Semiparametric estimation of moment condition models with weakly dependent data

Bravo, Francesco and Chu, Ba and Jacho-Chavez, David

University of York, Carleton University, Emory University

2013

Semiparametric Estimation of Moment Condition Models with Weakly Dependent Data

Francesco Bravo*
University of York

Ba M. Chu†
Carleton University

David T. Jacho-Chávez‡
Emory University

Abstract

This paper develops the asymptotic theory for the estimation of smooth semiparametric generalized estimating equations models with weakly dependent data. The paper proposes new estimation methods based on smoothed two-step versions of the Generalized Method of Moments and Generalized Empirical Likelihood methods. An important aspect of the paper is that it allows the first step estimation to have an effect on the asymptotic variances of the second-step estimators and explicitly characterizes this effect for the empirically relevant case of the so-called generated regressors. The results of the paper are illustrated with a partially linear model that has not been previously considered in the literature. The proofs of the results utilize a new uniform strong law of large numbers and a new central limit theorem for U-statistics with varying kernels that are of independent interest.

Keywords: Alpha-Mixing; Empirical Processes; Generalised Empirical Likelihood; Kernel Smoothing; Stochastic Equicontinuity; Uniform Law of Large Numbers.

*Department of Economics, University of York, Heslington, York YO10 5DD, UK. E-mail: francesco.bravo@york.ac.uk. Web Page: https://sites.google.com/a/york.ac.uk/francescobravo/.

†Department of Economics, Carleton University, B-857 Loeb Building, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. E-mail: ba_chu@carleton.ca. Web Page: http://http-server.carleton.ca/ bchu/.

‡Corresponding Author: Department of Economics, Emory University, Rich Building 306, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA. E-mail: djachocha@emory.edu. Web Page: https://sites.google.com/site/djachocha/.

1 Introduction

In this paper we consider estimation of *semiparametric* statistical models defined by a set of generalized estimating equations. These models, often called over-identified moment conditions models in the econometric literature, are very general and contain semiparametric extensions to generalized instrumental variable models used with economics and financial data and quadratic inference functions models used with longitudinal data. We develop two-step semiparametric extensions to the generalized method of moments (GMM) proposed by Hansen (1982), the generalized empirical likelihood (GEL) estimator of Newey and Smith (2004) and the exponentially tilted empirical likelihood (ETEL) estimator of Schennach (2007), where the first step is used to estimate an infinite dimensional nuisance parameters and the second-step is used to estimate a finite dimensional parameter of interest. The aforementioned methods have many desirable theoretical and practical properties. For example, GEL is a quasi-likelihood alternative to GMM that includes Owen's (1988) Empirical Likelihood (EL), and Kitamura and Stutzer's (1997) Exponential Tilting (ET) as special cases. It does not require estimation of the efficient metric as in GMM estimation, and allows for the construction of classical-type statistics such as likelihood ratio, and score for various hypotheses of interest. On the other hand GMM is computationally simpler than GEL, whereas ETEL is known to be robust to possible global misspecification of the estimating equations.

The theoretical properties of two-step semiparametric estimators have been considered both in the statistical and econometric literature for both cross section and time series data, see e.g. Truong and Stone (1994), Andrews (1994a), Newey (1994), Gao and Liang (1997), Chen and Shen (1998), Li and Wooldridge (2002), Chen et al. (2003) to name just a few among many others. Li and Racine (2007) and Gao (2007) provide further examples and references. The statistical model we consider includes all of these models as special cases and in particular it allows for the possibility that the first-step estimation can affect the asymptotic variance of the second step estimator (the so-called estimation effect). To be specific we consider the case where the infinite dimensional parameter can depend on an estimated finite dimensional random vector. This case is empirically relevant because it often arises in situations where an estimated variable is used as a proxy for an unobservable variable of interest, such as for example the risk term in finance, and it is also theoretically interesting because with weakly dependent data the characterization of the estimation effect is more complicated. As far as we are aware of, this is the first paper that fully considers the estimation effect in semiparametric generalized estimating equations models with weakly dependent observations (see Mammen et al., 2015 and Escanciano et al., 2014 for the case of just-identified semiparametric estimating equations models with independent and identically distributed (i.i.d.) observations).

The main methodological contribution of this paper is to derive the asymptotic properties of semiparametric two-step GEL, GMM and ETEL estimators under the weakest form of dependency, namely α (or strong) mixing (see for example Doukhan, 1994, for a review of statistical properties and applications of α-mixing processes) using the same kernel based smoothing[1] proposed by Kitamura and Stutzer (1997) for ET and generalized by Smith (1997) (see also Smith, 2011) to GEL. In our frame-

[1]For an asymptotically equivalent approach based on blocking techniques see for example Kitamura (1997).

work, smoothing the estimating equations is useful whether there is an estimation effect or not. In the latter case smoothing is necessary for both the GEL and ETEL estimators to achieve the same asymptotic lower bound established by Chamberlain (1987) for efficient GMM estimators with i.i.d. observations. In the former case smoothing is useful because it results in heteroskedasticity and autocorrelation robust variance matrix estimators alternative to those typically used in both empirical economics and finance, see for example Andrews (1991). In this situation we obtain explicit formulae for the resulting asymptotic variance that are based on pathwise derivatives as in Newey (1994), and rely on a linear representation of the first-step estimator. This linear representation is fairly general and is satisfied, for example, in the important cases of non-parametric regression and non-parametric density estimators.

This paper also contains a number of new technical contributions that are used in the proof of the main results and are of independent interest. To be specific we establish a new strong uniform law of large numbers (SULLN) for strictly stationary α-mixing processes with a sharp logarithmic bound that depends on an exponential decay rate of the α-mixing coefficient, a weak condition on the growth rate of the bracketing entropy of a polynomial class of functions (of which Vapnik-Červonenkis (V-C) classes are a special case), see e.g., van der Vaart and Wellner (1996, p. 86), and the existence of certain moments of the estimating equations. This result extends a number of ULLN available in both the econometric and statistical literature including those obtained by Andrews (1987), Yu (1993, 1994), Doukhan et al. (1994) and Adams and Nobel (2010). We also introduce two new central limit theorems (CLT) (see Appendix B in the supplemental material) for both degenerate and nondegenerate second-order generalized U-statistics (that is U-statistics with varying kernels). The resulting CLTs are important because they represent a nontrivial extension of the existing results that are valid for either i.i.d. or β-mixing sequences – see for example, de Jong (1987), Powell et al. (1989) and Mikosch (1993) for the i.i.d. case, and Yoshihara (1976, 1989) and Fan and Li (1999) for the β-mixing case. To establish these theorems, we impose mild regularity conditions directly on the kernel of the U-statistic and rely on Sun and Chiang's (1997) conditional expectation bound for α-mixing sequences and on Dvoretsky's (1972) central limit theorem for double arrays of dependent random variables.[2]

The theoretical results of the paper are illustrated by deriving the asymptotic properties of an estimator of a general partially linear regression model, where we allow for the unobservable error to be correlated with the regressors and the infinite dimensional parameter to depend on an unknown finite dimensional parameter. Other examples where the results of the paper can be used are the weighted instrumental variable model that adapt for unknown heteroskedasticity of Robinson (1987), the instrumental variable model of sample selection of Lee (1994), and the inverse-density-weighted moment model of Chu and Jacho-Chávez (2012) and Chu et al. (2013).

The rest of the paper is organized as follows: The next section introduces the statistical model and the estimators. Section 3 contains the asymptotic results. Sections 4 and 5, respectively, introduce

[2]We note that Yoshihara (1992) uses an alternative approach to the one we follow to obtain the CLTs (and more generally invariance principles) for α-mixing sequences. His approach relies on the Karhunen–Loève expansion of the kernel and is based on a set of regularity conditions that are not imposed directly on the kernel and thus could be very hard to verify in practice.

the new partially linear regression model and the results of the Monte Carlo simulations used to assess the finite sample properties of the proposed estimators. Section 6 contains some concluding remarks. The proofs of the theorems of Sections 3 and 4 are contained in the Appendix A. A supplement to this paper contains the new CLT's for second-order generalized U-statistics, a number of auxiliary technical lemmas and related proofs, which should be of independent interest.

The following notation is used in the text: a " $'$ " denotes a matrix or vector transpose; for any finite dimensional possibly random vector v or square matrix M, $\|\cdot\|$ denotes the Euclidean norm and $\|v\|_M := v'Mv$; for any measurable possibly vector valued function $f(\cdot)$, let $\|f(\cdot)\|_p$ denote the L_p norm, i.e., $\left(\int \|f(x)\|^p P(dx)\right)^{1/p}$, and more generally for a pseudo-metric space, say $\mathcal{H}$, $\|\cdot\|_{\mathcal{H}}$ denotes a function norm, such as the sup norm.

2 The Model and Estimators

Let $\{z_t, t = 1, 2, ...\}$ be a sequence of $\mathcal{Z}$-valued $\left(\mathcal{Z} \subset \mathbb{R}^d\right)$ weakly dependent random vectors defined on a probability space $(\Omega, \mathcal{B}, P)$. Let $\theta \in \Theta \subset \mathbb{R}^k$ denote the finite dimensional parameter of interest and $h \in \mathcal{H}$ denote the infinite dimensional nuisance parameter where $\mathcal{H}$ is a pseudo-metric space.

We consider a smooth semiparametric statistical model defined by

$$E\left[g\left(z_t, \theta, h\right)\right] = 0 \quad \text{iff } \theta = \theta_0 \in \text{int}(\Theta), \text{ and } h = h_0 \in \mathcal{H}, \tag{2.1}$$

where $g(\cdot) : \mathcal{Z} \times \Theta \times \mathcal{H} \to \mathbb{R}^l$ $(l \geq k)$ is a vector-valued measurable known function, and $\theta_0 \in \text{int}(\Theta)$ and $h_0 \in \mathcal{H}$ are the true unknown parameters. As in Andrews (1994a), h is allowed to depend on z_t and possibly on a finite dimensional parameter $\alpha \subset A \subset \mathbb{R}^p$, so that $h_0 =: h_0\left(z_t, \alpha_0\right)$ includes also the case of estimated random variables.

Let $g_t(\theta, h) := g(z_t, \theta, h)$; given a sample $\{z_t\}_{t=1}^T$ and a preliminary non-parametric estimator $\widehat{h}$ of h_0 a two-step GMM estimator $\widehat{\theta}$ for θ_0 is defined as

$$\widehat{\theta}^{\text{GMM}} = \arg\min_{\theta\in\Theta} \| \, \widehat{g}(\theta, \widehat{h}) \, \|_{\widehat{W}} \,, \tag{2.2}$$

where $\widehat{g}(\theta, \widehat{h}) := T^{-1} \sum_{t=1}^T g_t(\theta, \widehat{h})$ and $\widehat{W}$ is a positive semi-definite possibly random $\mathbb{R}^l \times \mathbb{R}^l$-valued matrix that may depend on θ, and $\widehat{h}$. The consistency of $\widehat{\theta}$ follows by the results of Andrews (1994a) and Chen et al. (2003), whereas its asymptotic normality follows by the results of Andrews (1994a) with weakly dependent observations under the assumption of asymptotic orthogonality - see Assumption 6 given below- and in full generality by the results of Chen et al. (2003) but only under the assumption of i.i.d. observations.

An alternative method for estimating θ_0 is to use GEL and/or ETEL instead. To handle the dependent structure of the estimating equation $g_t(\theta, h)$, we follow the same approach of Smith (1997) and consider the following smoothed version

$$g_{ts}(\theta, h) = \frac{1}{s_T} \sum_{j=t-T}^{t-1} \omega\left(\frac{j}{s_T}\right) g_{t-j}(\theta, h)\,, \quad t = 1, \ldots, T,$$

where s_T is a bandwidth parameter and $\omega(\cdot)$ is a kernel function. Examples of possible kernel functions include the Bartlett kernel $\omega_B(\cdot)$ used for example by Kitamura and Stutzer (1997) and the quadratic spectral kernel $\omega_{QS}(\cdot)$ considered by Andrews (1991), given, respectively, by

$$\omega_B(x) = \begin{cases} 1 - |x| & ; \quad |x| \le 1 \\ 0 & ; \quad \text{otherwise,} \end{cases} \tag{2.3}$$

$$\omega_{QS}(x) = \frac{25}{12\pi^2 x^2}\left[\frac{\sin(6\pi x/5)}{6\pi/5} - \cos\left(\frac{6\pi x}{5}\right)\right]. \tag{2.4}$$

Smith (2011) provides further examples and a detailed discussion of different choices of $\omega(\cdot)$.

Let $\rho(\cdot) : Q \to \mathbb{R}$ denote a twice continuously differentiable function that is concave in its domain Q - an open interval of the real line that contains 0. The smoothed two-step GEL criterion function for the semiparametric estimating equation satisfying (2.1) is

$$\Gamma(\theta, h, \lambda) = \frac{2}{T}\sum_{t=1}^{T}[\rho(\omega\lambda' g_{ts}(\theta, \widehat{h})) - \rho(0)],$$

where $\omega = \omega_1/\omega_2$ ($\omega_j := \int \omega(q)^j\, dq,\ j = 1, 2, \ldots$) is a normalization that has no effect on the GEL estimator for θ_0 but makes the scale of the estimator for λ comparable for different choices of $\omega(\cdot)$ and λ is a vector of unknown auxiliary parameters.

The GEL estimator for θ_0 is defined as the minimizer of the (profile) smoothed two-step GEL criterion function, that is

$$\widehat{\theta}^{\text{GEL}} = \arg\min_{\theta\in\Theta} \Gamma(\theta, \widehat{h}, \widehat{\lambda}), \tag{2.5}$$

where

$$\widehat{\lambda} := \arg\max_{\lambda\in\Lambda_T} \Gamma(\theta, \widehat{h}, \lambda), \tag{2.6}$$

for some fixed θ and $\Lambda_T = \{\lambda : \lambda' g_{ts}(\theta, \widehat{h}) \in Q\}$ is the restricted parameter space of λ (see for example Newey and Smith, 2004 and Smith, 2011).

We can also define the following two-step smoothed GMM estimator for θ_0,

$$\widehat{\theta}^{\text{s-GMM}} = \arg\min_{\theta\in\Theta} \| \widehat{g}_s(\theta, \widehat{h}) \|_{\widehat{W}}, \tag{2.7}$$

where $\widehat{g}_s(\theta, \widehat{h}) := T^{-1}\sum_{t=1}^{T} g_{ts}(\theta, \widehat{h})$, which is an extension of that proposed by Smith (2005) and, as opposed to the standard GMM estimator, takes directly into account the weakly dependent structure of the observations.[3]

[3] This implies that a consistent estimator of the efficient metric $W = \lim_{T\to\infty} \text{var}(T^{1/2}\widehat{g}(\theta_0, h_0))$ is given by an appropriately standardized version of the outer product of the smoothed estimating equations $g_{ts}(\widehat{\theta}, \widehat{h})$, viz.

$$\left\| \left[\frac{1}{s_T}\sum_{j=1-T}^{T-1} \omega\left(\frac{j}{s_T}\right)^2\right]^{-1} \frac{s_T}{T}\sum_{t=1}^{T} g_{ts}(\widehat{\theta}, \widehat{h}) g_{ts}(\widehat{\theta}, \widehat{h})' - \lim_{T\to\infty} \text{var}(T^{1/2}\widehat{g}(\theta_0, h_0)) \right\| = o_p(1),$$

see the proof of Theorem 3.2 for more details.

