exploit the available auxiliary information, as its variance is $\left(M(\beta_0)'\right)^{-1}S(\beta_0)\left(M(\beta_0)\right)^{-1}$. Clearly, the efficiency of the weighted estimator depends on the relevance of the auxiliary information and, therefore, on the covariance between the original moment function m and the vector of auxiliary moments f, $B(\beta)$: thus, the larger the covariance $B(\beta)$, the smaller the resulting asymptotic variance V_β. It is also quite obvious that if the covariance is zero $\hat{\beta}$ and $\hat{\beta}_\pi$ share the same variance.

An alternative approach is due to Qian and Schmidt (1999), and it consists of constructing a moment vector that includes the extra moments

$$\hat{m}^f(\beta) = \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} m_t(\beta) \\ f_t \end{pmatrix}. \tag{5}$$

The above model is overidentified, since $L_m + L_f > L_\beta$ (notice that we assumed $L_m = L_\beta$) and the associated parameter vector may be estimated by GMM. The resulting estimator is asymptotically equivalent to our weighted Z-estimator. Notice that the standard asymptotic variance for the GMM estimator is $\left(G(\beta_0)'\Omega(\beta_0)^{-1}G(\beta_0)\right)^{-1}$. In our case $G(\beta_0) = \left(E(\partial m_t(\beta_0)/\partial\beta'), 0\right)'$, where the presence of the zeros depends on the fact that the portion of the moment vector that carries the auxiliary information is independent of the estimand parameter vector. The matrix $\Omega(\beta_0)$ is a 2×2 block matrix, whose elements on the main diagonal are $S(\beta_0)$ and Σ, and the off diagonal entry is the covariance matrix $B(\beta_0)$. After some simple algebra the result follows, and

$V_\beta = \left(G(\beta_0)' \Omega(\beta_0)^{-1} G(\beta_0)\right)^{-1}$.[2] A further method that is similar to ours is due to Smith (2004), and consists of estimating the parameters, given the augmented vector of moments in (5), by means of (smoothed) GEL. Such procedure consists of augmenting the GEL criterion function by the vector of auxiliary moments and simultaneously compute the an estimate of the parameters of interest.[3]

These two alternative approaches are equivalent to ours, in the sense that they share the same first order properties. However, there might be some practical drawbacks. The estimator of Qian and Schmidt (1999) is a GMM estimator and it may carry all the problems that affect GMM estimators in small samples. On the other hand, the estimator of Smith (2004) has better theoretical features than the GMM one, but it is computationally very demanding. The estimator we propose treats the estimation of the probabilities and of the parameter of interest separately, in order to reduce the computational complexity and exploit the desirable small sample features of the BGEL estimator.

3 Asymptotic Theory

The following theorems establish consistency and asymptotic normality of the Z-estimator with auxiliary information. Proofs follow some results of Pakes and Pollard (1989), Pakes and Linton (2001), and Bravo (2008).

The following lemma establishes consistency and asymptotic normality of the BGEL estimator of the Lagrange multiplier in (4).

Lemma 1 *Assume 1)* $\{x_t\}_{t\in\mathbb{Z}}$ *is a strictly stationary strong mixing sequence, 2)* $E\|f_t\|^{2(1+\eta)}$ *for some small enough* $\eta > 0$, $\Sigma = E(f_t f_t')$ *is positive definite, 3)*$R(\lambda) = E(\rho(\lambda' f_t))$ *has*

[2] The result in Qian and Schmidt (1999) is slightly different, since the initial vector of moments, m in our notation, is overidentified.

[3] Smith (2004) assumes that the auxiliary set of moments also depends on β, while in our case it does not. The final result is different since the asymptotic variance includes extra terms that involve the first derivatives of the auxiliary moments. However, the substance is essentially the same.

a maximum for $\lambda = 0$ *and it is unique, 4) zero is in the interior of the convex set* Λ *and* $\rho(\nu)$ *is concave and twice continuously differentiable about zero and its* j*th derivative* $\rho_j(0) = -1$*,* $j = 1, 2$*, 5)* $\hat{R}(\lambda) \to_p R(\lambda)$ *for all* $\lambda \in \Lambda$*, then* $\hat{\lambda}$ *is consistent and normally distributed*

$$\frac{\sqrt{n}}{M}\hat{\lambda} \to_d N\left(0, \Sigma^{-1}\right)$$

Theorem 1 and Theorem 2 establish consistency and asymptotic Normality for the efficient Z-estimator $\hat{\beta}_\pi$.

Theorem 1 (Consistency of $\hat{\beta}_\pi$) *Assume 1)* $\mathcal{B}$ *is a compact set, 2)* $\forall\delta > 0$ *there exists* $\varepsilon(\delta)$ *such that* $\sup_{\|\beta-\beta_0\|>\delta} \|m(\beta)\| \geq \varepsilon(\delta) > 0$*, 3)* $\sup_{\beta\in\mathcal{B}} \|\hat{m}(\beta) - m(\beta)\| = o_p(1)$*. Then, if also the assumptions in Lemma 1 are satisfied,* $\hat{\beta}_\pi \to_p \beta_0$*.*

Theorem 2 (Asymptotic Normality of $\hat{\beta}_\pi$) *Assume* $\hat{\beta}_\pi$ *is consistent; moreover, assume 1)* $m_t(\beta)$ *being continuously differentiable in a neighborhood of* β_0*,* $\mathcal{N}(\beta_0, \delta)$*, 2)* $M(\beta_0) = E(\partial m_t(\beta_0)/\partial\beta)$ *is continuous and nonsingular,* $E\left(\|m_t(\beta_0)\| \|f_t\|^2\right) < \infty$*, and* $E\sup_{\beta\in\mathcal{N}(\beta_0,\delta)}(\|\partial m_t(\beta)/\partial\beta\| \|f_t\|) < \infty$*, 3)* $\sqrt{n}\hat{h}(\beta_0) \to_d N(0, S(\beta_0))$*. Then, if assumptions in Theorem 1 are satisfied* $\sqrt{n}\left(\hat{\beta}_\pi - \beta_0\right) \to N(0, V_\beta)$*, where*

$$V_\beta = \left(M(\beta_0)'\right)^{-1}\left(S(\beta_0) - B(\beta_0)\Sigma^{-1}B(\beta_0)'\right)(M(\beta_0))^{-1},$$

and $B(\beta_0) = E\sum_{s=-\infty}^{\infty}\left(f_t m_{t-s}(\beta_0)'\right)$*.*

The following corollary is a direct result of Theorem 2. It states that an estimator of the empirical distribution function based on the BGEL probabilities is more efficient than an estimator computed as $\hat{\mu}(x) = \frac{1}{n}\sum_{t=1}^{n} 1(x_t \leq x)$.

