Article:

Bravo, Francesco orcid.org/0000-0002-8034-334X (2024) Estimation of non smooth nonparametric estimating equations models with dependent data. Journal of Time Series Analysis. ISSN 1467-9892

https://doi.org/10.1111/jtsa.12758

Estimation of non smooth nonparametric estimating equations models with dependent data

Francesco Bravo*
University of York

May 2024

Abstract

This paper considers estimation of non smooth possibly overidentified nonparametric estimating equations models with weakly dependent data. The estimators are based on a kernel smoothed version of the generalized empirical likelihood and the generalized method of moments approaches. The paper derives the asymptotic normality of both estimators and shows that the proposed local generalized empirical likelihood estimator is more efficient than the local generalized moment estimator unless a two-step procedure is used. The paper also proposes novel tests for the correct specification of the considered model that are shown to have power against local alternatives and are consistent against fixed alternatives. Monte Carlo simulations and an empirical application illustrate the finite sample properties and applicability of the proposed estimators and test statistics.

Keywords: β-mixing, Kernel estimation, Quantile regression
MSC classification: 62G08, 62G20

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

1 Introduction

Estimating equations (EE) models (often called moment conditions models in the econometric literature) arise naturally in economics, finance and statistics: For example, many dynamic stochastic general equilibrium models used in macroeconomics, many treatment effects models used in microeconomics, many assets pricing models under the no arbitrage condition used in finance, and generalized estimating equations (GEE) for longitudinal data used in statistics, all give rise to a set of (possibly overidentified) estimating equations. Estimation of the unknown parameters in such models is typically carried out using Hansen's (1982) generalized method of moments (GMM) - see also Qu, Lindsay and Li (2000) for GEE models, or, alternatively, Newey and Smith's (2004) generalized empirical likelihood (GEL). When the unknown parameters are finite dimensional, the asymptotic properties of GEL and GMM estimators are well understood. Extensions to EE models with infinite dimensional parameters, which we call nonparametric EE (NPEE) models have been proposed in the literature: Severini and Staniswalis (1994) considered a nonparametric quasi-likelihood model, Cai (2003), Lewbel (2007) and Bravo (2022) considered different specifications of NPEE models, Cai and Li (2008) (see also Bravo (2016)) considered a nonparametric dynamic panel data model, Fang, Ren and Yuan (2011) of Cai, Ren and Sun (2015) considered a nonparametric stochastic discount factor model, among others.

All of the above papers are based on *smooth* (that is differentiable) EE; in this paper, we consider *non smooth* EE. Non smooth statistical models are theoretically interesting and empirically relevant, as they include least absolute deviations and more generally quantile regression models (Koenker and Bassett 1978), rank regression models (Cuzik 1988), copula models (Patton 2012) and receiver operating characteristic (ROC) curves models (Pepe 1997), all of which give rise to a set of non smooth EE.

This paper contributes to the literature on estimating non smooth EE models by considering GEL and GMM estimation of non smooth NPEE models. The estimators are based on local smoothing, hence we call them local GEL (LGEL) and local GMM (LGMM), respectively. It should be noted that local smoothing has been used before in the case of non smooth EE models both with unknown finite and infinite dimensional parameters. For the former, Chen and Hall (1993), Horowitz (1998), Whang (2006) and Otsu (2008) applied smoothing directly to the EE themselves, so that the asymptotic normality of the related unknown finite dimensional parameters estimators can be derived by direct (standard) methods. For the latter, Fan, Hu and Truong (1994), Yu and Jones (1998) and Cai and Xu (2008) (among others) used smoothing to directly estimate the unknown infinite dimensional parameters, however the asymptotic normality of the resulting estimators relies heavily on the structure of the underlying model. The main contribution of this paper is to derive the asymptotic normality of the proposed LGEL and LGMM estimators that does not rely on any particular structure of the underlying NPEE model. To obtain this result, we assume smoothness (local differentiability) of the expectation of the

NPEE model and use a stochastic equicontinuity argument. Stochastic equicontinuity[1] is the natural technical tool to use in the context of non smooth EE models, see for example Pakes and Pollard (1989) for *parametric* EE models under random sampling, however the assumed weakly dependent structure of the observations and the nonparametric nature of the model considered here create some additional technical challenges, which are addressed first under a set of high level conditions (see Section 3), and then under a set of more primitive conditions (see Section 4). As far as we are aware of, this is the first paper that addresses the issue of estimating general non smooth NPEE models with weakly dependent observations.