The last estimator we consider is the two-step semiparametric ETEL estimator for θ_0, that is defined as

$$\widehat{\theta}^{\text{ETEL}} = \arg\min_{\theta\in\Theta} \frac{1}{T}\sum_{t=1}^{T} T\log\widehat{\pi}_s(z_t,\theta,\widehat{h},\widehat{\lambda}), \tag{2.8}$$

where $\widehat{\pi}_s(z_t,\theta,\widehat{h},\widehat{\lambda}) = \rho_1(\omega\widehat{\lambda}' g_{ts}(\theta,\widehat{h}))/\sum_{t=1}^{T}\rho_1(\omega\widehat{\lambda}' g_{ts}(\theta,\widehat{h}))$, and $\widehat{\lambda}$ is as in (2.6) for $\rho(\cdot) = -\exp(\cdot)$.

3 Asymptotic Theory

3.1 Strong Uniform Law of Large Numbers

We begin this section by introducing some further notation: Let $\mathcal{F} := \{f(\theta,h) : \theta\in\Theta, h\in\mathcal{H}\}$ denote a class of functions indexed by an Euclidean parameter and an infinite dimensional parameter. Given a probability distribution P and $\mathcal{F}$ in $L_p(P)$, let $N_{[],p}(\epsilon,P,\mathcal{F})$ and $H_{[],p}(\epsilon,P,\mathcal{F})$ denote, respectively, the bracketing number and the ϵ-entropy with bracketing of $\mathcal{F}$ (see for example van der Vaart and Wellner, 1996, Section 2.1, pp. 80-94)

Assumption 1 *$\{z_t, t = 1,2,\ldots\}$ is a sequence of $\mathcal{Z}$-valued $\left(\mathcal{Z}\subset\mathbb{R}^d\right)$ stationary α-mixing random vectors with the mixing coefficient satisfying $\alpha(t) = O\left(\exp(-at^b)\right)$ for some positive a and b.*

Assumption 2 *The class of functions $\mathcal{F}$ satisfies*

$$H_{[],1}(\epsilon,P,\mathcal{F}) \le \upsilon\log\left(\frac{1}{\epsilon}\right) \text{ for some } \upsilon > 0, \tag{3.1}$$

$$E\left[\sup_{(\theta,h)\in\Theta\times\mathcal{H}} \|f_t(\theta,h)\|^{\mu}\right] < \infty \text{ for some } \mu\ge 4. \tag{3.2}$$

Assumption 1 specifies the dependent structure of the observations as α-mixing. Examples of time series models that are α-mixing can be found in Doukhan (1994). α-mixing dependency is considered by Andrews (1994a) in the context of semiparametric models, and by Kitamura (1997) and Smith (2011) in the context of EL and GEL estimation and inference for (finite dimensional) generalized estimating equations models. Assumption 1 imposes an exponential decay rate on the α-mixing coefficient $\alpha(t)$, which could be satisfied by many m-dependent stochastic processes, such as ARMA, GARCH, and bilinear processes; this same type of assumption has also been employed by Boente and Fraiman (1988) and Bonhomme and Manresa (2015) for example. Assumption 2 imposes a restriction on the complexity of the class of functions $\mathcal{F}$ and the existence of some moments of order greater than 4. Various types of function classes such as Hölder, Sobolev and many others can be shown to satisfy (3.1) (see, e.g., van der Vaart and Wellner, 1996, Section 2.7, pp. 154-165). Note that (3.2) is only used to establish the strong convergence rate in the following theorem.[4]

[4]Note that condition (3.1) combined with (3.2) for $\mu = 2+\zeta$ for some $\zeta > 0$ would suffice to prove a weaker version of the uniform law of large numbers given in Theorem 3.1.

Theorem 3.1 *Under Assumptions 1 and 2*

$$\sup_{(\theta,h)\in\Theta\times\mathcal{H}}\left|\frac{1}{T}\sum_{t=1}^{T}\{f_t(\theta,h)-E[f_t(\theta,h)]\}\right| = O_{a.s.}\left(\frac{\log T}{T^{\beta}}\right) \text{ for some } \beta\in\left(0,\frac{1}{4}\right).$$

Remark 3.1 *The proposed ULLN complements that of Yu (1993, 1994) who established a rate of convergence for a ULLN for strictly stationary β mixing (absolutely regular) empirical processes indexed by a general class of functions with its capacity measured via the empirical metric entropy.*

The above result is used repeatedly in the proofs of the Theorems 3.2 and 3.3. Its proof can be found in the supplemental material for this paper.

3.2 Asymptotic Normality

Let $\Theta_\delta = \{\theta \in \Theta : \|\theta-\theta_0\| \le \delta\}$, $\mathcal{H}_\delta = \{h \in \mathcal{H} : \|h-h_0\|_{\mathcal{H}} \le \delta\}$ (possibly uniformly in $\alpha \in A$), where $h := h(z_t)$ for some positive generic constant δ. Also let $\partial_\cdot$ denote a derivative operator with respect to $\cdot$, which corresponds to an ordinary partial derivative with respect to θ, and to the pathwise derivative in the direction of $h-h_0$, that is

$$\frac{\partial g(z_t,\theta,h_0)}{\partial h}[h-h_0] := \frac{\partial g(z_t,\theta,(1-\tau)h_0+\tau h)}{\partial \tau}|_{\tau=0}$$

(see Newey, 1994 for some examples). Assume that:

Assumption 3 *(a) $s_T \to \infty$ as $T\to\infty$, and $s_T = O(T^{\frac{1}{2}-\eta}) = o(T^{1/2})$ for some $\eta \in \left(\frac{1}{6},\frac{1}{2}\right)$ (cf. Smith, 2011); (b) $\omega(\cdot) : \mathbb{R}\to[-\overline{\omega},\overline{\omega}]$ for some $\overline{\omega}<\infty$, $\omega(0)\neq 0$, $\omega_1 \neq 0$, $\omega(x)$ is continuous at 0 and almost everywhere, $(2\pi)^{-1}\int_{-\infty}^{\infty}\exp(-\iota x u)\,\omega(x)\,dx \ge 0$ for each $\omega\in\mathbb{R}$ and all $u\in\mathbb{R}$, and $\int_{-\infty}^{0}\sup_{y\le x}|\omega(y)\,dx| + \int_{0}^{\infty}\sup_{y\ge x}|\omega(y)\,dx| < \infty$.*

Assumption 4 *(a) The class of functions $\mathcal{G}_1 := \{g_t(\theta,h) :\ \theta\in\Theta,\ h\in\mathcal{H}\}$ satisfies conditions (3.1) and (3.2) in Assumption 2; (b) $E\left[\sup_{\theta\in\Theta,h\in\mathcal{H}_\delta}\|\partial_\theta g_t(\theta,h)\|^{\alpha}\right]<\infty$ and $E[\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta}\|\partial_h g_t(\theta,h)\|^{\alpha}] < \infty$ for some $\alpha>2$; (c) the class of functions $\mathcal{G}_2 := \{\partial_{\theta h} g_t(\theta,h) :\ \theta\in\Theta,\ h\in\mathcal{H}\}$ satisfies conditions (3.1) and (3.2) in Assumption 2, $E[\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta}\left\|\partial^2_{\theta\theta}g_t(\theta,h)\right\|]<\infty$.*

Assumption 5 *(a) $\|\widehat{h}(z_t)-h_0(z_t)\|_{\mathcal{H}} = o_p\left(T^{-1/4}\right)$;*
(b) $\widehat{v}_T(\theta,h) := T^{-1/2}\sum_{t=1}^{T}\{g_t(\theta,h)-E[g_t(\theta,h)]\}$ is stochastically equicontinuous at $(\theta_0,h_0)\in\Theta\times\mathcal{H}$.

Assumption 3 imposes some standard mild regularity conditions on the kernel function $\omega(\cdot)$ used to smooth the observations and on the rate of growth of the related smoothing parameter s_T. Note that the latter is allowed to grow at the rate $O(T^{1/3})$, which is known to be optimal (in terms of minimizing the asymptotic mean squared error) for α-mixing processes for the Bartlett kernel. Examples of kernels satisfying Assumption 3 include the Bartlett and the quadratic one given in (2.3), (2.4) respectively and the Parzen kernel (see Andrews, 1991, for more details). Assumption 4 contains some mild moment conditions and requires that the classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$ satisfy the conditions of Theorem

3.1. Assumptions 2, 3 and 4(a) can be used to show the consistency of the estimators described above. Assumption 5(a) assumes uniform consistency (possibly also with respect to α) of the non-parametric estimator used for h_0. This is a standard assumption in the semiparametric literature of two-step estimation procedures, see, e.g., Chen et al. (2003), Escanciano et al. (2014, 2016), Chen et al. (2016), and Bravo et al. (2016). Similarly, Andrews (1995) provides sufficient conditions including the case of estimated random variables for kernel smoothing estimators. Assumption 5(b) is a high level assumption. It assumes stochastic equicontinuity of the empirical process $\widehat{v}_T(\theta, h)$. Although, sufficient conditions for Assumption 5(b) are provided for example in Andrews (1994a,b), Lemma C.3 in the Appendix C in the supplement provides a set of low level conditions that can be used to verify Assumption 5(b).

Assumption 6 *(a)* $\| E[g_t(\theta_0, \widehat{h})] \| = o_p\left(T^{-1/2}\right)$*; or (b)* $E[\partial g(z_t, \theta, \tau)/\partial\tau|_{\tau=h_0}\widetilde{h}(z_t)] = 0\ \forall \widetilde{h} \in \mathcal{H}$ *and* $z_{2t} \subset z_t$*.*

Assumption 7 *(a)* $\widehat{h}(w) - h_0(w) = T^{-1}\sum_{t=1}^{T} \Phi_T(z_{2t}, w) \odot \phi(z_t) + r_T(w)$*, where "*$\odot$*" is the Hadamard product,* $\Phi_T(z_{2t}, \cdot)$ *is some weighting function,* $\|r_T(w)\|_{\mathcal{H}} = o_p\left(T^{-1/2}\right)$ *(possibly uniformly in* $\alpha \in A$*);*
(b) $E\left[\phi(z_t) | \mathcal{F}_{t,z_{2t}}\right] = 0$*, where* $\mathcal{F}_{t,z_{2t}}$ *is the minimum* σ*-algebra generated by* z_{2t}*;* $E\left[\phi(z_t)\phi(z_t)'\right] < \infty$*; and* $\lim_{T\to\infty}\sup_w \mathrm{var}(T^{-\left(\frac{1}{2}+\delta\right)}\sum_{t=1}^{T}\Phi_T(z_{2t}, w) \odot \phi(z_t)) < \infty$ *for some* $\delta \in (0, 1/2)$*;*
(c) the class of functions $\mathcal{G}_3 := \{\partial^2_{hh} g(z_t, \theta_0, h) : h \in \mathcal{H}\}$ *satisfies conditions (3.1) and (3.2) Assumption in 2.*

Assumptions 6 and 7 account for the potential estimation effect from the first-step. When there is none, Assumption 6 implies the asymptotic orthogonality between the finite dimensional and the infinite dimensional parameter. In such case, it is not necessary to account for the presence of $\widehat{h}$ in the asymptotic distribution of $\widehat{\theta}$, which greatly simplifies the calculation of the asymptotic variance. Condition 6(a) is directly assumed by Andrews (1994a), while Assumption 6(b) is assumed by Newey (1994). Note that for $h = h(z_{2t})$ sufficient conditions for condition 6(a) are Assumptions 6(b) and 5(a). On the other hand, when there is estimation effect, Assumption 7 provides a generic way to account for it. For example, when h_0 represents a conditional mean function, Assumption 7(a) requires that the first-step estimator admits a certain asymptotic expansion which can be shown to hold when $\widehat{h}$ represents some kernel-based non-parametric regression estimator of h_0 (see for example Masry, 1996 and Kong, Linton, and Xia, 2010); or $\widehat{h} := h(\cdot, \widehat{\alpha})$ when $h_0(\cdot) = h(\cdot, \alpha_0)$ is known up to some vector of parameters α_0. For instance, when $\widehat{h}$ is the Nadaraya-Watson estimator of h_0 in a non-parametric regression model, say $z_{1t} = h_0(z_{2t}) + \xi_t$, then one can immediately show that Assumption 7(a) holds under some regularity conditions with $\phi(z_t) = z_{1t} - h_0(z_{2t})$ and $\Phi_T(z_{2t}, w_t) = f_{z_{2t}}(w_t)K_{b_T}(z_{2t} - w_t)$, where $f_{z_{2t}}(\cdot)$ is the pdf of z_{2t} and $K_{b_T}(\cdot)$ is a kernel function with bandwidth $b_T = b(T)$ that goes to zero as T diverges to infinity.

The following two theorems establish the asymptotic normality for the smoothed two-step GEL, both two-step efficient s-GMM, and smoothed two-step ETEL estimators under the asymptotic orthogonality Assumption 6, and under the presence of an estimation effect that can be characterized by Assumption 7, respectively.

Let $\Omega(\theta_0, h_0) = \lim_{T\to\infty} \mathrm{var}\big(T^{1/2}\widehat{g}(z_t, \theta_0, h_0)\big)$, $G(\theta_0, h_0) = E[\partial_\theta g(z_t, \theta_0, h_0)]$ and $\Sigma(\theta_0, h_0) = G(\theta_0, h_0)' \Omega(\theta_0, h_0)^{-1} G(\theta_0, h_0)$.