Corollary 1 *Let* $\mu(x) = \Pr(x_t \leq x)$*. If assumptions in Theorems 1 and 2, and assumption 1 in Lemma 1 hold, then* $\hat{\mu}_\pi(z) \to_p \mu(x)$ *and* $\sqrt{n}(\hat{\mu}_\pi(z) - \mu(x)) \to_d N(0, \sigma^2 - a'\Sigma^{-1}a)$*, where* $\hat{\mu}_\pi(x)$ *is the BGEL version of* $\hat{\mu}(x) = \frac{1}{n}\sum_{t=1}^{n} 1(x_t \leq x)$*, that is* $\hat{\mu}_\pi(z) = \sum_{i=1}^{b}\hat{\pi}_i 1_M(z_i \leq z)$*.*

Proofs are in the appendix.

4 Monte Carlo Experiments

In this section we study the small sample features of our weighted Z-estimator by means of a location parameter model. Given the auxiliary information $w_i = \sum_{j=1}^{M} u_{(i-1)L+j}$, the three BGEL estimators for the probabilities are defined as

$$\begin{aligned}
\hat{\pi}_i^{EL} &= \frac{1}{b\left(1+\hat{\lambda}^{EL} w_i\right)} \\
\hat{\pi}_i^{ET} &= \frac{\exp\left(\hat{\lambda}^{ET} w_i\right)}{\sum_{j=1}^{b} \exp\left(\hat{\lambda}^{ET} w_j\right)} \\
\hat{\pi}_i^{EU} &= \frac{1}{b}\left(1-\hat{\lambda}^{EU}\left(w_i-\bar{w}\right)\right)
\end{aligned}$$

where $\bar{w} = \frac{1}{b}\sum_{i=1}^{b} w_i$. $\hat{\lambda}^{EL}$ and $\hat{\lambda}^{ET}$ are computed numerically, while it is available a close form solution for $\hat{\lambda}^{EU}$. The measures we use are both bias and mean square error (MSE).

Let us consider the estimation of a location parameter as in Qian and Schmidt (1999)

$$y_t = \beta_0 + e_t$$

where β_0 is a scalar, and it is assumed to be equal to 1, and e_t is a zero mean disturbance. Thus, we want to find an estimate for $\beta_0 = E\left(y_t\right)$. We also assume there exists a certain random variable u_t, that is known to have zero mean and it is correlated with e_t. We define then the following equations:

$$
\begin{aligned}
y_t &= 1 + e_t \\
u_t &= \rho e_t + \sqrt{1-\rho^2}\eta_t
\end{aligned}
$$

where $\rho = 0.4, 0.8$. Moreover, we choose five specifications for the processes e_t and η_t:

DGP 1 $e_t = \alpha e_{t-1} + \varepsilon_t^e$, η_t is normally distributed with zero mean and unit variance

DGP 2 $e_t = \alpha \varepsilon_{t-1}^e + \varepsilon_t^e$, η_t is normally distributed with zero mean and unit variance

DGP 3 $e_t = \alpha e_{t-1} + \varepsilon_t^e$, $\eta_t = \alpha \varepsilon_{t-1}^\eta + \varepsilon_t^\eta$

DGP 4 $e_t = \alpha \varepsilon_{t-1}^e + \varepsilon_t^e$, $\eta_t = \alpha \eta_{t-1} + \varepsilon_t^\eta$

DGP 5 $e_t = \alpha e_{t-1} + \varepsilon_t^e$, $\eta_t = \alpha \eta_{t-1} + \varepsilon_t^\eta$.

Where $\varepsilon_t^i \sim N(0,1)$, $i = e, \eta$. The parameter that characterizes the AR process and the MA process is the same and we name it α, and takes values $\alpha = .4, .8$.

We compute an estimate for β_0 in five different ways. The first is a simple sample mean

$$
\hat{\beta} = \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t.
$$

The second is an efficient GMM estimator with two moment functions

$$
\hat{\beta} = \arg\min_{\beta} \hat{g}(\beta)' \hat{\Omega}\left(\bar{\beta}\right)^{-1} \hat{g}(\beta)
$$

where $\hat{g}(\beta) = \frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} y_t - \beta, & u_t \end{pmatrix}'$. The matrix of weights $\hat{\Omega}\left(\bar{\beta}\right)$ is a Newey-West matrix evaluated at a certain consistent estimator of β, $\bar{\beta}$. The remaining three estimators are weighted averages based on GEL estimators, i.e. the EL, the ET and the EU estimator,

$$\hat{\beta} = \sum_{i=1}^{b} \hat{\pi}_i^{GEL} z_i$$

where $z_i = \sum_{j=1}^{M} y_{(i-1)L+j}$.

Our simulations suggest that in most of the cases, weighted estimators deliver smaller MSEs than sample mean and also than efficient GMM estimator. The bias of the GEL estimators is almost always negative, and, specially for the smaller sample size ($n = 256$), tends to be larger, in absolute terms, than the competing estimators. On the other hand, weighted averages considerably reduce the MSE in most of the case we analyze. Some cases, though, do not deliver the expected results. In Table 4 for $\alpha = .4$ we see that auxiliary information is not capable of improve efficiency and our weighted averages are at best as good as the corresponding result for the sample mean. Similarly, in Table 4 for $\alpha = .4$ weighted averages improve the sample mean, but the GMM estimator remains a better alternative. We notice also that the choice of the block structure has a negligible effect on the estimates that derive from the second DGP, and also for the third DGP, when α is set to .4. For the fifth DGP, with $\alpha = .8$ the choice of the block structure is relevant. The combination that minimizes the MSE is $M = L = 4$. The effect of the block structure tends to disappear when $\rho = .8$ and as the sample size grows.

Let us analyze more in detail each case.

DGP 1: the results for this DGP aresummarized in Table 1. The weighted BGEL estimator is better than the two benchmarks (sample mean and GMM estimator) in the sense that have smaller MSE, and in a number of cases they display also a smaller bias. For either sample sizes the weighted means improve in terms of the MSE the sample mean and the GMM estimator, even though in the latter case the difference is small. On the other hand, the BGEL-based estimators tend to suffer from larger biases, when $n = 256$.

For $\alpha = .8$ the data are more persistent and we expect our results to worsen. Interestingly, we notice that for $\alpha = \rho = .8$ the results in terms of MSE are similar to the case with $\alpha = .4$ and $\rho = .8$, suggesting that the efficient estimator is to some extent indifferent to the presence of persistence in the data. In addition, if on the one hand the effect of the persistence parameter is milder when ρ is higher, on the other, for $\rho = \alpha = .4$ the GMM estimator does not benefit from that, and its MSE is about twice the size of the weighted estimators.

DGP 2: the results for this DGP aresummarized in Table 2. For $\rho = \alpha = .8$ the results of the improved estimators (GMM and BGEL-based) are quite similar, and display a smaller MSE with respect to the simple sample mean. In several cases the bias of the BGEL-based estimators is larger than the bias of the GMM estimator, and for $n = 256$ is larger also than the bias of the sample mean. For $\rho = .8$ the result indicate a reduction of the MSE for the weighted estimators and the GMM estimator. On the other hand, also in this case the bias of the weighted estimators is larger than the bias for the sample mean. Similar results are found for $\alpha = .8$. For $\alpha = .8$ and $\rho = .4$ the BGEL estimator is slightly better. However, they both approach the same value for $\rho = .8$, for either sample size.