The main contributions of the paper are as follows:

First, it considers weakly dependent observations, and specify the dependency of the observations as β (or absolutely regular) mixing (Volkonskii and Rozanov 1959). Many commonly used time series models can be shown to be β mixing: examples include various GARCH and stochastic volatility models (Carrasco and Chen 2002), and, more generally, Markov chains models under the Harris recurrence condition, see for example Davydov (1973) and Mokkadem (1990). The assumption of β mixing (rather than the weaker α mixing) dependency seems the natural one in the context of this paper, because of the stochastc equicontinuity argument used in the proofs of Theorems 1 and 2, which would require more stringent assumptions on the summability of the mixing coefficient and the complexity of the underlying (functional) parameter space under α mixing, see for example Andrews and Pollard (1994), than those assumed in this paper - see the discussion of Assumptions A1-A8 in Section 3 for more details about this important point.

Second, we consider overidentified NPEE models, that is models where the dimension of the EE can be larger than the dimension of the unknown infinite dimensional parameters, and allow for possible endogeneity, that is possible correlation between some or all of the variables which, if not accounted for, would result in a misspecified NPEE model.

Third, we obtain the asymptotic distributions of the LGEL and LGMM estimators and show that the former is always characterized by a smaller asymptotic covariance matrix than the latter, unless a two step procedure is used, see Remark 2 in Section 3 for more details.

Fourth, we propose new LGMM and LGEL based overidentification test statistics that can be used to test for the correct specificification of nonsmooth NPEE models, and are in the same spirit of those proposed by Hansen (1982) and Newey and Smith (2004) for finite dimensional overidentified smooth EE models. We show that under the null hypothesis the proposed statistics are characterized by a nonstandard asymptotic distribution which is asymptotically pivotal, hence easy to simulate; we also show that the proposed statistics have power against local Pitman alternative hypotheses and are consistent against any fixed alternatives.

Fifth, we provide simulations evidence about the finite sample properties of two examples of

[1]See the Appendix for a formal definition of stochastic equicontinuity.

the LGEL estimator (namely the local empirical likelihood (LEL) defined in (5) and the local exponential tilting (LET) defined in (6) estimators) and the two step LGMM estimator, which seems to suggest that, as for the case of finite dimensional overidentified smooth EE models, both LGEL estimators and their related test statistics are characterized by better (as measured by the mean squared error criterion, and size and power, respectively) finite sample properties compared to the two step LGMM estimator and related test statistic. In addition, the simulations seem to confirm that in the important case of NPEE models with instrumental variables - see Section 4 for an example - the mean squared error of the LEL estimator does not seem to increase with the number of instruments, a fact shown analytically by Newey and Smith (2004) and Bravo (2022) for parametric and nonparametric overidentified smooth EE models, respectively.
Finally, we consider an empirical application, which illustrates the applicability of the proposed estimators and test statistics.

The rest of the paper is organized as follows: next section introduces the model, presents a simple illustrative example and describes the estimators. Section 3 presents the main results; Section 4 illustrates how the results of Section 3 can be verified for an instrumental variables varying coefficients quantile regression model under more primitive conditions. Sections 5 and 6 contain, respectively, the results of the Monte Carlo simulation study and the empirical application. Section 7 contains some concluding remarks. An Appendix contains the formal definitions of the most technical concepts used in the paper; a supplemental Appendix contains all the proofs, some additional Monte Carlo simulations results and the UK data used in the empirical application.

The following notation is used throughout the paper: "$^\top$" indicates transpose, for any vector v, $v^{\otimes 2} = vv^\top$ and $||.||$ denotes the Euclidean norm for both vectors and matrices (also known as the Frobenious norm for the latter).

2 The model and estimators

Let $\left\{ \left(Z_t^\top, U_t \right)^\top, t \in \mathbb{Z} \right\}$ denote a strictly stationary sequence of random vectors taking values in $\mathcal{Z} \subseteq \mathbb{R}^{d_Z}$, where d_Z is the dimension of $\mathcal{Z}, \mathcal{U} \subseteq \mathbb{R}$, and let $h \in \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times ... \times \mathcal{H}_k$ denote a k dimensional vector of unknown functions, where each $\mathcal{H}_j$ $(j = 1, ..., k)$ is a pseudo-metric space of functions. The model we consider is

$$E \left[m \left(Z_t, h \left(U_t \right) \right) | U_t \right] = 0 \ a.s. \text{ for a unique } h = h_0, \tag{1}$$

where $m : \mathcal{Z} \times \mathcal{U} \times \mathcal{H} \to \mathbb{R}^l$ is a vector of known functions with $l \geq k$.
The following example provides a simple illustration of the type of models (1) encompasses; a more general example is given in (9) and in Section 4.