Theorem 3.2 *Assume that (a) $\theta_0 \in$int(Θ), (b) $\Omega(\theta_0, h_0)$ is positive definite, (c)* rank$(G(\theta_0, h_0)) = k$, *(d) $\Sigma(\theta_0, h_0)$ is nonsingular, (e) $\| \widehat{W} - \Omega(\theta_0, h_0)^{-1} \| = o_p(1)$ for the GMM defined in (2.2) and s-GMM estimator defined in (2.7). Then under Assumptions 1-6 for $\widehat{\theta}$ defined as in (2.2), (2.5), (2.7) and (2.8)*

$$T^{1/2}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma(\theta_0, h_0)^{-1}).$$

The following theorem establishes the asymptotic normality of the above estimator in the presence of estimation effect. Let

$$\Omega_d^e(\theta_0, h_0) = \lim_{T\to\infty} \mathrm{var}\left[\frac{1}{T^{1/2}} \sum_{t=2}^{T} \left(g_t(\theta_0, h_0) + \frac{1}{(T-1)} \sum_{s=1}^{t-1} \Psi(z_s, z_t, \theta_0, h_0)\right)\right], \tag{3.3}$$

$$\Omega_{nd}^e(\theta_0, h_0) = \lim_{T\to\infty} \mathrm{var}\left[\frac{1}{T^{1/2}} \left(\sum_{t=2}^{T} g_t(\theta_0, h_0) + h_T^{(1)}(z_t, \theta_0, h_0)\right)\right],$$

where

$$\Psi(z_s, z_t, \theta_0, h_0) = \partial_h g(z_t, \theta_0, h_0)' \Phi_T(z_{2s}, z_{2t}) \odot \phi(z_s) + \partial_h g(z_s, \theta_0, h_0)' \Phi_T(z_{2t}, z_{2s}) \odot \phi(z_t),$$

$$h_T^{(1)}(\cdot, \theta_0, h_0) = E[\Psi(\cdot, z_t, \theta_0, h_0)] = \int \Psi(\cdot, u, \theta_0, h_0(u)) f_{z_t}(u) du.$$

Theorem 3.3 *Assume that (a) $\theta_0 \in$int(Θ), (b) $\Omega(\theta_0, h_0)$, $\Omega_d^e(\theta_0, h_0)$ and $\Omega_{nd}^e(\theta_0, h_0)$ are positive definite, (c)* rank$(G(\theta_0, h_0)) = k$, *(d) $\Sigma(\theta_0, h_0)$ is nonsingular. Then under Assumptions 1-5, and 7 for $\widehat{\theta}$ defined in (2.5) or in (2.8)*

$$T^{1/2}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma(\theta_0, h_0)^{-1} \Sigma_*^v(\theta_0, h_0) \Sigma(\theta_0, h_0)^{-1}),$$

where

$$\Sigma_*^v(\theta_0, h_0) = G(\theta_0, h_0)' \Omega(\theta_0, h_0)^{-1} \Omega_*^e(\theta_0, h_0) \Omega(\theta_0, h_0)^{-1} G(\theta_0, h_0),$$

and $\Omega_*^e(\theta_0, h_0)$ is either $\Omega_d^e(\theta_0, h_0)$ or $\Omega_{nd}^e(\theta_0, h_0)$ given in (3.3).
For the two-step GMM estimator and its smoothed version, say $\widehat{\theta}^\ell$ for $\ell \in \{GMM, s\text{-}GMM\}$, defined in (2.2) and in (2.7) under (a)-(c) above, (d) $\Sigma^e(\theta_0, h_0)$ is nonsingular and Assumptions 2-5, 7 and (e) $\| \widehat{W} - \Omega_*^e(\theta_0, h_0)^{-1} \| = o_p(1)$,

$$T^{1/2}(\widehat{\theta}^\ell - \theta_0) \xrightarrow{d} N(0, \Sigma_*^e(\theta_0, h_0)^{-1}),$$

where

$$\Sigma_*^e(\theta_0, h_0) = G(\theta_0, h_0)' \Omega_*^e(\theta_0, h_0)^{-1} G(\theta_0, h_0).$$

Remark 3.2 *It is important to note that*

$$\Sigma_*^e(\theta_0,h_0)^{-1} \leq \Sigma(\theta_0,h_0)^{-1}\Sigma_*^v(\theta_0,h_0)\Sigma(\theta_0,h_0)^{-1}$$

in the matrix sense,[5] *implying that in the presence of an estimation effect, as long as condition (e) of Theorem 3.3 is satisfied, the two-step GMM estimator is more efficient than the smoothed two-step GEL or ETEL estimators. On the other hand, because of the explicit estimation of the efficient metric* $\Omega_*^e(\theta_0,h_0)^{-1}$ *both GMM estimators* $\widehat{\theta}^\ell$ *for* $\ell \in \{GMM, s\text{-}GMM\}$ *might be more prone to bias. The Monte Carlo evidence of Section 5 based on the model considered in Section 4 seems to provide some support to both points.*

4 Example: Partially Linear Instrumental Variable model

We consider a generalization of the partial linear model considered by Li and Wooldridge (2002)

$$y_t = x_{1t}'\theta_0 + m_0(x_{2t}) + \varepsilon_t \quad t = 1,\ldots,T, \tag{4.1}$$

where θ_0 is an $\mathbb{R}^k$-valued vector of unknown parameters, $m_0(\cdot)$ is an unknown real valued function, and the unobservable weakly dependent errors ε_t's are such that $E[\varepsilon_t|x_t] \neq 0$, where $x_t = [x_{1t}', x_{2t}']'$. Suppose that there exists an $\mathbb{R}^l$-valued $(l \geq k)$ vector w_t of instruments such that $E(\varepsilon_t|x_{2t}, w_t) = 0$; then the estimation of the parameter of interest θ_0 can be based on

$$g_t(\theta_0,h_0) = w_t\left[y_t - E(y_t|x_{2t}) - (x_{1t} - E(x_{1t}|x_{2t}))'\theta_0\right], \tag{4.2}$$

where $h_0 := h_0(x_{2t}) = [E(y_t|x_{2t}), E(x_{1t}|x_{2t})']'$.

For $v_t = y_t$ or x_{1t} let $\widehat{E}(v_t|x_{2t}) = \sum_{s\neq t=1}^T v_t K_{b_T}((x_{2s}-x_{2t})/b_T)/\sum_{s\neq t=1}^T K_{b_T}((x_{2s}-x_{2t})/b_T)$, where $K_{b_T}(\cdot) = K(\cdot)/b_T$ denotes a kernel estimator of the conditional expectation $E[v_t|x_{2t}]$ with bandwidth b_T and let

$$g_t(\theta,\widehat{h}) = w_t\left(\widetilde{y}_t - \widetilde{x}_{1t}'\theta\right),$$

where $\widetilde{y}_t = y_t - \widehat{E}(y_t|x_{2t})$, $\widetilde{x}_{1t} = x_{1t} - \widehat{E}(x_{1t}|x_{2t})$ denote the plug-in version of (4.2).

The following proposition establishes the asymptotic distribution of the two-step GMM, two-step GEL and two-step ETEL estimators when there is an estimation effect. To this end note that by the results of Andrews (1994a) and Newey (1994), an estimation effect in (4.2) is only possible in the case of a generated regressor. So we assume that x_{2t} is generated as a residual from the following linear regression model $s_t = v_t'\alpha_0 + x_{2t}$ where α_0 is a vector of unknown parameters and v_t is a vector of exogenous regressors so that $E[x_{2t}|v_t] = 0$. We also note that because the model is linear in both the finite and infinite dimensional parameters some of the regularity conditions (including a polynomial rate for the mixing coefficient $\alpha(t)$) are weaker than those assumed in the theorems of the previous section.

[5] This follows since $\Sigma^e(\theta_0,h_0) - \Sigma(\theta_0,h_0)\Sigma_*^v(\theta_0,h_0)^{-1}\Sigma(\theta_0,h_0) = X_0'[I - Z_0(Z_0'Z_0)^{-1}Z_0']X_0 \geq 0$, for $X_0 = \Omega_*^e(\theta_0,h_0)^{-1/2}G(\theta_0,h_0)$ and $Z_0 = \Omega_*^e(\theta_0,h_0)^{1/2}\Omega(\theta_0,h_0)^{-1}G(\theta_0,h_0)$.

Proposition 4.1 *Let $z_t := [y_t, x'_{1t}, x_{2t}, w'_t]'$, and assume that: (a) $\{z_t\}_{t=1}^{T}$ is a sequence of α-mixing random vectors with $\alpha(t) = o\left(t^{-2(2+\gamma)}\right)$; (b) the joint density $f(z_t)$ of z_t and the marginal density $f(x_{2t})$ of x_{2t} are twice continuously differentiable with bounded derivatives and $\inf_{x_{2t}\in\mathcal{X}_2^*} f(x_{2t}) > 0$, where $\mathcal{X}_2^*$ is an open bounded subset of $\mathbb{R}^{d_{x_2}}$ (c) $h_0(x_{2t})$ is twice continuously differentiable and $\sup_{x_{2t}\in\mathcal{X}_2^*} \| h_0^{(j)}(x_{2t}) \| < \infty$ $(j = 0, 1, 2)$ uniformly in A where $h_0^{(j)}(\cdot)$ is the jth derivative of $h_0(\cdot)$; (d) $E \| w_t(y_t - E(y_t|x_{2t}) - (x_{1t} - E(x_{1t}|x_{2t}))' \theta_0) \|^{4+\gamma} < \infty$; (e) $\mathrm{rank}\left(E\left[w_t(x_{1t} - E(x_{1t}|x_{2t}))'\right]\right) = k$, the matrices $\Omega(\theta_0, h_0)$ and $\Omega^e(\theta_0, h_0)$ defined in (4.3) are positive definite; (f) the function $K(\cdot)$ is a nonnegative second-order kernel with second order continuous bounded derivatives, and b_T satisfies $T^{1/2}b_T^2 \to \infty$, $T^{1/2}b_T^4 \to 0$. Moreover $\left|K(\cdot + u) - K(u) - K^{(1)}(\cdot)u\right| \leq \mathcal{K}(\cdot)u^2$ where $K^{(1)}(\cdot)$ is the first derivative of the kernel function and $\mathcal{K}(\cdot)$ is a bounded function, (f) $T^{1/2}(\widehat{\alpha} - \alpha_0) = \sum_{t=1}^{T} r(v_t)' x_{2t}/T^{1/2} + o_p(1)$. Then the two-step GMM, GEL and ETEL estimators have the same distribution as that given in Theorem 3.3 with*

$$G(\theta_0, h_0) = E\left[w_t(x_{1t} - E(x_{1t}|x_{2t}(\alpha_0)))'\right], \tag{4.3}$$

$$\Omega(\theta_0, h_0) = \lim_{T\to\infty} \mathrm{var}(T^{-1/2}\sum_{t=1}^{T} w_t[y_t - E(y_t|x_{2t}(\alpha_0)) - (x_{1t} - E(x_{1t}|x_{2t}(\alpha_0)))\theta_0]),$$

$$\Omega^e(\theta_0, h_0) = \lim_{T\longrightarrow\infty} \mathrm{var}\left\{\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(w_t\varepsilon_t + E\left[\frac{w_t}{f(x_{2t}(\alpha_0))}\partial_\alpha[f(x_{2t})h_0(x_{2t}(\alpha_0), \theta_0)] - \frac{w_t[h_0(x_{2t}(\alpha_0), \theta_0)]}{f(x_{2t}(\alpha_0))}\partial_\alpha f(x_{2t}(\alpha_0))\right] r(v_t)' x_{2t}(\alpha_0)\right)\right\},$$

where $h(x, \theta) := E[y_t - x'_{1t}\theta | x_{2t} = x]$ and $x_{2t}(\alpha_0) = s_t - v'_t\alpha_0$.

Proposition 4.1 generalizes some of the results of Li and Wooldridge (2002) to the possibly overidentified partial linear models with α-mixing errors. Note that in case of martingale difference errors, the above result simplifies to

$$\Omega(\theta_0, h_0) = E\left[w_t w'_t (y_t - E(y_t|x_{2t}(\alpha_0)) - (x_{1t} - E(x_{1t}|x_{2t}(\alpha_0)))\theta_0)^2\right],$$

$$\Omega^e(\theta_0, h_0) = \Omega(\theta_0, h_0) + E\left\{\frac{w_t}{f(x_{2t}(\alpha_0))}\partial_\alpha[f(x_{2t}(\alpha_0))h_0(x_{2t}(\alpha_0), \theta_0) - h(x_{2t}(\alpha_0), \theta_0)\partial_\alpha f(x_{2t}(\alpha_0))]\right\} \times E\left[r(v_t)'r(v_t)x_{2t}^2(\alpha_0)\right] E\left\{\frac{w_t}{f(x_{2t}(\alpha_0))}\partial_\alpha[f(x_{2t}(\alpha_0))h_0(x_{2t}(\alpha_0), \theta_0) - h_0(x_{2t}(\alpha_0), \theta_0)\partial_\alpha f(x_{2t}(\alpha_0))]\right\}'.$$

Let $\tau(x_{2t}(\alpha_0)) := \mathbb{I}(x_{2t}(\alpha_0) \in \mathcal{X}_2^*)$ denote a fixed trimming function that equals one whenever $x_{2t}(\alpha_0) \in \mathcal{X}_2^*$ and zero otherwise; then given the results of Proposition (4.1) the proposed two-step semiparametric GEL, GMM, s-GMM and ETEL estimators can be based on the following trimmed smoothed criterion functions

$$
\begin{aligned}
\Gamma^{\text{GEL}}(\theta,\widehat{h},\lambda) &= \sum_{t=1}^{T} \tau\left(\widehat{x}_{2t}\right)\left[\rho(\omega\lambda' g_{ts}(\theta,\widehat{h})) - \rho\left(0\right)\right],\\
\Gamma^{\text{GMM}}(\theta,\widehat{h},\lambda) &= \parallel \tau\left(\widehat{x}_{2t}\right)\widehat{g}(\theta,\widehat{h}) \parallel_{\widehat{\Omega}^{e}(\widetilde{\theta},\widehat{h})^{-1}},\\
\Gamma^{\text{s-GMM}}(\theta,\widehat{h},\lambda) &= \parallel \tau\left(\widehat{x}_{2t}\right)\widehat{g}_{s}(\theta,\widehat{h}) \parallel_{\widehat{\Omega}^{e}(\widetilde{\theta},\widehat{h})^{-1}},\\
\Gamma^{\text{ETEL}}(\theta,\widehat{h},\lambda) &= \log\left\{\frac{1}{T}\sum_{t=1}^{T}\tau\left(\widehat{x}_{2t}\right)\exp[\lambda' g_{ts}^{c}(\theta,\widehat{h})]\right\},
\end{aligned}
$$

where $\widehat{x}_{2t} = x_{2t}(\widehat{\alpha})$ and $\widehat{\Omega}^{e}(\widetilde{\theta},\widehat{h})$ is a consistent estimator of $\Omega^{e}(\theta_0,h_0)$.

5 Monte Carlo Results

In this section we present results for the partial linear regression model with endogenous covariates in its parametric component discussed in Section 4. Specifically, we focus on

$$
\begin{aligned}
y_t &= x_{11t}\theta_{10} + x_{12t}\theta_{20} + m_0(x_{2t}) + \varepsilon_t,\\
x_{11t} &= \pi_{10}v_{1t} + \pi_{20}v_{2t} + u_t,
\end{aligned}
$$

where $v_{1t} = \rho_1 v_{1t-1} + \epsilon_{1t}$, $v_{2t} = \rho_2 v_{2t-1} + \epsilon_{2t}$, $\varepsilon_t = \rho_\varepsilon \varepsilon_{t-1} + \epsilon_{\varepsilon t}$, $u_t = \rho_u u_{t-1} + \epsilon_{ut}$ and

$$
\begin{bmatrix}\epsilon_{1t}\\ \epsilon_{2t}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right),\quad \begin{bmatrix}\epsilon_{\varepsilon t}\\ \epsilon_{ut}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \rho_{\varepsilon u}\\ \rho_{\varepsilon u} & 1\end{bmatrix}\right).
$$

Let $\omega_{lt} \sim N\left(0,1\right)$ $(l = 2,3,4)$ independent of v_{1t} and v_{2t}, and set $x_{12t} = v_{2t} + \omega_{2t}$, $x_{2t} = v_{1t} + v_{2t} + \omega_{3t}$ such that $s_t = \omega_{4t}\alpha_0 + x_{2t}$. For $\rho_1 = \rho_2 = 0.5$, $\rho_\varepsilon = \rho_u = 0.95$, and $m_0\left(v\right) = \Phi\left(v\right)$ ($\Phi\left(\cdot\right)$ is the CDF of a standard normal), we generate 2000 samples, $\{y_t, x_{11t}, x_{12t}, s_t, \omega_{4t}, v_{1t}, v_{2t}\}_{t=1}^{T}$, with $T \in \{200, 400, 800\}$, two different scenarios $\rho_{\varepsilon u} \in \{0.1, 0.9\}$ representing an increasing degree of endogeneity and $\theta_0 = [1,1]'$, $\pi_0 = [1,-1]'$, $\alpha_0 = 1$.