DGP 3: the results for this DGP aresummarized in Table 3. By comparing the results in Table 1 and Table 3 for $\rho = \alpha = .8$, and for $\rho = .4$ and $\alpha = .8$ we notice that the introduction of a MA process has a detrimental effect on the size of the MSE for the improved estimators, even though they are still smaller than the MSE of the sample mean. However, differently for the efficiently weighted case, we notice that the MSE of the GMM estimator is particularly high in the presence of low correlation ($\rho = .4$) for $\alpha = .8$.

DGP 4: the results in Table 4 are quite interesting since in some cases the improved estimators provide only small improvements over the GMM estimator (although the weighted estimators tend to display a smaller bias overall). An increase in persistence ($\alpha = .8$) higlights how the efficiently weigthed estimators display a lower MSE with respect to its

GMM counterpart.

DGP 5: the results summarized in Table 5 describe another interesting scenario: the improved estimators have smaller MSEs than the sample mean, but the MSE for the BGEL-based estimators are at most equal to the MSE for the GMM estimator.

5 Conclusion

In this chapter we propose a two step procedure for Z-estimators in the presence of weakly dependent data and auxiliary information based on the estimation of BGEL probabilities. This procedure is attractive from different points of view. First of all, the computation of the BGEL probabilities is very simple, as it contemplates only the convex part of the problem (this is the estimation of the Lagrange multiplier λ), contrarily to the improved GEL estimator of Smith (2004), where the auxiliary information is included in the saddle point problem. Moreover, whenever the Z-estimator is equivalent to a GMM estimator (Qian and Schmidt, 1999), it does not entail the well-known small sample effects that stem from the estimation of the efficient weight matrix (see for example Altonji and Segal, 1996). Our asymptotic results state that the resulting Z-estimator is consistent and Normally distributed. The resulting variance depends on the relevance of the auxiliary information. In addition, we demonstrate that the estimator of a distribution based on the BGEL weights enjoys the same favourable features of the abovementioned Z-estimator. Moreover, by means of Monte Carlo experiments, we describe how to apply our approach to some standard time series problems. In fact, the strand of literature we refer to (Imbens and Hellerstein 1994, and related papers) focusses exclusively on micro data. Qian and Schmidt (1999) apply their GMM estimators (Improved or Augmented) to a type of setting that may imply the presence of weakly dependent data, even though they develop their results under the *iid* assumption. We argue that a BGEL-weighted Z-estimator is desirable since it may provide improvements over unweighted or, more generally, alternative techniques. The small sample properties of the weighted Z-estimator are assessed by means of Monte Carlo experiments. The laboratory we set is a location parameter estimation, similar to what is described in Qian and Schmidt (1999, see also Zhang, 1994). We compare three BGEL weighted estimators against a simple sample mean and an augmented GMM estimator. The examples we consider define a range of possible dependence structures. The results that follow clearly indicate a preference for the weighted estimator, which turns out to have, generally, a smaller MSE. The corresponding bias is mostly negative and quite similar to the bias of the benchmark estimators.

References

[1] Altonji, J. G., L. M. Segal (1996): Small-sample bias in GMM estimation of covariance structures, *Journal of Business and Economic Statistics*, **14**, 353-366

[2] Anatolyev, S., (2005): GMM, GEL, serial correlation, and asymptotic bias, *Econometrica*, **73**, 983-1002

[3] Back, K, D. Brown (1993): Implied probabilities in GMM estimation, *Econometrica*, **61**, 971-975.

[4] Bravo, F. (2005): Blockwise empirical entropy tests for time series regressions. *Journal of Time Series Analysis*, **26**, 185-210.

[5] Bravo, F. (2008): Efficient M-estimators with auxiliary information, University of York Working Paper.

[6] Chamberlain, G. (1987): Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics*, **34**, 305-334.

[7] Chen, J. J. Qin (1993): Empirical likelihood estimation for finite populations and the effective usage of auxiliary information, *Biometrika*, **80**, 107-116.

[8] Crudu, F. (2009): GMM, Generalized Empirical Likelihood, and Time Series, *submitted.*

[9] Fitzenberger, B. (1997), The moving blocks bootstrap and robust inference for linear least squares and quantile regression, *Journal of Econometrics*, **82**, 235-287.

[10] Hansen, L. P. (1982): Large sample properties of generalized method of moments estimator, *Econometrica*, **50**, 1029-1054.

[11] Hansen, L. P., J. Heaton, A. Yaron (1996): Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics*, **14**, 262-280.

[12] Hellerstein, J. G. W. Imbens (1999): Imposing moment restrictions from auxiliary data by weighting, *Review of Economics and Statistics*, **81**, 1-14.

[13] Huber, P. (1964): Robust estimation of a location parameter, *Annals of Mathematical Statistics*, **18**, 121-140.

[14] Huber, P. (1973): Robust regression: asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, **1**, 799-821.

[15] Ibragimov, I. A., Y. V. Linnik (1971): *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

[16] Imbens, G. W. (1992): An efficient method of moments estimator for discrete choice models with choice-based sampling, *Econometrica*, **60**, 1187-1214.

[17] Imbens, G. W. (1997): One-step estimators for over-identified generalized method of moments models, *Review of Economic Studies*, **64**, 359-383.

[18] Imbens, G. W., T. Lancaster (1994): Combining micro and macro data in microeconometric models, *Review of Economic Studies*, **61**, 655-680.

[19] Imbens, G. W., R. H. Spady, P. Johnson (1998): Information theoretic approaches to inference in moment condition, *Econometrica*, **66**, 333-357.

[20] Kitamura, Y. (1997): Empirical likelihood methods with weakly dependent processes, *The Annals of Statistics*, **25**, 2084-2102.

[21] Kuk, A. Y. C., T. K. Mak (1989): Median estimation in the presence of auxiliary information, *Journal of the Royal Statistical Society B*, **51**, 261-269.

[22] Künsch, H. R. (1989): The jackknife and the bootstrap for general stationary observations, *The Annals of Statistics*, **17**, 1217-1241.

[23] Newey, W. K., D. McFadden (1994): Large sample estimation and hypothesis testing, in *Handbook of Econometrics vol. IV*, ed. R. Engle and D. McFadden. North Holland.

[24] Newey, W. K., R. J. Smith (2004): Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, **72**, 219-255.

[25] Newey, W. K., K. West (1987): A simple positive semidefinite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703-708.

[26] Owen, A. B. (2001): *Empirical Likelihood*, Chapman-Hall.