Example *Consider the following varying coefficients median regression model:*

$$Y_t = X_t^\top h_0(U_t) + \varepsilon_t,$$

where X_t *is an* $\mathbb{R}^k$*- valued random vector of covariates and the unobservable error term* ε_t *satisfies the conditional median restriction* $Pr(\varepsilon_t \leq 0|X_t, U_t) = 1/2$ a.s., *which is equivalent to* $E\left(sign_{1/2}(\varepsilon_t)|X_t, U_t\right) = 0$ a.s., *where* $sign_{1/2}(\varepsilon_t) = 1/2 - I(\varepsilon_t \leq 0)$. *By iterated expectations, the conditional median restriction is equivalent to*

$$E\left(g(X_t)\, sign_{1/2}(\varepsilon_t)|U_t\right) = 0,$$

where g is an $\mathbb{R}^l-$ *valued vector of known functions, which corresponds to* (1) *with* $Z_t = \left\{\left(Y_t, X_t^\top\right)^\top, t \in \mathbb{Z}\right\}$, $m(Z_t, h(U_t)) = g(X_t)\, sign_{1/2}\left(Y_t - X_t^\top h(U_t)\right)$ *and the non smoothness in the EE* $m(.)$ *comes from the presence of the indicator function in the definition of the function* $sign_{1/2}(.)$.

Given the conditional nature of (1), we propose to estimate h_0 locally, that is for U_t in a neighborhood of u we assume that $h_0(U_t) = h_0(u) := a$, which implies that

$$E\left[m(Z_t, a)|U_t\right] \approx 0 \tag{2}$$

and base the estimators on the sample local NPEE

$$\frac{1}{Tb}\sum_{t=1}^{T} m(Z_t, a)\, K\left(\frac{U_t - u}{b}\right),$$

where $K : \mathcal{U} \rightarrow \mathbb{R}$ is a kernel function with bandwidth $b =: b(T)$.

To define the LGMM estimator, let $\widehat{W}(u)$ denote a, possibly random, positive semidefinite $\mathbb{R}^l \times \mathbb{R}^l$ -valued matrix; then the LGMM estimator is the approximate solution of the minimization problem

$$\widehat{a} \leq \inf_{a\in\mathcal{A}} \frac{1}{Tb}\sum_{t=1}^{T} m(Z_t, a)^\top K\left(\frac{U_t - u}{b}\right)\widehat{W}(u)\frac{1}{Tb}\sum_{s=1}^{T} m(Z_s, a)\, K\left(\frac{U_s - u}{b}\right) + o_p\left(\frac{1}{Tb}\right), \tag{3}$$

where $\mathcal{A} \supseteq \mathcal{H}$ is the space a is assumed to belong to (the relationship between $\mathcal{A}$ and $\mathcal{H}$ is discussed after Assumptions A1-A8 in Section 3). Note that the definition of the approximate minimizer $\widehat{a}$ is standard in the non smooth EE literature, see for example Pakes and Pollard (1989, Theorem 3.1 (i)) and Newey and McFadden (1994, Theorem 7.2).

To define the LGEL estimator, let

$$\rho\left(\lambda(u)^\top m(Z_t, a)\, K\left(\frac{U_t - u}{b}\right)\right) := \rho\left(s_{t,K}(u, \lambda, a)\right),$$

where ρ is a concave function in λ on its domain, an open set Λ_0 containing 0, and the auxiliary parameter $\lambda(u)$ can be thought of as an $\mathbb{R}^l$- valued vector of unknown Lagrange multipliers associated with the local constraint (2), that is

$$\sum_{t=1}^{T} \pi_t (u, \lambda, a)\, m(Z_t, a)\, K\left(\frac{U_t - u}{b}\right) \approx 0,$$

with

$$\pi_t (u, \lambda, a) = \frac{\partial \rho (s_{t,K} (u, \lambda, a)) / \partial s_t}{\sum_{t=1}^{T} \partial \rho (s_{t,K} (u, \lambda, a)) / \partial s_t}$$

playing the role of the "implied probabilities". The LGEL estimator is then defined as