Let $z_t := [y_t, x_{11t}, x_{12t}, \widehat{x}_{2t}, v_{1t}, v_{2t}]'$, $w_t := [x_{12t}, \widehat{x}_{2t}, v_{1t}, v_{2t}]'$, $h_0\left(z_t\right) := [\widetilde{y}_t, \widetilde{x}_{11t}, \widetilde{x}_{12t}]'$, $\widetilde{y}_t := y_t - \widehat{E}[y_t|\widehat{x}_{2t}]$, $\widetilde{x}_{11t} := x_{1t} - \widehat{E}[x_{11t}|\widehat{x}_{2t}]$, $\widetilde{x}_{12t} := x_{12t} - \widehat{E}[x_{12t}|\widehat{x}_{2t}]$ and $\widehat{x}_{2t} := s_t - \omega_{4t}\widehat{\alpha}$, so that

$$
g_t(\theta,\widehat{h}) = w_t(\widetilde{y}_t - \widetilde{x}_{11t}\theta_1 - \widetilde{x}_{12t}\theta_2),
$$

where $\widehat{h}$ is the Nadaraya-Watson estimator with bandwidths chosen as $c \in \{0.5, 1, 1.5\}$ times the Silverman's rule-of-thumb bandwidth, and $\widehat{\alpha}$ is an estimator of α_0 obtained from regressing s_t on ω_{4t} by ordinary least squares.

The GEL estimators we consider are the Empirical Likelihood (EL), Exponential Tilting (ET) and

Continuous Updated (CU) estimators; for the GMM estimators we use the following estimator

$$\widehat{\Omega}^e(\widetilde{\theta},\widehat{h}) = \left[\frac{1}{s_T}\sum_{j=1-T}^{T-1}\omega^2\left(\frac{j}{s_T}\right)\right]^{-1}\left(\frac{s_T}{T}\right)\tau\left(\widehat{x}_{2t}\right)\sum_{t=1}^{T} g_{ts}(\widetilde{\theta},\widehat{h})g_{ts}(\widetilde{\theta},\widehat{h})', \tag{5.1}$$

$$g_{ts}(\widetilde{\theta},\widehat{h}) = \frac{1}{s_T}\sum_{j=t-T}^{t-1}\omega\left(\frac{j}{s_T}\right)\left\{w_t\widetilde{\varepsilon}_t + \frac{1}{T}\sum_{t=1}^{T}\left[\frac{w_t}{\widehat{f}(\widehat{x}_{2t})}\partial_\alpha\widehat{f}(\widehat{x}_{2t})\widehat{h}(\widehat{x}_{2t},\widetilde{\theta}) - \frac{w_t[\widehat{h}(\widehat{x}_{2t},\widetilde{\theta})]}{\widehat{f}(\widehat{x}_{2t})}\partial_\alpha f(\widehat{x}_{2t})\right]\widehat{r}(\omega_{4t})\widehat{x}_{2t}\right\},$$

where $\widetilde{\varepsilon}_t = \widetilde{y}_t - \widetilde{x}_{11t}\widetilde{\theta}_1 - \widetilde{x}_{12t}\widetilde{\theta}_2$, $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ are preliminary consistent estimators of θ_{10} and θ_{20}, $\widehat{f}(\widehat{x}_{2t})$ is a kernel estimator of the marginal density of $\widehat{x}_{2t}$ and $\widehat{r}(\omega_{4t}) = \omega_{4t}/\left(\sum_{t=1}^{T}\omega_{4t}^2/T\right)$. In the Monte Carlo we use a Bartlett smoothing kernel with bandwidth parameter s_T chosen by the method suggested in Andrews (1991). The same bandwidths and kernels are used to estimate the asymptotic standard errors based on (4.3) and to compute the estimator $\widehat{\Omega}^e(\widetilde{\theta},\widehat{h})$ given in (5.1).

The Monte Carlo Bias (Bias), Standard Deviation (Std. Dev.), Average Ratios of Standard Errors (Ratio) with respect to that of a standard GMM and Coverage Probability (Cov. Prob.) are reported in Tables 1-2 for the estimator of the endogenous regressor parameter θ_{10}. We use the standard GMM partly because of its efficiency property discussed in Remark (3.2) and partly because it would probably be the most popular estimator given its (relatively) computational simplicity.

Tables 1 and 2 approx. here

We first consider the bias reported for the estimator of the endogenous regressor parameter and note that the bandwidth choice has some finite sample effect especially for $T = 200$ and 400, but it is also important to note that the magnitude of the bias of all of the proposed estimators is statistically insignificant. As expected, the degree of endogeneity has some negative effect on the bias for the smaller sample sizes. Second the standard and smoothed efficient GMM estimators are characterized by the largest bias but smallest standard deviations, whereas the EL estimator has the smallest bias, especially in the case of low endogeneity. Turning to the Monte Carlo standard deviation, we first note that in this case the degree of endogeneity have a less significant finite sample effect. Second the standard and smoothed GMM estimators seem to have an edge compared to the other estimators especially for $T = 200$ and 400. Third, as pointed out in Remark 3.2, the standard and smoothed GMM estimators have the smallest standard errors. Finally we note that the asymptotic approximation of all estimators seem appropriate for small samples as measured by the Monte Carlo coverage probability.

Figures 1-2 report the Q-Q plots that are used to illustrate the quality of the asymptotic normal approximation for the estimator of the exogenous regressor parameter θ_{20}.

Figures 1 and 2 approx. here

The figures show that the asymptotic approximation is good across models especially for samples $T = 400$ and 800 for all estimators across low and high degrees of endogeneity. The approximation improves with the sample size and seems to be robust to bandwidth choice for the first step estimator. Taking these results together, they suggest that the smoothed two-step estimators we are proposing seem to be characterized by good finite sample properties.

6 Conclusions

In this paper we consider the problem of estimating parameters of interest in semiparametric moment condition models with dependent data. We propose two-step GMM, GEL and ETEL estimators for the finite dimensional parameter and use smoothing to take the dependency into consideration. We show that as long as there is no estimation effect from the first step estimation all of the proposed estimators are asymptotically equivalent to the efficient GMM estimator of Hansen (1982). On the other hand, when there is estimation effect, this equivalence does not hold any longer for GEL and ETEL estimators, which become less efficient. Our proofs rely on a new uniform law of large numbers that generalizes that of Andrews' (1987) and use two new CLT's for both degenerate and non-degenerate second-order U-statistics with varying kernels. These results are of independent interest. We illustrate the results with an instrumental variable partial linear model with a nonparametric generated regressor and use simulations to assess the finite sample properties of some of the proposed estimators. The results of the simulations suggest that overall all of the proposed estimators have good finite sample properties. Finally, we would like to mention that the results of this paper could be readily used in the context of quadratic inference functions for certain type of longitudinal data structures $\{z_{it_i}, i = 1, ...n,\ t_i = 1, ..., T\}$. In particular, under the additional assumption that the data are independent and identically distributed across i for fixed t_i, and are α-mixing with the same mixing coefficient as that given in Assumption 1 for a fixed i, it can be shown that the conclusion of Theorem 3.2 is still valid for an appropriately smoothed version of the quadratic inference function $g(z_{it_i}, \theta, h)$. The case for Theorem 3.3 is considerably more complicated and we leave it for future research.

Acknowledgements

We thank an Associate Editor and two anonymous referees for various useful suggestions that improve the readability and clarity of the paper. We also acknowledge the usage of the `np` and `gmm` packages by Hayfield and Racine (2008) and Chaussé (2010) respectively in the statistical computing environment `R`. Chu and Jacho-Chávez gratefully acknowledge support by the Social Science and Humanities Research Council of Canada grant (MBF Grant 410-2011-1700).

References

Adams, T.M., and Nobel, A.B. (2010), 'Uniform Convergence of Vapnik-Čhervonenkis Classes under Ergodic Sampling', *Annals of Probability*, 38, 1345–1367.

Andrews, D.W.K. (1987), 'Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers', *Econometrica*, 55, 1465–1471.

Andrews, D.W.K. (1991), 'Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation', *Econometrica*, 59, 817–858.

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

Andrews, D.W.K. (1994b), 'Empirical Process Methods in Econometrics', in *Handbook of Econometrics, Vol. IV*, eds. R. Engle and D. McFadden. New York: North Holland, pp. 2247–2294.

Andrews, D.W.K. (1995), 'Nonparametric Kernel Estimation for Semiparametric Models', *Econometric Theory*, 11, 560–596.

Boente, G., and Fraiman, R. (1988), 'Consistency of a Nonparametric Estimate of a Density Function for Dependent Variables', *Journal of Multivariate Analysis*, 25, 90–99.

Bonhomme, S., and Manresa, E. (2015), 'Grouped Patterns of Heterogeneity in Panel Data', *Econometrica*, 83, 1147–1184.

Bravo, F., Chu, B.M., and Jacho-Chávez, D.T. (2016), 'Generalized Empirical Likelihood M Testing for Semiparametric Models with Time Series Data', Unpublished manuscript.

Chamberlain, G. (1987), 'Asymptotic Efficiency in Estimation with Conditional Moment Restrictions', *Journal of Econometrics*, 34, 305–344.

Chaussé, P. (2010), 'Computing Generalized Method of Moments and Generalized Empirical Likelihood with R', *Journal of Statistical Software*, 34, 1–35.

Chen, X., and Shen, X. (1998), 'Sieve Extremum Estimates for Weakly Dependent Data', *Econometrica*, 66, 289–314.

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

Chen, X., Jacho-Chávez, D.T., and Linton, O. (2016), 'Averaging of an Increasing Number of Moment Condition Estimators', *Econometric Theory*, 32, 30–70.

Chu, B.M., and Jacho-Chávez, D.T. (2012), 'k-Nearest Neighbour Estimation of Inverse-Density-Weighted Expectations with Dependent Data', *Econometric Theory*, 28, 769–803.

Chu, B.M., Huynh, K.P., and Jacho-Chávez, D.T. (2013), 'Functionals of Order Statistics and their Multivariate Concomitants with Application to Semiparametric Estimation by Nearest Neighbors', *Sankhya: Series B*, 75, 238–292.

de Jong, P. (1987), 'A Central Limit Theorem for Generalized Quadratic Forms', *Probability Theory and Related Fields*, 75, 261–277.

Doukhan, P. (1994), *Mixing: Properties and Examples*, Lecture Notes in Statistics, Vol. 85, New York: Springer & Verlag.

Doukhan, P., Massart, P., and Rio, E. (1994), 'The Functional Central Limit Theorem for Strongly Mixing Processes', *Annales de L'Institut Henri Poincaré (B) – Probabilités et Statistiques*, 30, 63–82.

Dvoretsky, A. (1972), 'Asymptotic Normality for Sums of Dependent Random Variables', in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, eds. L.M.L. Cam, J. Neyman, and E.L. Scott. University of California, pp. 513–535.

Escanciano, J.C., Jacho-Chávez, D.T., and Lewbel, A. (2014), 'Uniform Convergence of Weighted Sums of Non- and Semi-parametric Residuals for Estimation and Testing', *Journal of Econometrics*, 178, 426–443.

Escanciano, J.C., Jacho-Chávez, D.T., and Lewbel, A. (2016), 'Identification and Estimation of Semiparametric Two Step Models', *Quantitative Economics*, 7, 561–589.

Fan, Y., and Li, Q. (1999), 'Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing', *Journal of Nonparametric Statistics*, 10, 245–271.

Gao, J. (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman and Hall/CRC.

Gao, J., and Liang, H. (1997), 'Statistical Inference in Single-Index and Partially Nonlinear Models', *Annals of the Institute of Statistical Mathematics*, 49, 493–517.

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

Hayfield, T., and Racine, J.S. (2008), 'Nonparametric Econometrics: The np Package', *Journal of Statistical Software*, 27, 1–32.

Kitamura, Y. (1997), 'Empirical Likelihood Methods with Weakly Dependent Processes', *Annals of Statistics*, 25, 2084–2102.

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

Kong, E., Linton, O.B., and Xia, Y. (2010), 'Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model', *Econometric Theory*, 26, 1529–1564.

Lee, L. (1994), 'Semiparametric Instrumental Variable Estimation of Simultaneous Equation Sample Selection Models', *Journal of Econometrics*, 63, 341–388.

Li, Q., and Racine, J.S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Li, Q., and Wooldridge, J. (2002), 'Semiparametric Estimation for Partially Linear for Dependent Data with Generated Regressors', *Econometric Theory*, 18, 625–645.

Liebscher, E. (1998), 'Estimation of the Density and the Regression Function under Mixing Conditions', *Statistics & Decisions*, 19, 9–26.

Mammen, E., Rothe, C., and Schienle, M. (2015), 'Semiparametric Estimation with Generated Covariates', *Econometric Theory*, 32, 1140–1177.

Masry, E. (1996), 'Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates', *Journal of Time Series Analysis*, 17, 571–599.

Mikosch, T. (1993), 'A Weak Invariance Principle for Weighted U-Statistics with Varying Kernels', *Journal of Multivariate Analysis*, 47, 82–102.

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

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

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

Powell, J., Stock, J., and Stoker, T. (1989), 'Semiparametric Estimation of Weighted Average Derivatives', *Econometrica*, 57, 1403–1430.

Robinson, P. (1987), 'Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form', *Econometrica*, 55, 875–891.

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

Smith, R.J. (1997), 'Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation', *Economic Journal*, 107, 503–519.

Smith, R.J. (2005), 'Automatic Positive HAC Covariance Matrix and GMM Estimation', *Econometric Theory*, 21, 158–170.

Smith, R.J. (2011), 'GEL Criteria for Moment Condition Models', *Econometric Theory*, 27, 1192–1235.

Sun, S., and Chiang, C. (1997), 'Limiting Behaviour of the Perturbed Empirical Distribution Functions Evaluated at U-Statistics for Strongly Mixing Sequences of Random Variables', *Journal of Applied Mathematics and Stochastic Analysis*, 10, 3–20.

Truong, Y., and Stone, C. (1992), 'Nonparametric Function Estimation Involving Time Series', *Annals of Statistics*, 20, 77–97.

Truong, Y.K., and Stone, C.J. (1994), 'Semiparametric Time Series Regression', *Journal of Time Series Analysis*, 15, 405–428.

van der Vaart, A.W., and Wellner, J.A. (1996), *Weak Convergence and Empirical Processes*, New York, Berlin: Springer.

Yoshihara, K. (1976), 'Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes', *Probability Theory and Related Fields*, 35, 237–252.

Yoshihara, K. (1989), 'Limiting Behavior of Generalized Quadratic Forms Generated by Absolutely Regular Sequences II', *Yokohama Mathematical Journal*, 37, 109–123.

Yoshihara, K. (1992), 'Limiting Behavior of Generalized Quadratic Forms Generated by Regular Sequences III', *Yokohama Mathematical Journal*, 40, 1–9.