[27] Pakes, A., O. Linton (2001): *Nonlinear Methods for Econometrics*, LSE lecture notes http://econ.lse.ac.uk/staff/olinton/ec481/aet.pdf.

[28] Pakes, A., D. Pollard (1989): Simulation and the asymptotics of optimization estimators, *Econometrica*, **57**, 1027-1057.

[29] Qian, H., P. Schmidt (1999): Improved instrumental variables and generalized method of moments estimators, *Journal of Econometrics*, **91**, 145-169.

[30] Qin, J., J. Lawless (1994): Empirical likelihood and general estimating equations, *The Annals of Statistics*, **22**, 300-325.

[31] Smith, R. J. (1997): Alternative semi-parametric likelihood approaches to generalised method of moments estimation, *The Economic Journal*, **107**, 503-519.

[32] Smith, R. J. (2004): GEL criteria for moment condition models, cemmap Working Paper.

[33] Van der Vaart, A. (2007): *Asymptotic Statistics*, CUP.

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

[35] Zhang, B. (1999): Bootstrapping with auxiliary information, *The Canadian Journal of Statistics*, **27**, 237-249.

6 Appendix: Proofs and Tables

In what follows we present the proofs of the theorems presented in Section 3 and some auxiliary results. In addition, we use the following notation: $\rightarrow_p$ and $\rightarrow_d$ denote convergence in probability and convergence in distribution; C is a generic positive constant; CS and T denote Cauchy-Schwarz inequality and triangular inequality respectively; $\|\cdot\|$ is the Euclidean norm of $\cdot$. The CLT is meant to be a CLT for strong mixing sequences (see e.g. Ibragimov and Linnik, 1971).

Proof of Lemma 1. Consider again

$$\hat{R}(\lambda) = \frac{1}{b}\sum_{i=1}^{b} \rho(\lambda' g_i).$$

$\hat{R}(\lambda)$ is concave through $\rho(\cdot)$, since it is a linear combination of concave functions, namely. Moreover, assumptions 2 to 4 match assumptions (i)-(iii) from Theorem 2.7 of Newey and McFadden (1994). Then, consistency of $\hat{\lambda}$ follows.

Consider now a mean value expansion of the first order conditions of the GEL criterion function,

$$\begin{aligned} 0 &= \frac{\partial \hat{R}\left(\hat{\lambda}\right)}{\partial \lambda} = \frac{1}{b}\sum_{i=1}^{b} \rho_1\left(\dot{\lambda}' g_i\right) g_i \\ &= -\bar{g} + \frac{\hat{\lambda}'}{M}\frac{M}{b}\sum_{i=1}^{b} \rho_1\left(\dot{\lambda}' g_i\right) g_i g_i' \end{aligned}$$

Since $\hat{\lambda}$ is consistent and $\left\|\dot{\lambda}\right\| \leq \left\|\hat{\lambda}\right\|$, we have that $\rho_1\left(\dot{\lambda}' g_i\right) = -1 + o_p(1)$. Thus, multiplying by $\sqrt{n}$

$$0 = -\sqrt{n}\bar{g} - \sqrt{n}\frac{\hat{\lambda}'}{M}\hat{\Sigma} - \sqrt{n}\frac{\hat{\lambda}'}{M}\hat{\Sigma} o_p(1)$$

where $\hat{\Sigma} = M\sum_{i=1}^{b} g_i g_i'/b$ and $\bar{g} = \sum_{i=1}^{b} g_i/b$. Notice that $\sqrt{n}\frac{\hat{\lambda}'}{M}\hat{\Sigma} = O_p(1)$; therefore, by rearranging

$$\sqrt{n}\frac{\hat{\lambda}}{M} = -\hat{\Sigma}^{-1}\sqrt{n}\bar{g} + o_p(1). \tag{6}$$

Finally, by applying CLT to $\sqrt{n}\bar{g}$ and Slutsky theorem, the result follows. ■

Proof of Theorem 1 (Consistency of $\hat{\beta}_\pi$). Let us compute a mean value expansion of

$$\hat{\pi}_i = \frac{\rho_1\left(\hat{\lambda}' g_i\right)}{\sum_j \rho_1\left(\hat{\lambda}' g_j\right)}$$

about $\lambda = 0$, where $\hat{\lambda}$ is a consistent estimator for λ:

$$\begin{aligned}\hat{\pi}_i &= \frac{1}{b} + \frac{1}{b}\left(\frac{\rho_2\left(\dot{\lambda}' g_i\right) g_i'}{\frac{1}{b}\sum_j \rho_1\left(\dot{\lambda}' g_j\right)} - \frac{\rho_1\left(\hat{\lambda}' g_i\right) \frac{1}{b}\sum_j \rho_2\left(\dot{\lambda}' g_j\right) g_j'}{\left(\frac{1}{b}\sum_j \rho_1\left(\dot{\lambda}' g_j\right)\right)^2}\right)\left(\hat{\lambda} - 0\right) \\ &= \frac{1}{b} + \frac{1}{b}\left(\frac{\rho_2\left(\dot{\lambda}' g_i\right) \hat{\lambda}' g_i}{\frac{1}{b}\sum_j \rho_1\left(\dot{\lambda}' g_j\right)} - \frac{\rho_1\left(\hat{\lambda}' g_i\right) \frac{1}{b}\sum_j \rho_2\left(\dot{\lambda}' g_j\right) \hat{\lambda}' g_j}{\left(\frac{1}{b}\sum_j \rho_1\left(\dot{\lambda}' g_j\right)\right)^2}\right).\end{aligned}$$

From results of Lemma 1 we obtain

$$\hat{\pi}_i = \frac{1}{b} + \frac{1}{b}\left(\hat{\lambda}' g_i + o_p(1)\right) \tag{7}$$

and

$$\hat{\pi}_i = \frac{1}{b}\left(1 + o_p(1)\right). \tag{8}$$

From Lemma 1 in Crudu (2009) we have $\hat{h}_\pi(\beta) = \hat{m}(\beta) + O_p(M/n)$. Then, by adding and subtracting $\hat{h}_\pi\left(\hat{\beta}_\pi\right)$ and T

$$\left\|m\left(\hat{\beta}_\pi\right)\right\| \leq \left\|m\left(\hat{\beta}_\pi\right) - \hat{h}_\pi\left(\hat{\beta}_\pi\right)\right\| + \left\|\hat{h}_\pi\left(\hat{\beta}_\pi\right)\right\|$$

Moreover, by optimality of $\hat{\beta}_\pi$ and since $m(\beta_0) = 0$, and by repeated application of Lemma 1 in Crudu (2009) and T

$$\begin{aligned}\left\|m\left(\hat{\beta}_\pi\right)\right\| &\leq \left\|m\left(\hat{\beta}_\pi\right) - \hat{m}\left(\hat{\beta}_\pi\right)\right\| + (1 + o_p(1))\left\|\hat{m}(\beta_0) - m(\beta_0)\right\| + O_p\left(\frac{M}{n}\right) \\ &\leq \sup_{\beta\in\mathcal{B}}\left\|m(\beta) - \hat{m}(\beta)\right\| + (1 + o_p(1))\sup_{\beta\in\mathcal{B}}\left\|\hat{m}(\beta) - m(\beta)\right\| + O_p\left(\frac{M}{n}\right).\end{aligned}$$