$$\widehat{a}^{\rho} (u) \leq \inf_{a \in \mathcal{A}} \sup_{\lambda \in \Lambda_T(a)} \frac{1}{Tb} \sum_{t=1}^{T} \rho (s_{t,K} (u, \lambda, a)) + o_p \left(\frac{1}{Tb}\right), \tag{4}$$

where $\Lambda_T (a) = \{\lambda (u) : s_{t,K} (u, \lambda, a) \in \Lambda_0, t = 1, ..., T\}$. For example, the local version of empirical likelihood (LEL) is

$$\widehat{a}^{el} (u) \leq \inf_{a \in \mathcal{A}} \sup_{\lambda \in \Lambda_T(a)} \frac{1}{Tb} \sum_{t=1}^{T} \log (1 - s_{t,K} (u, \lambda, a)) + o_p \left(\frac{1}{Tb}\right), \tag{5}$$

whereas the local version of exponential tilting (LET) is

$$\widehat{a}^{et} (u) \leq \inf_{a \in \mathcal{A}} \sup_{\lambda \in \Lambda_T(a)} -\frac{1}{Tb} \sum_{t=1}^{T} \exp (s_{t,K} (u, \lambda, a)) + o_p \left(\frac{1}{Tb}\right). \tag{6}$$

3 Asymptotic results

Let $m(Z_t, h) := m_t(h)$ and define

$$\begin{aligned} G(u) &= \frac{\partial}{\partial h^{\top}} E[m_t(h_0) | U_t = u] f(u), \qquad (7)\\ \Omega(u) &= E\left[m_t(h_0)^{\otimes 2} | U_t = u\right] f(u) \int K^2(v)\, dv, \\ \Sigma(u)_W &= G(u)^{\top} W(u) G(u), \\ \Sigma(u) &= G(u)^{\top} \Omega(u)^{-1} G(u). \end{aligned}$$

Remark 1 *It is important to note that the "derivative" matrix $G(u)$ should be interpreted as a linear approximation of $E(m_t(h) | U_t = u) f(u)$ at $h_0(u)$ in the sense that for all h evaluated at u the following holds:*

$$lim_{||h(u) - h_0(u)|| \to 0} \frac{\|E(m_t(h_0) | U_t = u) f(u) - G(u)(h(u) - h_0(u))\|}{||h(u) - h_0(u)||} = 0. \tag{8}$$

Section 4 provides an example of how to calculate $G(u)$ in practice.

In what follows, let $m_{t,K}(.,.) := m_t(.,.)\,K((U_t - .)/b)$, where the first dot can be either a or h and the second dot can be either u or v in a neighborhood of $u \in \mathcal{U}$, so for example $m_{t,K}(a, v) := m_t(a(v))\,K((U_t - v)/b)$. Assume that:

A1 The sequence $\left\{\left(Z_t^{\top}, U_t\right)^{\top}, t \in \mathbb{Z}\right\}$ is strictly stationary β mixing with mixing coefficient $\beta(t) = O(t^{-c})$ for some $c > 0$.

A2 (i) The LGMM and LGEL estimators $\widehat{a}$ and $\widehat{a}^{\rho}$ exist (with probability approaching 1), (ii) there exists a unique h_0 such that $E[m_t(h_0)|U_t = u]\,f(u) = 0$, (iii) for all v in a neighborhood of $u \in \mathcal{U}$ and for each $\epsilon > 0$

$$\inf_{\substack{a \in \mathcal{A} \\ \|a(v)-a\| \geq \epsilon}} \left\| E\frac{(m_{t,K}(a, v))}{b} \right\| > 0,$$

(iv) h_0 is twice continuously differentiable.

A3 (i) $m_t(h, u)$ is continuous for each $h \in \mathcal{H}$, *a.s.*[2], (ii) for all v in a neighborhood of $u \in \mathcal{U}$, the classes of functions

$$\begin{aligned} \mathcal{M}_1^K &= \{m_{t,K}(a, v), v \in \mathcal{U}, a \in \mathcal{A}, \}, \\ \mathcal{M}_2^K &= \left\{(m_{t,K}(a, v))^{\otimes 2},\ v \in \mathcal{U}, a \in \mathcal{A}, \right\} \end{aligned}$$

are Glivenko-Cantelli[3].