Yu, B. (1993), 'Density Estimation in the L_∞ Norm for Dependent Data with Application to the Gibbs sampler', *Annals of Statistics*, 21, 711–735.

Yu, B. (1994), 'Rates of Convergence for Empirical Processes of Stationary Mixing Sequences', *Annals of Probability*, 22, 94–116.

Appendix A Main Proofs

Throughout this section "FOC" and "CMT" stand for, respectively, First Order Conditions and Continuous Mapping Theorem; unless otherwise stated "CLT" denotes a Central Limit Theorem for α-mixing sequences (see for example Doukhan, 1994, Chapter 1.5). C and $C(\cdot)$ represent generic constants that may depend on additional quantities and may be different from line to line.

Proof of Theorem 3.1: See the supplemental material to this paper. ■

Proof of Theorem 3.2: We first show the consistency of $\widehat{\theta}$ and $\widehat{\lambda}$ for the GEL criterion function. Without loss of generality we normalize the first two derivatives $\rho_j(0) = -1$ $(j = 1, 2)$ of $\rho(\cdot)$, where $\rho_j(0) := \partial^j \rho(q) / \partial q^j|_{q=0}$. Let $\Lambda_T^r = \{\lambda : \|\lambda\| \leq R_T\}$ where $R_T = O_p(s_T/T)^\xi$ for $\xi < 1/2$; as in Smith (2011) it suffices to show that

$$\sup_{\theta\in\Theta} \| \widehat{g}_{ts}(\theta, \widehat{h}) - \omega_1 E[g_t(\theta, h_0)] \| = o_p(1), \tag{A-1}$$

$$\max_{1\leq t\leq T} \sup_{\lambda\in\Lambda_T^r} \sup_{\theta\in\Theta} \| \lambda' g_{ts}(\theta, \widehat{h}) \| = o_p(1), \tag{A-2}$$

$$\left\| \left(\frac{1}{s_T} \sum_{t=1-T}^{T-1} \omega \left(\frac{t}{s_T} \right)^2 \right)^{-1} \frac{s_T}{T} \sum_{t=1}^{T} g_{ts}(\widehat{\theta}, \widehat{h}) g_{ts}(\widehat{\theta}, \widehat{h})' - \Omega(\theta_0, h_0) \right\| = o_p(1). \tag{A-3}$$

To verify (A-1) note that by triangle inequality, Theorem 3.1 and dominated convergence

$$\sup_{\theta\in\Theta} \| \widehat{g}_s(\theta, \widehat{h}) - \omega_1 E[g_t(\theta, h_0)] \| \le \left| \sum_{j=1-T}^{T-1} \frac{1}{s_T} \omega\left(\frac{j}{s_T}\right) \right| \sup_{\theta\in\Theta, h\in\mathcal{H}_\delta} \left\| \frac{1}{T} \sum_{t=1}^{T} g_t(\theta, h) - E[g_t(\theta, h)] \right\| +$$
$$\left| \sum_{j=1-T}^{T-1} \frac{1}{s_T} \omega\left(\frac{j}{s_T}\right) - \omega_1 \right| E \sup_{\theta\in\Theta, h\in\mathcal{H}_\delta} \| g_t(\theta, h) \| +$$
$$+ \omega_1 \sup_{\theta\in\Theta} \left\| E[g_t(\theta, \widehat{h})] - E[g_t(\theta, h_0)] \right\| = o_p(1),$$

since $\left| \sum_{j=1-T}^{T-1} s_T^{-1} \omega(j/s_T) - \omega_1 \right| \to 0$. To show (A-2) note that by triangle inequality and the (functional) mean value theorem one has

$$\max_{1\le t\le T} \sup_{\lambda\in\Lambda_T^r} \sup_{\theta\in\Theta} \| \lambda' g_{ts}(\theta, \widehat{h}) \| \le R_T \left| \sum_{j=1-T}^{T-1} \frac{1}{s_T} \omega\left(\frac{j}{s_T}\right) \right| \times$$
$$\max_{1\le t\le T} \sup_{\theta\in\Theta} \left[\| g_t(\theta, h_0) \| + \sup_{h\in\mathcal{H}_\delta} \| \partial_h g_t(\theta, h) \| \, \| \widehat{h} - h_0 \|_{\mathcal{H}} \right] = o_p(1),$$

by Assumptions 4(a0-(b) and 5(a) since $\max_{1\le t\le T} \sup_{\theta\in\Theta} \| g_t(\theta, h_0) \| = o_{a.s}\left(T^{1/\mu}\right)$ and

$$\max_{1\le t\le T} \sup_{\theta\in\Theta} \sup_{h\in\mathcal{H}_\delta} \| \partial_h g_t(\theta, h) \| = o_{a.s}\left(T^{1/2}\right)$$

by the Borel-Cantelli lemma. Finally, to show (A-3), it follows from the triangle inequality

$$\left\| \frac{s_T}{T} \sum_{t=1}^{T} g_{ts}(\widehat{\theta}, \widehat{h}) g_{ts}(\widehat{\theta}, \widehat{h})' - \omega_2 \Omega(\theta_0, h_0) \right\|$$
$$\le \left\| \frac{s_T}{T} \sum_{t=1}^{T} g_{ts}(\theta_0, h_0) g_{ts}(\theta_0, h_0)' - \omega_2 \Omega(\theta_0, h_0) \right\| + 2 \left\| \frac{s_T}{T} \sum_{t=1}^{T} g_{ts}(\theta_0, h_0) [g_{ts}(\widehat{\theta}, \widehat{h}) - g_{ts}(\theta_0, h_0)]' \right\|$$
$$+ \left\| \frac{s_T}{T} \sum_{t=1}^{T} [g_{ts}(\widehat{\theta}, \widehat{h}) - g_{ts}(\theta_0, h_0)][g_{ts}(\widehat{\theta}, \widehat{h}) - g_{ts}(\theta_0, h_0)]' \right\| = \mathfrak{T}_1^* + \mathfrak{T}_2^* + \mathfrak{T}_3^*.$$

$\mathfrak{T}_1^* = o_p(1)$ by Lemma A.3 of Smith (2011). Calculations along the lines of Lemma A.3 of Smith (2011) and Cauchy-Schwarz inequality yield

$$\mathfrak{T}_2^* \le \left| \frac{1}{s_T} \sum_{s=1-T}^{T-1} \omega\left(\frac{t-s}{s_T}\right) \omega\left(\frac{t}{s_T}\right) \right| \left[\| \widehat{g}(\theta_0, h_0) \|_2 \, \| \widehat{g}(\widehat{\theta}, \widehat{h}) - \widehat{g}(\theta_0, h_0) \|_2 + O\left(\frac{t}{T}\right) \right],$$

and by the functional mean value theorem, Assumptions 4 and 5(a)

$$\frac{1}{T} \sum_{t=1}^{T} \left\| g_t(\widehat{\theta}, \widehat{h}) - g_t(\theta_0, h_0) \right\|^2 \le \| \widehat{h} - h_0 \|_{\mathcal{H}}^2 \sup_{\theta\in\Theta_\delta, h\in\mathcal{H}_\delta} \left| \frac{1}{T} \sum_{t=1}^{T} \| \partial_h g_t(\theta, h) \|^2 \right| = o_p(1),$$

hence $\mathfrak{T}_2^* = o_p(1)$ since

$$\lim_{T\to\infty}\left|\frac{1}{s_T}\sum_{t=1-T}^{T-1}\frac{t}{T}\omega\left(\frac{t-s}{s_T}\right)\omega\left(\frac{t}{s_T}\right)\right| = 0$$

by Lemma C.1 of Smith (2011). Similar arguments yield $\mathfrak{T}_3^* = o_p(1)$. Clearly $\Pr(\Lambda_T^r \in \Lambda_T) \to 1$ and note that by (A-2) and CMT

$$\sup_{\theta\in\Theta,\ \lambda\in\Lambda_T^r}\max_{1\le t\le T}\left|\rho_j(\lambda' g_{ts}(\theta,\widehat{h})) - \rho_j(0)\right| = o_p(1),\ j = 1,2. \tag{A-4}$$

Given (A-1)-(A-4), the consistency of the GEL estimator $\widehat{\theta}$ follows by the same arguments of Newey and Smith (2004) and Smith (2011). First note that

$$\sup_{\lambda\in\Lambda_T^r}\frac{1}{s_T}\Gamma(\theta_0,\widehat{h},\lambda) \le \|\widehat{g}_s(\theta_0,h_0)\|^2 + \frac{1}{s_T}\left[\Gamma(\theta_0,\widehat{h},\lambda) - \Gamma(\theta_0,h_0,\lambda)\right], \tag{A-5}$$

and that by a Taylor expansion along the continuous connected path $h^*(\epsilon) = h_0 + \epsilon(\widehat{h} - h_0)$ such that $h^*(\epsilon) \in \mathcal{H}_\delta$, $\forall\epsilon\in[0,1]$ we have

$$\Gamma(\theta_0,\widehat{h},\lambda) = \Gamma(\theta_0,h_0,\lambda) + \omega(1-\overline{\epsilon})\frac{1}{T}\sum_{t=1}^{T}\rho_1\left(\omega\lambda' g_{ts}(\theta_0,h^*(\overline{\epsilon}))\right)\lambda'\partial_h g_{ts}(\theta_0,h^*(\overline{\epsilon}))(\widehat{h} - h_0),$$

where $\overline{\epsilon}\in(0,1)$. Then by triangle inequality

$$\left|\Gamma(\theta_0,\widehat{h},\lambda) - \Gamma(\theta_0,h_0,\lambda)\right| \le \left|(1-\overline{\epsilon})\omega\frac{1}{T}\sum_{t=1}^{T}\lambda'\partial_h g_{ts}(\theta_0,h^*(\overline{\epsilon}))(\widehat{h}-h_0)\right| +$$

$$\left|(1-\overline{\epsilon})\omega\frac{1}{T}\sum_{t=1}^{T}\left[\rho_1\left(\omega\lambda' g_{ts}(\theta_0,h^*(\overline{\epsilon}))\right) - \rho_1(0)\right]\lambda'\partial_h g_{ts}(\theta_0,h^*(\overline{\epsilon}))(\widehat{h}-h_0)\right| \le$$

$$R_T \parallel \widehat{h} - h_0 \parallel_{\mathcal{H}} \sup_{h\in\mathcal{H}_\delta}\frac{1}{T}\sum_{t=1}^{T}\|\partial_h g_{ts}(\theta_0,h)\|\left|\sup_{\lambda\in\Lambda_T^r,\ h\in\mathcal{H}_\delta}\left|\rho_1\left(\omega\lambda' g_{ts}(\theta_0,h)\right) - \rho_1(0)\right|\right| = o_p(1)$$

by Assumptions 4 and 5(a). Let $\lambda_T = -\widehat{g}(\widehat{\theta},\widehat{h})\xi_T / \parallel \widehat{g}(\widehat{\theta},\widehat{h}) \parallel$ where $|\xi_T| < R_T$, so that $\Pr(\lambda_T\in\Lambda_T^r)\to 1$. A Taylor expansion of $\Gamma(\widehat{\theta},\widehat{h},\lambda_T)$ with respect to $\lambda_T' g_{ts}(\widehat{\theta},\widehat{h})$ about 0 gives

$$\Gamma(\widehat{\theta},\widehat{h},\lambda_T) \ge -\omega\lambda_T'\widehat{g}_s(\widehat{\theta},\widehat{h}) - C\omega^2\lambda_T'\lambda_T = \omega\xi_T \parallel \widehat{g}(\widehat{\theta},\widehat{h}) \parallel - C\omega^2\xi_T^2.$$

Since

$$\Gamma(\widehat{\theta},\widehat{h},\lambda_T) \le \sup_{\lambda\in\Lambda_T^r}\Gamma(\widehat{\theta},\widehat{h},\lambda) \le \sup_{\lambda\in\Lambda_T^r}\Gamma(\theta_0,\widehat{h},\lambda),$$

we have by (A-5), the CLT and some algebra yield

$$\parallel \widehat{g}(\widehat{\theta},\widehat{h}) \parallel \le \frac{s_T}{\xi_T}\|\widehat{g}(\theta_0,h_0)\|^2 + o_p(1) = o\left(T^{1/2}\right)O_p\left(T^{-1}\right),$$

$\parallel \widehat{g}(\widehat{\theta},\widehat{h}) \parallel = o_p(1)$. The consistency of $\widehat{\theta}$ follows now by Lemma C.1 and the identification condition (2.1). A similar expansion can be used to show that $\parallel \widehat{\lambda} \parallel = O_p\left(s_T/T^{1/2}\right)$ where $\widehat{\lambda} = \arg\max_{\lambda\in\Lambda_T^r}\Gamma_\tau(\widehat{\theta},\widehat{h},\lambda)$. The asymptotic distribution is obtained by a standard mean value expansion of the FOC

$$0 = \left[\partial_\theta\Gamma(\widehat{\theta},\widehat{h},\lambda), \partial_\lambda\Gamma(\widehat{\theta},\widehat{h},\lambda)\right]'$$

that hold with probability $\to 1$ by (a), and gives

$$T^{1/2}[(\widehat{\theta}-\theta_0)',\widehat{\lambda}/s_T]' = \frac{1}{\omega_1}\widehat{M}(\overline{\lambda},\overline{\theta},\widehat{h})^{-1}[0',-T^{1/2}\widehat{g}_s(\theta_0,\widehat{h})']' + o_p(1),$$

where

$$\widehat{M}(\overline{\lambda},\overline{\theta},\widehat{h}) = \begin{bmatrix} 0 & \frac{1}{T}\sum_{t=1}^{T}\rho_1(\omega\overline{\lambda}'g_{ts}(\overline{\theta},\widehat{h}))\partial_\theta g_{ts}(\overline{\theta},\widehat{h}) \\ \frac{1}{T}\sum_{t=1}^{T}\rho_1(\omega\overline{\lambda}'g_{ts}(\overline{\theta},\widehat{h}))\partial_\theta g_{ts}(\overline{\theta},\widehat{h}) & \frac{s_T}{T}\sum_t \rho_2(\omega\overline{\lambda}'g_{ts}(\overline{\theta},\widehat{h}))g_{ts}(\overline{\theta},\widehat{h})g_{ts}'(\overline{\theta},\widehat{h}) \end{bmatrix}.$$

Note that

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\left[\rho_1(\omega\overline{\lambda}'g_{ts}(\overline{\theta},\widehat{h}))-\rho_1(0)\right]\partial_\theta g_{ts}(\overline{\theta},\widehat{h}) + \rho_1(0)\frac{1}{T}\sum_{t=1}^{T}\partial_\theta g_{ts}(\overline{\theta},\widehat{h}) \le \\ &\sup_{\lambda\in\Lambda_T^r,\ h\in\mathcal{H}_\delta}\left|\rho_1\left(\omega\lambda' g_{ts}(\theta_0,h)\right)-\rho_1(0)\right|\left\|\frac{1}{T}\sum_{t=1}^{T}\partial_\theta g_{ts}(\overline{\theta},\widehat{h})\right\| + \\ &\left\|\frac{1}{T}\sum_{t=1}^{T}\partial_\theta g_{ts}(\overline{\theta},\widehat{h}) - \frac{1}{T}\sum_{t=1}^{T}\partial_\theta g_{ts}\left(\overline{\theta},h_0\right)\right\|,\end{aligned}$$

and

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\partial_\theta g_{ts}(\overline{\theta},\widehat{h}) - \frac{1}{T}\sum_{t=1}^{T}\partial_\theta g_{ts}\left(\overline{\theta},h_0\right)\right\| \le \|\,\widehat{h}-h_0\,\|_{\mathcal{H}}\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta}\frac{1}{T}\sum_{t=1}^{T}\left\|\partial^2_{\theta h}g_{ts}(\theta,h)\right\| = o_p(1).$$