By Assumption 3 $\sup_{\beta\in\mathcal{B}}\left\|m(\beta) - \hat{m}(\beta)\right\| = o_p(1)$; hence

$$\left\|m\left(\hat{\beta}_\pi\right)\right\| \leq o_p(1).$$

Since $m(\beta)$ is bounded away from zero for $\|\beta - \beta_0\| > \delta$ (assumption 2), it follows that $\hat{\beta}_\pi \in \|\beta - \beta_0\| < \delta$. As δ is arbitrary, $\hat{\beta}_\pi \rightarrow_p \beta_0$. ■

Proof of Theorem 2 (Asymptotic Normality of $\hat{\beta}_\pi$). Let us consider $\sum_i \hat{\pi}_i h_i\left(\hat{\beta}_\pi\right) = 0$, by replacing the probabilities with the expression in 7

$$0 = \frac{1}{b}\sum_i \left(1 + \hat{\lambda}' g_i + o_p(1)\right) h_i\left(\hat{\beta}_\pi\right)$$

and mean value expand $h_i\left(\hat{\beta}_\pi\right)$ about β_0, for $\hat{\beta}_\pi$ being consistent

$$\begin{aligned} 0 &= \sum_i \left(1 + \hat{\lambda}' g_i + o_p(1)\right) h_i\left(\hat{\beta}_\pi\right) \\ &= \sum_i \left(1 + \hat{\lambda}' g_i\right)\left(h_i(\beta_0) + \frac{\partial h_i\left(\dot{\beta}\right)}{\partial \beta}\left(\hat{\beta}_\pi - \beta_0\right)\right) + o_p(1)\,\hat{h}\left(\hat{\beta}_\pi\right) \end{aligned}$$

where $\left\|\dot{\beta} - \beta_0\right\| \leq \left\|\hat{\beta}_\pi - \beta_0\right\|$. Let us define $\hat{B}(\beta) = M \sum_i g_i h_i(\beta)/b$. Then, by appropriate rescaling and (6)

$$\begin{aligned} 0 &= \sqrt{n}\hat{h}(\beta_0) + \hat{\Sigma}^{-1}\sqrt{n}\bar{g}'\hat{B}(\beta_0) \\ &\quad + \left(\frac{1}{b}\sum_i \frac{\partial h_i\left(\dot{\beta}\right)}{\partial \beta} + \hat{\lambda}' \sum_i g_i \frac{\partial h_i(\beta_0)}{\partial \beta}/b\right)\sqrt{n}\left(\hat{\beta}_\pi - \beta_0\right) \\ &\quad + o_p(1)\sqrt{n}\hat{h}\left(\hat{\beta}_\pi\right) \\ &= A_1 + A_2 + A_3 \end{aligned}$$

where

$$A_1 = \sqrt{n}\hat{h}(\beta_0) + \hat{\Sigma}^{-1}\sqrt{n}\bar{g}'\hat{B}(\beta_0),$$

$$A_2 = \left(\sum_i \partial h_i\left(\dot{\beta}\right)/\partial\beta + \hat{\lambda}' \sum_i g_i \partial h_i(\beta_0)/\partial\beta/b\right)\sqrt{n}\left(\hat{\beta}_\pi - \beta_0\right),$$

$$A_3 = o_p(1)\sqrt{n}\hat{h}\left(\hat{\beta}_\pi\right).$$

From assumption 3 and Lemma 1 $\sqrt{n}\ \hat{h}(\beta_0) \to_d N(0, S(\beta_0))$ and $\sqrt{n}\bar{g} \to_d N(0, \Sigma)$.Then, after simple calculations, we get $A_1 \to_d N(0, W)$, where

$$\begin{aligned} W &= \begin{pmatrix} I, & -B(\beta_0)'\Sigma^{-1} \end{pmatrix} \begin{pmatrix} S(\beta_0) & B(\beta_0)' \\ B(\beta_0) & \Sigma \end{pmatrix} \begin{pmatrix} I \\ -\Sigma^{-1}B(\beta_0) \end{pmatrix} \\ &= S(\beta_0) - B(\beta_0)'\Sigma^{-1}B(\beta_0) \end{aligned}$$

Let us now focus attention on A_2. By Lemma 1 in Crudu (2009) and T

$$\begin{aligned} \left\| \hat{\lambda}'\frac{1}{b}\sum_i g_i \frac{\partial h_i\left(\dot{\beta}\right)}{\partial \beta'} \right\| &\le \left\|\hat{\lambda}\right\| \left\| \frac{1}{n}\sum_t f_t \frac{\partial m_t\left(\dot{\beta}\right)}{\partial \beta'} + O_p\left(\frac{M}{n}\right) \right\| \\ &\le \left\|\hat{\lambda}\right\| \frac{1}{n}\sum_t \sup_{\beta\in\mathcal{B}} \left\| f_t \frac{\partial m_t(\beta)}{\partial \beta'} \right\| + o_p(1). \end{aligned}$$

Thus,

$$\left\| \hat{\lambda}'\frac{1}{b}\sum_i g_i \frac{\partial h_i\left(\dot{\beta}\right)}{\partial \beta} \right\| \le o_p(1).$$

By CMT and assumption 3 $\sqrt{n}\hat{h}\left(\hat{\beta}_\pi\right)$ is Normally distributed. Thus, its order of magnitude is $O_p(1)$ and $A_3 = o_p(1)$. Finally,

$$M(\beta_0)'\sqrt{n}\left(\hat{\beta}_\pi - \beta_0\right) + o_p(1) = -\begin{pmatrix} I, & -\hat{B}(\beta_0)\Sigma^{-1} \end{pmatrix} \begin{pmatrix} \sqrt{n}\hat{h}(\beta_0) \\ \sqrt{n}\bar{g} \end{pmatrix} + o_p(1)$$

then

$$
\begin{aligned}
M(\beta_0)' \sqrt{n}\left(\hat{\beta}_\pi - \beta_0\right) &= -\left(\begin{array}{cc} I, & -\hat{B}(\beta_0)\Sigma^{-1} \end{array} \right) \\
&\quad \times \left(\begin{array}{c} \sqrt{n}\hat{h}(\beta_0) \\ \sqrt{n}\bar{g} \end{array} \right) + o_p(1)
\end{aligned}
$$

and

$$
\sqrt{n}\left(\hat{\beta}_\pi - \beta_0\right) = -M(\beta_0)' \left(\begin{array}{cc} I, & -\hat{B}(\beta_0)\Sigma^{-1} \end{array} \right) \left(\begin{array}{c} \sqrt{n}\hat{h}(\beta_0) \\ \sqrt{n}\bar{g} \end{array} \right) + o_p(1)
$$

which implies, by CLT applied to $\sqrt{n}\bar{g}$, assumption 3 and CMT,

$$
\sqrt{n}\left(\hat{\beta}_\pi - \beta_0\right) \to_d N\left(0, \left(M(\beta_0)'\right)^{-1} \left(S(\beta_0) - B(\beta_0)\Sigma^{-1}B(\beta_0)'\right)(M(\beta_0))^{-1}\right).
$$