Thus by (A-5), Theorem 3.1, condition 5(a) and the CMT

$$\frac{1}{T}\sum_{t=1}^{T}\rho_1(\omega\overline{\lambda}'g_{ts}(\overline{\theta},\widehat{h}))\partial_\theta g_{ts}(\overline{\theta},\widehat{h}) \xrightarrow{p} \omega_1 E\left[\partial_\theta g_t(\theta_0,h_0)\right].$$

Similarly note that

$$\sup_{\lambda\in\Lambda_T^r,h\in\mathcal{H}_\delta}\left|\rho_2\left(\omega\lambda' g_{ts}(\theta_0,h)\right)-\rho_2(0)\right| = o_p(1), \tag{A-6}$$

and

$$\begin{aligned}&\left\|\frac{s_T}{T}\sum_{t=1}^{T}\rho_2(\omega\overline{\lambda}'g_{ts}(\overline{\theta},\widehat{h}))g_{ts}(\overline{\theta},\widehat{h})g_{ts}(\overline{\theta},\widehat{h})' - \frac{s_T}{T}\sum_{t=1}^{T}\rho_2(0)\,g_{ts}(\theta_0,h_0)\,g_{ts}(\theta_0,h_0)'\right\| \qquad \text{(A-7)}\\ &\le \left\|\overline{\theta}-\theta_0\right\|\left[\frac{1}{T}\sup_{\theta\in\Theta_\delta,\ h\in\mathcal{H}_\delta}\sum_{t=1}^{T}\|g_t(\theta,h)\|^2\right]^{1/2}\left[\frac{1}{T}\sup_{\theta\in\Theta_\delta,\ h\in\mathcal{H}_\delta}\sum_{t=1}^{T}\|\partial_\theta g_t(\theta,h)\|^2\right]^{1/2} + \\ &\|\,\widehat{h}-h_0\,\|_{\mathcal{H}}\left[\frac{1}{T}\sup_{\theta\in\Theta_\delta,\ h\in\mathcal{H}_\delta}\sum_{t=1}^{T}\|g_t(\theta,h)\|^2\right]^{1/2}\left[\frac{1}{T}\sup_{\theta\in\Theta_\delta,\ h\in\mathcal{H}_\delta}\sum_{t=1}^{T}\|\partial_\theta g_t(\theta,h)\|^2\right]^{1/2} \xrightarrow{p} 0,\end{aligned}$$

so that by (A-3), $\|\, s_T T^{-1}\sum_t \rho_2(\omega\overline{\lambda}'g_{ts}(\overline{\theta},\widehat{h}))g_{ts}(\overline{\theta},\widehat{h})g_{ts}(\overline{\theta},\widehat{h})' - \Omega(\theta_0,h_0)\,\| = o_p(1)$. Thus by triangle inequality and the CMT

$$\widehat{M}(\overline{\lambda},\overline{\theta},\widehat{h})^{-1} \xrightarrow{p} \begin{bmatrix}\Sigma(\theta_0,h_0) & H(\theta_0,h_0)\\ H(\theta_0,h_0)' & P(\theta_0,h_0)\end{bmatrix} =: M(\theta_0,h_0)^{-1},$$

where

$$H(\theta_0,h_0)=\Sigma(\theta_0,h_0)\,G(\theta_0,h_0)'\,\Omega(\theta_0,h_0)^{-1}$$
$$P(\theta_0,h_0)=\Omega(\theta_0,h_0)^{-1}\left[I-G(\theta_0,h_0)\,\Sigma(\theta_0,h_0)\,G(\theta_0,h_0)'\,\Omega(\theta_0,h_0)^{-1}\right].$$

Then by Assumptions 5(b) and 6(a) we have

$$T^{1/2}[(\widehat{\theta}-\theta_0)',\widehat{\lambda}'/s_T]'=\frac{1}{\omega_1}M(\theta_0,h_0)^{-1}\left[0',-T^{1/2}\omega_1\widehat{g}(\theta_0,h_0)'\right]'+o_p(1) \tag{A-8}$$

and by CLT and CMT

$$T^{1/2}[(\widehat{\theta}-\theta_0)',\widehat{\lambda}'/s_T]'\xrightarrow{d}N\left(0,\operatorname{diag}\left[\Sigma(\theta_0,h_0),P(\theta_0,h_0)\right]\right).$$

The consistency of the two step smoothed semiparametric GMM based estimator $\widehat{\theta}$, in (2.7), follows by the identification condition (2.1), and the uniform convergence of $\|\widehat{g}_s(\theta,\widehat{h})\|_{\widehat{W}}$, which follows by (A-1), $\|\widehat{W}-W\|=o_p(1)$ for any positive definite matrix W, and

$$\begin{aligned}&\left|\|\widehat{g}_s(\theta,\widehat{h})\|_{\widehat{W}}-\|\omega_1E[g_t(\theta,h_0)]\|_W\right|\le\|\widehat{g}_s(\theta,\widehat{h})-\omega_1E[g_t(\theta,h_0)]\|^2\|\widehat{W}\|+\\&\|\omega_1E[g_t(\theta,h_0)]\|\,\|\widehat{W}-W\|+2\|\omega_1E[g_t(\theta,h_0)]\|\times\\&\|\widehat{g}_s(\theta,\widehat{h})-\omega_1E[g_t(\theta,h_0)]\|\,\|\widehat{W}\|=o_p(1),\end{aligned}$$

by the triangle inequality. The asymptotic normality follows by a standard Taylor expansion about θ_0 of the FOC

$$0=T^{1/2}[\partial_\theta\widehat{g}_s(\widehat{\theta},\widehat{h})]\widehat{W}\widehat{g}_s(\widehat{\theta},\widehat{h})$$

that hold with probability $\to 1$ by assumption (a). The conclusion follows by (2.1) (applied to $\partial_\theta\widehat{g}_s(z_t,\theta,\widehat{h})$), assumption 5(b), CLT and CMT. The consistency of the two-step smoothed semiparametric ETEL estimator $\widehat{\theta}$ follows by a two step argument: First, for any $\overline{\lambda}$ such that $\Pr\left(\overline{\lambda}\in\Lambda_T^r\right)\to 1$, the same arguments as those used to show the consistency of the GEL estimator show that the ETEL estimator

$$\widehat{\theta}=\arg\min_{\theta\in\Theta}\sup_{\lambda\in\Lambda_T^r}\log\left\{\frac{1}{T}\sum_{t=1}^{T}\exp\{\omega\lambda'[g_{ts}(\theta,\widehat{h})-\widehat{g}_s(\theta,\widehat{h})]\}\right\} \tag{A-9}$$

is consistent. Next the consistency of $\widehat{\lambda}$ defined as

$$\widehat{\lambda}:=\arg\max_{\lambda\in\Lambda_T^r}\frac{1}{T}\sum_{t=1}^{T}-\exp\{\omega\widehat{\lambda}'g_{ts}(\theta,\widehat{h})\}$$

follows noting that by a second order Taylor expansion about 0, (A-4) and (A-6) we have

$$\begin{aligned}&0\le 1-\frac{1}{T}\sum_{t=1}^{T}\exp\{\omega\widehat{\lambda}'g_{ts}(\widehat{\theta},\widehat{h})\}\le 1-\frac{1}{T}\sum_{t=1}^{T}\exp\{\omega\widehat{\lambda}'g_{ts}(\theta_0,\widehat{h})\}=\\&\omega\widehat{\lambda}'\widehat{g}_s(\theta_0,\widehat{h})+\frac{\omega^2}{2}\widehat{\lambda}'\sup_{h\in\mathcal{H}_\delta}\sum_{t=1}^{T}\rho_2\left(\omega\check{\lambda}'g_{ts}(\theta_0,h)\right)g_{ts}(\theta_0,h)\,g_{ts}(\theta_0,h)'\,\widehat{\lambda}\le\\&\|\omega\widehat{\lambda}\|\,\|\widehat{g}_s(\theta_0,\widehat{h})\|-s_T\|\omega\widehat{\lambda}\|^2C,\end{aligned}$$

where the last inequality follows by the triangle inequality, a similar argument as that used in (A-7) and $T^{-1}\sum_{t=1}^{T}\rho_2(\omega\check{\lambda}' g_{ts}(\theta_0,h))g_{ts}(\theta_0,h_0)\,g_{ts}(\theta_0,h_0)\leq -CI$ (Smith, 2011, p. 1224). By condition 5(b) and the CLT $\|\,\widehat{g}_s(\theta_0,\widehat{h})\,\| = O\left(T^{-1/2}\right)$ hence $\|\,\widehat{\lambda}\,\| = O_p\left(s_T T^{-1/2}\right)$. Thus $\Pr(\widehat{\lambda}\in\Lambda_T^r)\to 1$, which in turn implies the consistency of $\widehat{\theta}$ given in (A-9) with $\widehat{\lambda}=\overline{\lambda}$. The asymptotic normality follows using the same Taylor expansion and the same arguments as those used to obtain (A-8) (see also Schennach, 2007). ■

Proof of Theorem 3.3: The consistency of $\widehat{\theta}$ and $\widehat{\lambda}$ follows by the same arguments as those used in the proof of Theorem 3.2, so we assume consistency and derive the asymptotic distribution of $T^{1/2}(\widehat{\theta}-\theta_0)$. By a Taylor expansion with Cauchy remainder

$$g_t(\theta_0,\widehat{h}) = g_t(\theta_0,h_0) + \partial_h g_t(\theta_0,h_0)(\widehat{h}-h_0) + \frac{1}{2}\int_0^1 \partial^2_{hh} g_t^h(\theta_0,h_0+\xi(\widehat{h}-h_0))d\xi,$$

where $\partial^2_{hh} g_t^h(\cdot) = \sum_{j=1}^{l_h}(\widehat{h}-h_0)_j\partial^2_{hh_j} g_t(\cdot)(\widehat{h}-h_0)'$, so that by Assumption 7 it follows that

$$\begin{aligned}
&T^{1/2}\widehat{g}_s(\theta_0,\widehat{h}) = T^{1/2}\widehat{g}(\theta_0,h_0) + \frac{1}{T^{3/2}}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)\partial_h g_{t-s}(\theta_0,h_0)\times\\
&\sum_{\tau=1,\tau\neq t}^{T}\Phi_T(z_{2t},z_{2t-\tau})\odot\phi(z_t) + \frac{1}{T^{3/2}}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)r_T(z_{2t-s}) +\\
&\frac{1}{T^{3/2}}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)\int_0^1\partial^2_{hh}g_t^h(\theta_0,h_0+\xi(\widehat{h}-h_0))d\xi\\
&:= \mathfrak{T}_4^* + \mathfrak{T}_5^* + \mathfrak{T}_6^* + \mathfrak{T}_7^*.
\end{aligned}$$

By CLT $\mathfrak{T}_4^* \xrightarrow{d} N\left(0,\omega_1^2\Omega(\theta_0,h_0)\right)$ whereas by Assumption 7(a) and Lemma C.1 of Smith (2011)

$$\lim_{T\to\infty}\frac{1}{s_T}\sum_{s=1-T}^{T-1}\left|\omega\left(\frac{s}{s_T}\right)\right|\frac{\|r_T(z_{2t-s})\|_{\mathcal{H}}}{T^{1/2}} = o_p(1).$$

The term $\mathfrak{T}_5^*$ can be written as

$$\begin{aligned}
\mathfrak{T}_5^* &= \frac{1}{s_T}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)\frac{1}{T^{3/2}}\sum_{t=\max(1,1-s)}^{\min(T,T-s)}\sum_{\tau=1,\tau\neq t}^{T}\partial_h g_{t-s}(\theta_0,h_0)\,\Phi_T(z_{2t},z_{2t-\tau})\odot\phi(z_t)\\
&:= \frac{1}{s_T}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)U_{T,s},
\end{aligned} \tag{A-10}$$

and note that the difference between $U_{T,s}$ and $U_T := \sum_{t=1}^{T}U_T/T^{3/2}$ consists of s terms. The Markov inequality yields

$$P\left(\frac{1}{T^{3/2}}\left|\sum_{t=1}^{s}\sum_{\tau=1,\tau\neq t}^{T}\partial_h g_{t-s}\left(\theta_0,h_0\right)\Phi_T\left(z_{2t},z_{2t-\tau}\right)\odot\phi\left(z_t\right)\right|\geq\epsilon\right)\leq$$
$$\frac{1}{\epsilon T^{3/2}}\sum_{t=1}^{s}\sum_{\tau=1,\tau\neq t}^{T}E\left|\partial_h g_{t-s}\left(\theta_0,h_0\right)\Phi_T\left(z_{2t},z_{2t-\tau}\right)\odot\phi\left(z_t\right)\right|\leq$$
$$\frac{1}{\epsilon T^{3/2}}\sum_{t=1}^{s}\left\|\partial_h g_t\left(\theta_0,h_0\right)\right\|_2\sup_{z_{2t}}\left\|\sum_{\tau=1}^{T}\Phi_T\left(z_{2t},z_{2t-\tau}\right)\odot\phi\left(z_t\right)\right\|_2\leq O\left(\frac{|s|}{T^{1-\delta}}\right),$$

where the last equality follows from Assumptions 7(b)-(c). It then follows that

$$P\left(\left|\frac{1}{s_T}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)\{U_{T,s}-U_T\}\right|>\epsilon\right)\leq\frac{1}{\epsilon}\frac{1}{s_T}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)E\left|U_{T,s}-U_T\right| \tag{A-11}$$
$$\leq CT^{\delta}\frac{1}{s_T}\sum_{s=1-T}^{T-1}\frac{|s|}{T}\left|\omega\left(\frac{s}{s_T}\right)\right|=O(T^{\delta-\eta-1/2})=o(1),$$

where the last equality follows from Lemma C.1 of Smith (2011).[6] Thus by (A-10) and (A-11)

$$\mathfrak{T}_5^*=(\omega_1+o(1))\sqrt{T}U_T^*+O(T^{\delta-\eta-1/2}),$$

where $U_T^*:=\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\widetilde{\Phi}_T\left(z_{2s},z_{2t}\right)/T\left(T-1\right)$ and $\widetilde{\Phi}_T\left(z_{2s},z_{2t}\right)=\partial_h g_t\left(\theta_0,h_0\right)\Phi_T\left(z_{2s},z_{2t}\right)\odot\phi\left(z_s\right)$. Since U_T^* can be represented as a U-statistic with a varying symmetric kernel, that is

$$U_T^*=\frac{1}{T(T-1)}\sum_{1\leq s<t\leq T}\Psi_T\left(z_{2s},z_{2t}\right),$$

where $\Psi_T\left(z_{2s},z_{2t}\right):=\widetilde{\Phi}_T\left(z_{2s},z_{2t}\right)+\widetilde{\Phi}_T\left(z_{2t},z_{2s}\right)$, the asymptotic normality of $T^{1/2}U_T^*$ follows by either Lemma B.1 or B.2, so that the asymptotic normality of $T^{1/2}\widehat{g}_s(\theta_0,\widehat{h})$ follows by the CMT as long as $\|\mathfrak{T}_7^*\|=o_p(1)$. Note that Theorem 3.1 and A7(c) yield $\sup_{h\in\mathcal{H}_\delta}\sum_{t=1}^{T}\partial_{hh}^2 g_t\left(\theta_0,h\right)/T\overset{a.s.}{\to}E[\sup_{h\in\mathcal{H}_\delta}\partial_{hh}^2 g_t\left(\theta_0,h\right)]$, which implies that

$$\frac{1}{T}\sum_{t=1}^{T}\left|\int_0^1(1-\xi)\partial_{hh}^2 g_t\left(\theta_0,\xi\left(h-h_0\right)\right)d\xi\right|=O_{a.s.}(1).$$

Thus $\|\mathfrak{T}_7^*\|=o_p\left(1\right)$ follows by Assumption 5(a) and the CMT. The asymptotic equivalence between the GEL estimator $\widehat{\theta}$ defined in (2.5), and the ETEL estimator $\widehat{\theta}$ defined in (2.8) implies that the latter has the same asymptotic covariance as that of the former. Finally for the GMM estimator $\widehat{\theta}$ defined in (2.7), the result follows by the above arguments and those used in the proof of Theorem 3.2 using the metric $\widehat{\Omega}^e(\widetilde{\theta},\widehat{h})^{-1}$, Assumption (e) and the CMT. ■

[6]In fact Lemma C.1 in Smith (2011) states that $\lim_{T\to\infty}s_T^{-1}\sum_{t=1-T}^{T-1}|t|T^{-1}\left|\omega\left(t/s_T\right)\right|=0$. However, an examination of the proof reveals that, actually, $s_T^{-1}\sum_{t=1-T}^{T-1}|t|T^{-1}\left|\omega\left(t/s_T\right)\right|=O\left((T/s_T)^{-1}\right)$.