■

Proof of Corollary 1. From results in Lemma 1 and Theorem 1 we have

$$
\begin{aligned}
\hat{\mu}_\pi(z) &= \frac{1}{b}\sum_{t=1}^{b} 1_M(z_i \le z)\left(1 + \hat{\lambda}' g_i + o_p(1)\right) \\
&\quad \hat{\mu}_b(z) - \bar{g}'\hat{\Sigma}^{-1}\frac{M}{\sqrt{n}b}\sum_{t=1}^{b} g_i 1_M(z_i \le z) + o_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

Then, by adding and subtracting $\mu(x)$ and multiplying both sides by $\sqrt{n}$, we get

$$
\begin{aligned}
\sqrt{n}\left(\hat{\mu}_\pi(z) - \mu(x)\right) &= \sqrt{n}\left(\hat{\mu}_b(z) - \mu(x)\right) - \sqrt{n}\bar{g}'\hat{\Sigma}\frac{1}{b}\sum_{t=1}^{b} g_i 1_M(z_i \le z) + o_p(1) \\
&= \left(\begin{array}{cc} 1, & -\hat{a}' \end{array} \right) \left(\begin{array}{c} \sqrt{n}\left(\hat{\mu}_b(z) - \mu(x)\right) \\ \sqrt{n}\bar{g} \end{array} \right) \\
&\to {}_d N\left(0, \sigma^2 - a'\Sigma^{-1}a\right).
\end{aligned}
$$

The result follows by CLT applied to $\sqrt{n}\left(\hat{\mu}_b(z) - \mu(x)\right)$ and $\sqrt{n}\bar{g}$ and Slutsky theorem.

■

	$\bar{y}$		GMM			EL		ET		EU	
	$BIAS$	MSE	$BIAS$	MSE	M,L	$BIAS$	MSE	$BIAS$	MSE	$BIAS$	MSE
$n = 256$											
$\rho = .4,\ \alpha = .4$	-.0012	.0960	-.0014	.0069	$8,1$	-.0018	.0067	-.0018	.0066	-.0018	.0066
					$8,8$	-.0023	.0067	-.0022	.0066	-.0021	.0066
					$4,1$	-.0018	.0065	-.0018	.0065	-.0018	.0065
					$4,4$	-.0023	.0067	-.0022	.0066	-.0021	.0066
$\rho = .8,\ \alpha = .4$	-.0012	.0096	-.0012	.0019	$8,1$	-.0015	.0018	-.0015	.0018	-.0013	.0018
					$8,8$	-.0018	.0018	-.0018	.0018	-.0016	.0018
					$4,1$	-.0014	.0018	-.0014	.0018	-.0013	.0018
					$4,4$	-.0014	.0018	-.0014	.0018	-.0014	.0018
$\rho = .4, \alpha = .8$	-.0034	.0853	-.0038	.0369	$8,1$	-.0047	.0180	-.0046	.0180	-.0040	.0190
					$8,8$	-.0049	.0180	-.0046	.0180	-.0047	.0190
					$4,1$	-.0041	.0209	-.0040	.0208	-.0037	.0211
					$4,4$	-.0044	.0208	-.0044	.0207	-.0044	.0211
$\rho = .8, \alpha = .8$	-.0034	.0853	-.0021	.0039	$8,1$	-.0022	.0023	-.0021	.0023	-.0023	.0027
					$8,8$	-.0019	.0023	-.0019	.0022	-.0030	.0028
					$4,1$	-.0457	.0463	-.0456	.0463	-.0468	.0464
					$4,4$	-.0021	.0023	-.0022	.0023	-.0025	.0025
$n = 512$											
$\rho = .4,\ \alpha = .4$	.0007	.0054	-.0002	.0040	$8,1$	-.0006	.0037	-.0005	.0037	-.0005	.0037
					$8,8$	-.0008	.0037	-.0008	.0037	-.0008	.0037
					$4,1$	-.0004	.0037	-.0004	.0037	-.0004	.0037
					$4,4$	-.0007	.0037	-.0006	.0037	-.0006	.0037
$\rho = .8,\ \alpha = .4$	-.0007	.0054	-.0007	.0011	$8,1$	-.0009	.0010	-.0008	.0010	-.0008	.0010
					$8,8$	-.0009	.0010	-.0008	.0010	-.0008	.0010
					$4,1$	-.0008	.0010	-.0007	.0010	-.0008	.0010
					$4,4$	-.0008	.0010	-.0005	.0010	-.0008	.0010
$\rho = .4, \alpha = .8$	.0031	.0471	.0007	.0212	$8,1$	-.0018	.0095	-.0009	.0104	-.0017	.0096
					$8,8$	-.0020	.0094	.0001	.0095	-.0017	.0095
					$4,1$	-.0009	.0113	-.0005	.0120	-.0008	.0113
					$4,4$	-.0012	.0112	.0003	.0118	-.0010	.0112
$\rho = .8, \alpha = .8$	.0031	.0471	-.0004	.0028	$8,1$	-.0011	.0011	-.0011	.0017	-.0010	.0012
					$8,8$	-.0012	.0011	-.0045	.0015	-.0010	.0012
					$4,1$	-.0011	.0012	-.0006	.0017	-.0010	.0012
					$4,4$	-.0011	.0012	-.0011	.0012	-.0011	.0012

Table 1: DGP 1, the error term associated to the variable of interest follows a AR(1) process, while the error term associated to the auxiliary variable is iid Normal.