Proof of Proposition 4.1: Note that by the consistency of $\widehat{\alpha}$

$$T^{1/2}\widehat{g}_s(\theta_0,\widehat{h}(\widehat{x}_{2t}))\tau(\widehat{x}_{2t}) = T^{1/2}\widehat{g}_s(\theta_0,\widehat{h}(\widehat{x}_{2t}))\left(\tau(\widehat{x}_{2t})-\tau(x_{2t})\right)+T^{1/2}\widehat{g}_s(\theta_0,\widehat{h}(\widehat{x}_{2t}))\tau(x_{2t})$$
$$= T^{1/2}\widehat{g}_s(\theta_0,\widehat{h}(\widehat{x}_{2t}))\tau(x_{2t})+o_p(1)$$

and

$$T^{1/2}\widehat{g}_s(\theta_0,\widehat{h}(\widehat{x}_{2t}))\tau(x_{2t}) = T^{1/2}\omega_1\widehat{g}(\theta_0,h(x_{2t},\theta_0))\tau(x_{2t}) - \frac{\omega_1}{T^{1/2}}\sum_{t=1}^{T} w_t[\widehat{h}(x_{2t},\theta_0)-h(x_{2t},\theta_0)]\tau(x_{2t}) -$$
$$\frac{\omega_1}{T^{1/2}}\sum_{t=1}^{T} w_t[\widehat{h}(\widehat{x}_{2t},\theta_0)-\widehat{h}(x_{2t},\theta_0)]\tau(x_{2t})$$
$$:= \mathcal{T}_{g;1}+\mathcal{T}_{g;2}+\mathcal{T}_{g;3},$$

where $\widehat{h}(\cdot)$ is a kernel estimator for $h(\cdot)$, $\widehat{x}_{2t}=s_t-v_t'\widehat{\alpha}$ is the regression residual and, for notational simplicity, for $x_{2t}(\alpha_0)=s_t-v_t'\alpha_0$, $x_{2t}(\alpha_0):=x_{2t}$ The asymptotic normality of $\mathcal{T}_{g;1}$ follows by CLT; furthermore $\mathcal{T}_{g;2}=o_p(1)$ by an application of the Cauchy-Schwartz inequality, a covariance inequality for α mixing processes (see e.g., Truong and Stone, 1992), a standard law of large numbers and results of Liebscher (1998), which show that

$$\operatorname{var}(\mathcal{T}_{g;2}) \le \sup_t E\left[|\widehat{h}(x_{2t},\theta_0)-h(x_{2t},\theta_0)|^2\tau(x_{2t})\right]\frac{\omega_1}{T}\sum_{t=1}^{T} w_t w_t'$$
$$+\sup_t E\left[|\widehat{h}(x_{2t},\theta_0)-h(x_{2t},\theta_0)|^2\tau(x_{2t})\right]^2\frac{\omega_1}{T}\sum_{s\neq t}^{T}\left|E[w_s w_t']\right|$$
$$= O\left(\sup_t E\left[|\widehat{h}(x_{2t},\theta_0)-h(x_{2t},\theta_0)|^2\tau(x_{2t})\right]\right)\to 0$$

For $\mathcal{T}_{g;3}$ since x_{2t} is estimated parametrically we have the following linear representation:

$$\widehat{h}(\widehat{x}_{2t},\theta_0)-\widehat{h}(x_{2t},\theta_0) = \frac{1}{D_T(x_{2t},\theta_0)}\left(1-D_T(\widehat{x}_{2t},\theta_0)[D_T(\widehat{x}_{2t},\theta_0)-D_T(x_{2t},\theta_0)]\right)$$
$$\left[N_T(\widehat{x}_{2t},\theta_0)-N_T(x_{2t},\theta_0)-\frac{N_T(x_{2t},\theta_0)}{D_T(x_{2t},\theta_0)}[D_T(\widehat{x}_{2t},\theta_0)-D_T(x_{2t},\theta_0)]\right],$$

where $D_T(x_{2t},\theta):=T^{-1}\sum_{s\neq t}^{T}K_{b_T}(x_{2s}-x_{2t})$, $N_T(x_{2t},\theta):=T^{-1}\sum_{s\neq t}^{T}(y_s-x_{1t}'\theta)K_{b_T}(x_{2s}-x_{2t})$, $K_{b_T}(x_{2s}-x_{2t}):=K\left((x_{2s}-x_{2t})/b_T\right)/b_T$, $K(\cdot)$ is a kernel function and b_T is the bandwidth. Then again by the results of Liebscher (1998) we have

$$\widehat{h}(\widehat{x}_{2t},\theta_0)-\widehat{h}(x_{2t},\theta_0) = \frac{1}{D_T(x_{2t},\theta_0)}(1+o_p(1))$$
$$\times\left[N_T(\widehat{x}_{2t},\theta_0)-N_T(x_{2t},\theta_0)-\frac{N_T(x_{2t},\theta_0)}{D_T(x_{2t},\theta_0)}\left(D_T(\widehat{x}_{2t},\theta_0)-D_T(x_{2t},\theta_0)\right)\right]$$

uniformly in $x_{2t}\in\mathcal{X}_2^*$. Since $\widehat{x}_{2s}-\widehat{x}_{2t}=x_{2s}-x_{2t}+(v_t-v_s)'(\widehat{\alpha}-\alpha_0)$, a Taylor expansion and the same argument as in Li and Wooldridge (2002) yield

$$N_T(\widehat{x}_{2t},\theta_0) - N_T(x_{2t},\theta_0) = \frac{1}{(T-1)}\sum_{s\neq t}^{T}(y_s - x_{1s}'\theta_0)K_{b_T}^{(1)}(x_{2s}-x_{2t})\left(\frac{v_t - v_s}{b_T}\right)'(\widehat{\alpha}-\alpha_0) + o_p(1),$$

$$D_T(\widehat{x}_{2t},\theta_0) - D_T(x_{2t},\theta_0) = \frac{1}{(T-1)}\sum_{s\neq t}^{T}K_{b_T}^{(1)}(x_{2s}-x_{2t})\left(\frac{v_t - v_s}{b_T}\right)'(\widehat{\alpha}-\alpha_0) + o_p(1).$$

Thus $\mathcal{T}_{g;3}$ can be represented as

$$\begin{aligned}\mathcal{T}_{g;3} &= \frac{\omega_1}{T^{1/2}}\sum_{t=1}^{T}\tau(x_{2t})\frac{w_t(1+o_p(1))}{D_T(x_{2t},\theta_0)}\frac{1}{(T-1)}\sum_{s\neq t}^{T}(y_s - x_{1s}'\theta_0)K_{b_T}^{(1)}(x_{2s}-x_{2t})\left(\frac{v_t-v_s}{b_T}\right)'(\widehat{\alpha}-\alpha_0)\\ &- \frac{\omega_1}{T^{1/2}}\sum_{t=1}^{T}\tau(x_{2t})\frac{w_t(1+o_p(1))}{D_T(x_{2t},\theta_0)}\widehat{h}(x_{2t},\theta_0)\frac{1}{(T-1)}\sum_{s\neq t}^{T}K_{b_T}^{(1)}(x_{2s}-x_{2t})\left(\frac{x_{3t}-x_{3s}}{b_T}\right)'(\widehat{\alpha}-\alpha_0)\\ &= \mathcal{T}_{g;3;a} - \mathcal{T}_{g;3;b}.\end{aligned}$$

An application of the triangle inequality yields

$$\begin{aligned}&\left|\mathcal{T}_{g;3;a} - \left\{\frac{\omega_1}{T}\sum_{t=1}^{T}\tau(x_{2t})\frac{w_t(1+o_p(1))}{D_T(x_{2t},\theta_0)}\partial_\alpha[f(x_{2t})h_0(x_{2t}),\theta_0)]\right\}T^{1/2}(\widehat{\alpha}-\alpha_0)\right|\\ &\quad\leq \frac{\omega_1}{T}\sum_{t=1}^{T}\left|\frac{w_t(1+o_p(1))}{D_T(x_{2t},\theta_0)}\tau(x_{2t})\right|\max_t\left|\frac{1}{(T-1)}\sum_{s\neq t}^{T}(y_s-x_{1s}'\theta_0)K_{b_T}^{(1)}(x_{2s}-x_{2t})\left(\frac{x_{3t}-x_{3s}}{b_T}\right)'\right.\\ &\qquad\qquad\left. - f(x_{2t})\partial_\alpha[f(x_{2t})h_0(x_{2t},\theta_0)]\right|T^{1/2}(\widehat{\alpha}-\alpha_0).\end{aligned}$$

The uniform convergence results of Andrews (1995) and the extended Continuous Mapping Theorem (CMT) (see, e.g., van der Vaart and Wellner, 1996) imply that

$$\begin{aligned}&\max_t\left|\frac{1}{(T-1)}\sum_{s\neq t}^{T}(y_s-x_{1s}'\theta_0)K_{b_T}^{(1)}(x_{2s}-x_{2t})\left(\frac{x_{3t}-x_{3s}}{b_T}\right)' - f(x_{2t})\partial_\alpha h(x_{2t}),\theta_0)\right|\\ &\quad= \max_t\left|\partial_\alpha\left\{\frac{1}{(T-1)}\sum_{s\neq t}^{T}(y_s-x_{1s}'\theta_0)K_{b_T}((x_{2s})-x_{2t}))\right\} - \partial_\alpha[f(x_{2t})h_0(x_{2t},\theta_0)]\right|\\ &\quad= O\left(\max_t\left|\left\{\frac{1}{(T-1)}\sum_{s\neq t}^{T}(y_s-x_{1s}'\theta_0)K_{b_T}(x_{2s}-x_{2t}))\right\} - [f(x_{2t})h_0(x_{2t},\theta_0)]\right|\right) = o_p(1),\end{aligned}$$

and therefore

$$\mathcal{T}_{g;3;a} = \left\{\frac{\omega_1}{T}\sum_{t=1}^{T}\frac{w_t(1+o_p(1))}{f(x_{2t})+o_p(1)}\tau(x_{2t})\partial_\alpha[f(x_{2t})h_0(x_{2t},\theta_0)]\right\}T^{1/2}(\widehat{\alpha}-\alpha_0) + o_p(1). \tag{A-12}$$

Similarly, we can verify that

$$\mathcal{T}_{g;3;b} = \left\{ \frac{\omega_1}{T} \sum_{t=1}^{T} \frac{w_t[h_0(x_{2t},\theta_0) + o_p(1)]}{f(x_{2t}) + o_p(1)} \tau(x_{2t}) \partial_\alpha f(x_{2t})) \right\} T^{1/2}(\widehat{\alpha} - \alpha_0) + o_p(1). \tag{A-13}$$

Equations (A-12) and (A-13) then imply that

$$\begin{aligned}\mathcal{T}_{g;3} = \Bigg\{ \frac{\omega_1}{T} \sum_{t=1}^{T} \Bigg(& \frac{w_t(1 + o_p(1))}{f(x_{2t}) + o_p(1)} \tau(x_{2t}) \partial_\alpha [f(x_{2t}) h_0(x_{2t}), \theta_0)] - \\ & - \frac{w_t[h_0(x_{2t},\theta_0) + o_p(1)]}{f(x_{2t}) + o_p(1)} \tau(x_{2t}) \partial_\alpha f(x_{2t})) \Bigg) \Bigg\} T^{1/2}(\widehat{\alpha} - \alpha_0) + o_p(1),\end{aligned}$$

which is the Bahadur representation of the *degenerate* U-statistic $\mathfrak{T}_5^*$ in the Taylor expansion involving $\widehat{g}_{ts}(\theta_0, h)$ in the proof of Theorem 3.3. Note that to apply Theorem B.1 in Appendix B in the supplement, one can define

$$\begin{aligned}\Psi_T(z_s, z_t) := {} & \tau(x_{2t}) \frac{w_t}{f(x_{2t})} \{\partial_\alpha \{f(x_{2t}) h_0\,(x_{2t}), \theta_0)\} - h_0\,(x_{2t}, \theta_0)\, \partial_\alpha f(x_{2t})\} r(v_s) x_{2s} + \\ & \tau(x_{2s}) \frac{w_s}{f(x_{2s})} \{\partial_\alpha \{f(x_{2s}) h_0\,(x_{2s}, \theta_0)\} - h_0\,(x_{2s}, \theta_0)\, \partial_\alpha f(x_{2s})\} r(v_t) x_{2t} + o_p\left(T^{-1/2}\right).\end{aligned}$$

By Hölder's and Minkowski's inequalities, one can readily verify that Conditions (B-1)-(B-4) are satisfied. Therefore, we obtain that $\widehat{g}_{ts}(\theta_0, \widehat{h}(\widehat{x}_{2t}))\tau(x_{2t})/\omega_1 \xrightarrow{d} N(0, \Omega_d^e)$ and the conclusion follows by the same arguments as those used in the proof of Theorem 3.3. ∎