	$\bar{y}$		GMM			EL		ET		EU	
	$BIAS$	MSE	$BIAS$	MSE	M,L	$BIAS$	MSE	$BIAS$	MSE	$BIAS$	MSE
$n = 256$											
$\rho = .4,\ \alpha = .4$	.0003	.0069	-.0001	.0052	8, 1	-.0009	.0053	-.0009	.0053	-.0009	.0053
					8, 8	-.0001	.0053	.0000	.0052	.0001	.0052
					4, 1	-.0004	.0052	-.0004	.0052	-.0004	.0052
					4, 4	-.0004	.0051	-.0004	.0051	-.0004	.0051
$\rho = .8,\ \alpha = .4$	.0003	.0069	-.0003	.0017	8, 1	-.0011	.0017	-.0010	.0017	-.0009	.0017
					8, 8	-.0008	.0017	-.0008	.0017	-.0006	.0017
					4, 1	-.0006	.0017	-.0006	.0017	-.0006	.0017
					4, 4	-.0004	.0017	-.0004	.0017	-.0004	.0017
$\rho = .4, \alpha = .8$	.0004	.0115	-.0001	.0077	8, 1	-.0014	.0076	-.0014	.0075	-.0013	.0075
					8, 8	-.0004	.0075	-.0003	.0074	-.0002	.0074
					4, 1	-.0007	.0074	-.0007	.0074	-.0007	.0074
					4, 4	-.0007	.0073	-.0007	.0073	-.0007	.0073
$\rho = .8, \alpha = .8$	.0004	.0115	-.0003	.0019	8, 1	-.0010	.0019	-.0011	.0019	-.0011	.0019
					8, 8	-.0011	.0019	-.0003	.0019	-.0007	.0019
					4, 1	-.0007	.0019	-.0007	.0019	-.0007	.0019
					4, 4	-.0004	.0019	-.0005	.0019	-.0004	.0019
$n = 512$											
$\rho = .4,\ \alpha = .4$	.0006	.0038	-.0002	.0029	8, 1	-.0004	.0029	-.0003	.0029	-.0003	.0029
					8, 8	-.0005	.0028	-.0005	.0028	-.0005	.0028
					4, 1	-.0003	.0028	-.0003	.0028	-.0003	.0028
					4, 4	-.0005	.0028	-.0005	.0028	-.0005	.0028
$\rho = .8,\ \alpha = .4$	.0006	.0038	-.0007	.0009	8, 1	-.0007	.0009	-.0006	.0009	-.0006	.0009
					8, 8	-.0007	.0009	-.0007	.0009	-.0007	.0009
					4, 1	-.0007	.0009	-.0007	.0009	-.0007	.0009
					4, 4	-.0008	.0009	-.0008	.0009	-.0008	.0009
$\rho = .4, \alpha = .8$	.0007	.0063	-.0004	.0042	8, 1	-.0006	.0040	-.0006	.0040	-.0006	.0040
					8, 8	-.0009	.0040	-.0008	.0040	-.0008	.0040
					4, 1	-.0006	.0040	-.0006	.0040	-.0006	.0040
					4, 4	-.0008	.0040	-.0008	.0040	-.0008	.0040
$\rho = .8, \alpha = .8$	.0007	.0063	-.0008	.0010	8, 1	-.0008	.0010	-.0007	.0010	-.0007	.0010
					8, 8	-.0008	.0010	-.0008	.0010	-.0008	.0010
					4, 1	-.0009	.0010	-.0008	.0010	-.0009	.0010
					4, 4	-.0009	.0010	-.0009	.0010	-.0009	.0010

Table 2: DGP 2, the error term associated to the variable of interest follows a MA(1) process, while the error term associated to the auxiliary variable is iid Normal.

	$\bar{y}$		GMM			EL		ET		EU	
	$BIAS$	MSE	$BIAS$	MSE	M,L	$BIAS$	MSE	$BIAS$	MSE	$BIAS$	MSE
$n = 256$											
$\rho = .4,\ \alpha = .4$	-.0012	.0096	-.0018	.0076	8, 1	-.0017	.0080	-.0016	.0079	-.0015	.0079
					8, 8	-.0022	.0080	-.0020	.0079	-.0019	.0079
					4, 1	-.0018	.0077	-.0018	.0077	-.0018	.0077
					4, 4	-.0017	.0078	-.0018	.0077	-.0019	.0077
$\rho = .8,\ \alpha = .4$	-.0012	.0096	-.0015	.0029	8, 1	-.0019	.0030	-.0018	.0030	-.0017	.0030
					8, 8	-.0022	.0030	-.0022	.0030	-.0020	.0030
					4, 1	-.0017	.0029	-.0017	.0029	-.0017	.0029
					4, 4	-.0018	.0029	-.0018	.0029	-.0018	.0029
$\rho = .4, \alpha = .8$	-.0034	.0853	-.0042	.0509	8, 1	-.0049	.0409	-.0046	.0405	-.0039	.0411
					8, 8	-.0065	.0399	-.0053	.0396	-.0049	.0402
					4, 1	-.0041	.0444	-.0041	.0443	-.0039	.0445
					4, 4	-.0043	.0442	-.0045	.0440	-.0045	.0442
$\rho = .8, \alpha = .8$	-.0034	.0853	-.0031	.0090	8, 1	-.0035	.0070	-.0035	.0069	-.0033	.0074
					8, 8	-.0032	.0070	-.0033	.0069	-.0040	.0074
					4, 1	-.0034	.0074	-.0019	.0074	-.0032	.0076
					4, 4	-.0036	.0073	-.0036	.0073	-.0036	.0075
$n = 512$											
$\rho = .4,\ \alpha = .4$	.0007	.0054	-.0003	.0043	8, 1	-.0003	.0044	-.0003	.0044	-.0003	.0044
					8, 8	-.0006	.0044	-.0006	.0044	-.0006	.0044
					4, 1	-.0005	.0043	-.0005	.0043	-.0004	.0043
					4, 4	-.0005	.0043	-.0005	.0043	-.0004	.0043
$\rho = .8,\ \alpha = .4$	.0007	.0054	-.0009	.0016	8, 1	-.0010	.0016	-.0008	.0016	-.0009	.0016
					8, 8	-.0010	.0016	-.0010	.0016	-.0009	.0016
					4, 1	-.0009	.0016	-.0008	.0016	-.0009	.0016
					4, 4	-.0009	.0016	-.0009	.0016	-.0009	.0016
$\rho = .4, \alpha = .8$	.0031	.0471	.0007	.0284	8, 1	-.0009	.0219	-.0004	.0227	-.0009	.0219
					8, 8	-.0010	.0216	-.0009	.0216	-.0008	.0217
					4, 1	.0001	.0242	.0004	.0247	.0001	.0242
					4, 4	-.0002	.0240	-.0001	.0240	-.0001	.0241
$\rho = .8, \alpha = .8$	.0031	.0471	-.0012	.0049	8, 1	-.0019	.0036	-.0009	.0043	-.0016	.0036
					8, 8	-.0018	.0035	-.0017	.0035	-.0016	.0036
					4, 1	-.0015	.0039	.0006	.0042	-.0014	.0039
					4, 4	-.0015	.0039	-.0015	.0038	-.0015	.0039

Table 3: DGP 3, the error term associated to the variable of interest follows a AR(1) process, while the error term associated to the auxiliary variable follows a MA(1).

	$\bar{y}$		GMM			EL		ET		EU	
	$BIAS$	MSE	$BIAS$	MSE	M,L	$BIAS$	MSE	$BIAS$	MSE	$BIAS$	MSE
$n=256$											
$\rho=.4,\ \alpha=.4$	.0003	.0069	-.0006	.0062	8, 1	-.0007	.0064	-.0006	.0064	-.0005	.0063
					8, 8	-.0009	.0063	-.0007	.0063	-.0006	.0063
					4, 1	-.0004	.0063	-.0004	.0063	-.0004	.0063
					4, 4	-.0002	.0062	-.0003	.0062	-.0003	.0062
$\rho=.8,\ \alpha=.4$	.0003	.0069	-.0012	.0033	8, 1	-.0016	.0033	-.0016	.0033	-.0014	.0033
					8, 8	-.0012	.0033	-.0012	.0033	-.0010	.0033
					4, 1	-.0013	.0033	-.0014	.0033	-.0013	.0033
					4, 4	-.0009	.0033	-.0010	.0033	-.0010	.0033
$\rho=.4,\alpha=.8$	.0004	.0115	-.0015	.0141	8, 1	-.0010	.0128	-.0005	.0127	-.0001	.0125
					8, 8	-.0015	.0125	-.0008	.0124	-.0003	.0123
					4, 1	-.0012	.0131	-.0007	.0129	-.0006	.0128
					4, 4	-.0011	.0130	-.0007	.0128	-.0006	.0127
$\rho=.8,\alpha=.8$	.0004	.0115	-.0036	.0157	8, 1	-.0026	.0111	-.0022	.0110	-.0019	.0108
					8, 8	-.0025	.0108	-.0020	.0107	-.0018	.0106
					4, 1	-.0028	.0126	-.0022	.0125	-.0027	.0124
					4, 4	-.0028	.0125	-.0025	.0123	-.0024	.0122
$n=512$											
$\rho=.4,\ \alpha=.4$	.0006	.0038	-.0006	.0034	8, 1	-.0002	.0034	.0000	.0034	-.0002	.0034
					8, 8	-.0004	.0034	-.0004	.0034	-.0004	.0034
					4, 1	-.0004	.0034	.0001	.0034	-.0004	.0034
					4, 4	-.0004	.0034	-.0003	.0034	-.0003	.0034
$\rho=.8,\ \alpha=.4$	.0006	.0038	-.0011	.0018	8, 1	-.0010	.0018	-.0005	.0018	-.0008	.0018
					8, 8	-.0009	.0018	-.0009	.0018	-.0008	.0018
					4, 1	-.0010	.0018	.0004	.0018	-.0009	.0018
					4, 4	-.0009	.0018	-.0009	.0018	-.0009	.0018
$\rho=.4,\alpha=.8$	.0007	.0063	-.0015	.0075	8, 1	-.0003	.0064	.0008	.0063	-.0003	.0064
					8, 8	.0001	.0064	.0007	.0063	-.0001	.0063
					4, 1	-.0007	.0066	.0008	.0063	-.0007	.0066
					4, 4	-.0006	.0066	-.0005	.0066	-.0006	.0066
$\rho=.8,\alpha=.8$	.0007	.0063	-.0029	.0085	8, 1	-.0015	.0057	-.0003	.0056	-.0012	.0057
					8, 8	-.0015	.0057	-.0009	.0056	-.0012	.0057
					4, 1	-.0020	.0066	.0008	.0063	-.0019	.0065
					4, 4	-.0020	.0065	-.0018	.0065	-.0019	.0065

Table 4: DGP 4, the error term associated to the variable of interest follows a MA(1) process, while the error term associated to the auxiliary variable follows a AR(1).

	$\bar{y}$		GMM			EL		ET		EU	
	$BIAS$	MSE	$BIAS$	MSE	M,L	$BIAS$	MSE	$BIAS$	MSE	$BIAS$	MSE
$n = 256$											
$\rho = .4,\ \alpha = .4$	-.0012	.0096	-.0018	.0080	8, 1	-.0006	.0084	-.0007	.0085	-.0006	.0084
					8, 8	-.0015	.0083	-.0019	.0084	-.0017	.0083
					4, 1	-.0011	.0082	-.0012	.0082	-.0011	.0082
					4, 4	-.0014	.0082	-.0017	.0082	-.0016	.0082
$\rho = .8,\ \alpha = .4$	-.0012	.0096	-.0018	.0037	8, 1	-.0015	.0038	-.0013	.0038	-.0014	.0038
					8, 8	-.0021	.0038	-.0013	.0037	-.0019	.0037
					4, 1	-.0016	.0037	-.0014	.0037	-.0016	.0037
					4, 4	-.0019	.0037	-.0015	.0037	-.0019	.0037
$\rho = .4, \alpha = .8$	-.0034	.0853	-.0037	.0753	8, 1	-.0022	.0801	-.0019	.0786	-.0016	.0793
					8, 8	-.0018	.0783	.0004	.0767	-.0020	.0771
					4, 1	-.0030	.0775	-.0025	.0765	-.0028	.0769
					4, 4	-.0038	.0765	-.0028	.0757	-.0033	.0761
$\rho = .8, \alpha = .8$	-.0034	.0853	-.0032	.0331	8, 1	-.0034	.0353	-.0014	.0358	-.0018	.0351
					8, 8	-.0049	.0345	-.0020	.0341	-.0026	.0341
					4, 1	-.0033	.0342	-.0015	.0345	-.0023	.0340
					4, 4	-.0038	.0339	-.0017	.0337	-.0029	.0336
$n = 512$											
$\rho = .4,\ \alpha = .4$	.0007	.0054	-.0005	.0046	8, 1	-.0001	.0047	.0002	.0047	.0000	.0047
					8, 8	-.0006	.0047	-.0003	.0047	-.0004	.0047
					4, 1	-.0003	.0046	.0004	.0046	-.0002	.0046
					4, 4	-.0004	.0046	-.0003	.0046	-.0003	.0046
$\rho = .8,\ \alpha = .4$	.0007	.0054	-.0011	.0020	8, 1	-.0008	.0021	-.0003	.0021	-.0010	.0020
					8, 8	-.0010	.0021	-.0007	.0021	-.0012	.0021
					4, 1	-.0009	.0020	.0008	.0021	-.0010	.0020
					4, 4	-.0010	.0020	-.0008	.0020	-.0010	.0020
$\rho = .4, \alpha = .8$	.0031	.0471	-.0006	.0400	8, 1	-.0005	.0410	.0004	.0410	-.0001	.0410
					8, 8	-.0007	.0400	.0001	.0400	-.0004	.0400
					4, 1	-.0006	.0404	.0039	.0404	-.0003	.0404
					4, 4	-.0007	.0400	.0000	.0400	-.0004	.0401
$\rho = .8, \alpha = .8$	.0031	.0471	-.0036	.0177	8, 1	-.0036	.0180	.0027	.0184	-.0033	.0181
					8, 8	-.0040	.0178	-.0020	.0180	-.0036	.0179
					4, 1	-.0038	.0178	.0054	.0183	-.0035	.0179
					4, 4	-.0039	.0177	-.0024	.0179	-.0036	.0178

Table 5: DGP 5, the error term associated to the variable of interest and the error term associated to the auxiliary variable follow a AR(1) process.