Table 1: Monte Carlo Results - Endogenous Variable - $\rho_{eu} = 0.1$

		$c = 0.5$				$c = 1$				$c = 1.5$			
Estimator	T	Bias	Std. Dev.	Ratio	Cov. Prob.	Bias	Std. Dev.	Ratio	Cov. Prob.	Bias	Std. Dev.	Ratio	Cov. Prob.
GMM	200	-0.0199	0.2518	1	0.955	-0.0191	0.2491	1	0.950	-0.0286	0.2455	1	0.950
	400	0.0159	0.1740	1	0.949	0.0165	0.1734	1	0.949	0.0188	0.1729	1	0.949
	800	0.0138	0.1240	1	0.944	0.0135	0.1237	1	0.949	0.0130	0.1236	1	0.948
s-GMM	200	-0.0187	0.2490	0.9961	0.953	-0.0078	0.2469	0.9965	0.951	-0.0174	0.2424	0.9976	0.949
	400	0.0151	0.1738	1.0020	0.950	0.0058	0.1731	1.0019	0.949	0.0155	0.1726	1.0023	0.947
	800	0.0128	0.1238	1.0006	0.943	0.0032	0.1236	1.0005	0.944	0.0128	0.1235	1.0005	0.945
EL	200	0.0069	0.2520	1.0034	0.947	0.0059	0.2602	1.0025	0.946	0.0053	0.2554	1.0036	0.949
	400	0.0060	0.1785	1.0030	0.943	0.0050	0.1795	1.0032	0.945	0.0051	0.1789	1.0037	0.946
	800	0.0035	0.1258	1.0012	0.941	0.0031	0.1254	1.0009	0.944	0.0035	0.1250	1.0007	0.941
ET	200	-0.0094	0.2495	1.0038	0.954	-0.0081	0.2503	1.0029	0.954	-0.0084	0.2731	1.0032	0.951
	400	0.0058	0.1846	1.0014	0.949	0.0053	0.1835	1.0017	0.949	0.0058	0.1835	1.0024	0.948
	800	-0.0037	0.1267	1.0013	0.942	-0.0036	0.1259	1.0003	0.946	0.0035	0.1263	1.0010	0.945
CUE	200	-0.0161	0.2406	1.0019	0.959	-0.0151	0.2469	0.9924	0.952	-0.0111	0.2732	1.0201	0.948
	400	0.0089	0.1887	1.0015	0.955	0.0074	0.1903	1.0037	0.944	0.0104	0.1758	1.0190	0.951
	800	-0.0064	0.1261	1.0002	0.939	-0.0042	0.1256	1.0001	0.942	0.0042	0.1256	1.0001	0.943
ETEL	200	0.0201	0.2591	1.0145	0.950	0.0212	0.2492	1.0155	0.953	0.0296	0.2459	1.0147	0.954
	400	0.0112	0.1749	1.0155	0.948	0.0143	0.1773	1.0150	0.946	0.0188	0.1754	1.0143	0.941
	800	0.0128	0.1266	1.0153	0.950	0.0133	0.1254	1.0145	0.944	0.0136	0.1589	1.0132	0.948

Note: Table displays Monte Carlo Bias (bias), Standard Deviation (Std. Dev.), Average Standard Errors Ratios (Ratio) and Coverage Probability (Cov. Prob.) for the Generalized Method of Moments (GMM), its smoothed version (s-GMM), Empirical Likelihood (EL), Exponential Tilting (ET), Continuous Updated (CU) and Exponentially Tilted Empirical Likelihood (ETEL) estimator of $\theta_{0;1} = 1$.

Table 2: Monte Carlo Results - Endogenous Variable - $\rho_{eu} = 0.9$

		$c = 0.5$				$c = 1$				$c = 1.5$			
Estimator	T	Bias	Std. Dev.	Ratio	Cov. Prob.	Bias	Std. Dev.	Ratio	Cov. Prob.	Bias	Std. Dev.	Ratio	Cov. Prob.
GMM	200	0.0267	0.2812	1	0.967	0.0234	0.2714	1	0.967	0.0262	0.2613	1	0.965
	400	0.0159	0.1757	1	0.948	0.0185	0.1748	1	0.949	0.0201	0.1738	1	0.949
	800	0.0088	0.1267	1	0.952	0.0086	0.1264	1	0.951	0.0185	0.1260	1	0.950
s-GMM	200	0.0204	0.2775	0.9979	0.965	0.0228	0.2667	1.0012	0.966	0.0215	0.2570	1.0007	0.963
	400	0.0181	0.1761	1.0035	0.946	0.0156	0.1749	1.0033	0.947	0.0185	0.1739	1.0037	0.949
	800	0.0145	0.1268	1.0017	0.952	0.0131	0.1265	1.0016	0.952	0.0132	0.1261	1.0016	0.950
EL	200	-0.0202	0.2921	1.0117	0.965	-0.0213	0.2825	1.0113	0.960	-0.0211	0.2749	1.0088	0.957
	400	-0.0150	0.1825	1.0076	0.951	-0.0143	0.1780	1.0072	0.950	-0.0142	0.1804	1.0072	0.950
	800	-0.0101	0.1296	1.0051	0.948	-0.0113	0.1316	1.0057	0.953	-0.0101	0.1280	1.0051	0.951
ET	200	-0.0231	0.3042	1.0492	0.979	-0.0252	0.2994	1.0343	0.977	-0.0235	0.2749	1.0498	0.979
	400	-0.0150	0.1909	1.0102	0.948	-0.0195	0.1897	1.0099	0.946	-0.0186	0.1894	1.0094	0.945
	800	-0.0107	0.1320	1.0054	0.956	-0.0108	0.1317	1.0054	0.956	-0.0108	0.1312	1.0054	0.955
CUE	200	-0.0241	0.3115	1.0287	0.966	-0.0221	0.5250	1.0325	0.988	-0.0206	0.5634	1.0301	0.991
	400	-0.0111	0.1994	1.0107	0.956	-0.0198	0.1908	1.0101	0.952	-0.0188	0.1895	1.0095	0.950
	800	-0.0094	0.1314	1.0052	0.955	-0.0105	0.1310	1.0053	0.956	-0.0085	0.1307	1.0053	0.955
ETEL	200	0.0276	0.3515	1.0186	0.960	0.0240	0.3307	1.0168	0.950	0.0231	0.3328	1.0213	0.951
	400	0.0152	0.2047	1.0150	0.947	0.0199	0.2047	1.0148	0.950	0.0171	0.2033	1.0154	0.945
	800	0.0104	0.1868	1.0146	0.949	0.0137	0.1852	1.0147	0.945	0.0143	0.1869	1.0140	0.943

Note: Table displays Monte Carlo Bias (bias), Standard Deviation (Std. Dev.), Average Standard Errors Ratios (Ratio) and Coverage Probability (Cov. Prob.) for the Generalized Method of Moments (GMM), its smoothed version (s-GMM), Empirical Likelihood (EL), Exponential Tilting (ET), Continuous Updated (CU) and Exponentially Tilted Empirical Likelihood (ETEL) estimator of $\theta_{0;1} = 1$.

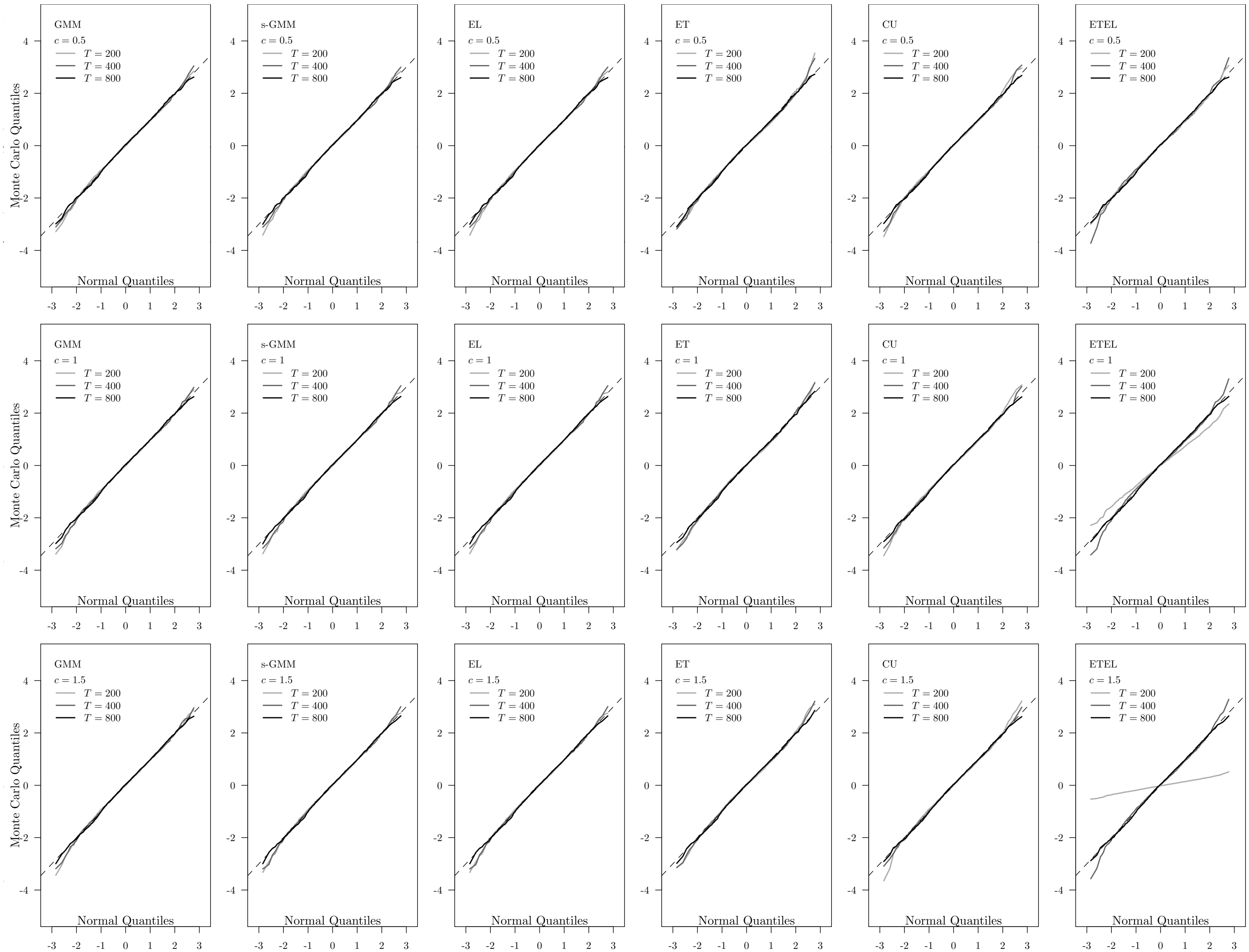

Figure 1: Q-Q Plots of standardized Monte Carlo sample versus the theoretical quantiles of a standard normal distribution, with a 45-degrees (dashed) line. Results are based on 2000 Monte Carlo replications of the Generalized Method of Moments (GMM), its smoothed version (s-GMM), Empirical Likelihood (EL), Exponential Tilting (ET), Continuous Updated (CU) and Exponentially Tilted Empirical Likelihood (ETEL) estimator of $\theta_{0;2} = 1$ when $\rho_{eu} = 0.1$.

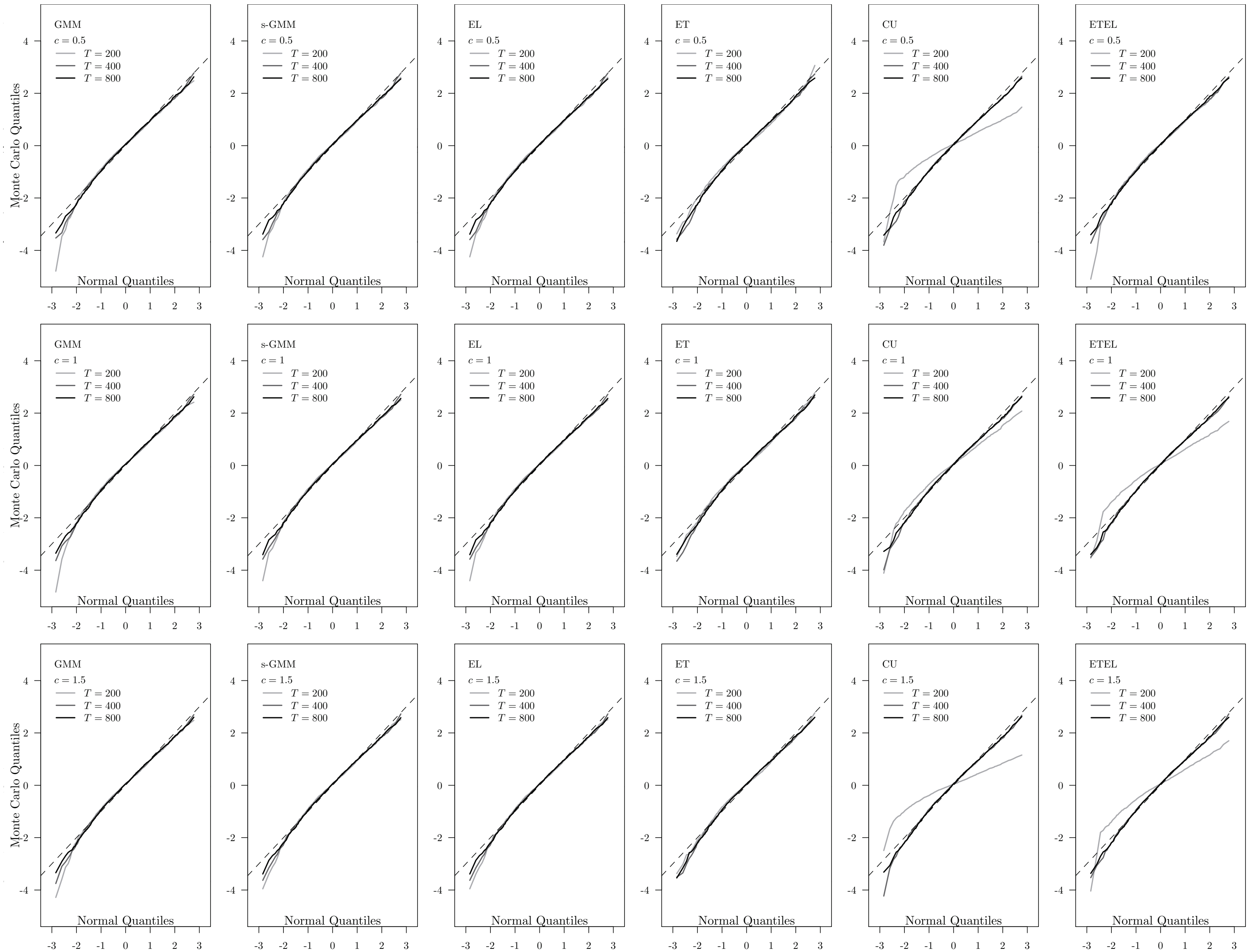

Figure 2: Q-Q Plots of standardized Monte Carlo sample versus the theoretical quantiles of a standard normal distribution, with a 45-degrees (dahsed) line. Results are based on 2000 Monte Carlo replications of the Generalized Method of Moments (GMM), its smoothed version (s-GMM), Empirical Likelihood (EL), Exponential Tilting (ET), Continuous Updated (CU) and Exponentially Tilted Empirical Likelihood (ETEL) estimator of $\theta_{0;2} = 1$ when $\rho_{eu} = 0.9$.