A4 (i) The matrices $G(u)$ and $\Omega(u)$ are continuous in $u \in \mathcal{U}$, with $rank(G(v)) = k$ and $\Omega(v)$ positive definite for all v in a neighborhood of $u \in \mathcal{U}$, (ii) $\Sigma(u)$ is nonsingular, (iii) $E(m_t(h)|U_t = v)$ is differentiable in the sense of (8) at h_0 for all v in a neighborhood of $u \in \mathcal{U}$ with derivative $G(v)$, (iv)

$$(Tb)^{1/2}\left(\frac{1}{Tb}\sum_{t=1}^{T} m_{t,K}(h_0, u) - E\left(\frac{m_{t,K}(h_0, u)}{b}\right)\right) \xrightarrow{d} N(0, \Omega(u)).$$

A5 The empirical process $v_{Tb}^K(h) = \sum_{t=1}^{T}(m_{t,K}(h, u) - Em_{t,K}(h, u))/(Tb)$ satisfies

$$\sup_{\|h-h_0\| \leq \delta_{Tb}} \left|(Tb)^{1/2}\left(v_{Tb}^K(h) - v_{Tb}^K(h_0)\right)\right| = o_p(1)$$

for all $\delta_{Tb} \to 0$ as $Tb \to \infty$.

A6 (i) $\rho(s_t(u, \lambda, .))$ is twice continuously differentiable in s_t in a neighborhood of 0, with $\rho_j = -1$ $(j = 1, 2)$ and $\rho_j = \partial^j \gamma(s_t(u, \lambda, .))/\partial s_t^j|_{\lambda=0}$.

[2]That is, for a generic fixed (localized) $h_*(u)$, $\lim_{||h(u)-h_*(u)||\to 0} m_t(h, u) = m_t(h_*, u)$ for all Z_t such that $\Pr(Z_t \in \mathcal{Z}_1) = 1$, where the set $\mathcal{Z}_1$ can be a proper subset of the support $\mathcal{Z}$.

[3]See the Appendix for a formal definition of a Glivenko-Cantelli class of functions.

A7 (i) The kernel function $K : \mathcal{U} \to \mathbb{R}$ is symmetric and has a compact support, say $[-1, 1]$, (ii) the marginal density f of U_t is continuously differentiable and strictly positive at $U_t = u$, (iii) the joint density $f_{1,2}$ of U_1 and U_2 is Lipschitz continuous at $u \in \mathcal{U}$, (iv) $Tb^5 \to 0$.

A8 (i) For all v in a neighborhood of $u \in \mathcal{U}$ and a random $\widehat{W}$, $\left\| \widehat{W}(v) - W(v) \right\| = o_p(1)$, where $W(v)$ is a positive semidefinite matrix, (ii) $\Sigma(u)_W$ is nonsingular.

Assumption A1 excludes deterministic and stochastic trends and specifies the dependence structure of the sequence $\{ (Z_t^\top, U_t)^\top, t \in \mathbb{Z} \}$ as β mixing with a polynomial rate c that is left unspecified because it is typically related to a moment condition on $m_t(h)$ for a central limit theorem to apply, and more generally to the complexity of the function space $\mathcal{H}$, as measured by its entropy[4] for a uniform central limit theorem to apply (which in turn is related to the stochastic equicontinuity assumption A5). For example, for a central limit theorem to apply, a sufficient condition is that $\sum_{t=1}^{\infty} t^{2/(\delta-2)} \beta(t) < \infty$, with $\| E(m_{t,K}(h_0)) \|^\delta < \infty$ for some $\delta > 2$, which is satisfied for $c > \delta / (\delta - 2)$. On the other hand, for a uniform central limit theorem to hold, the rate c crucially depends on the assumed structure and/or the (finite) entropy dimension of $\mathcal{H}$. For example, Doukhan, Massart and Rio (1995) showed that for α mixing (and hence β mixing) sequences a sufficient condition is that $c = 32d + 3$, where d is the entropy dimension of $\mathcal{H}$. By contrast, Arcones and Yu (1994) showed that if $\mathcal{H}$ is a Vapnick-Chervonenkis (V-C henceforth) subgraph class of functions (see Van der Vaart and Wellner (1996, Section 2.6.2) for a definition), then, a sufficient condition is that c satisfies $t^{\delta/(\delta-2} \log(t)^{2(\delta-1)/(\delta-2)} \beta(t) \to 0$, which is only slightly stronger than the rate given above for the central limit theorem to hold. Assumption A2(i) requires a suitable restriction on $\mathcal{A}$, such as assuming directly that it is a compact set. Given the local nature of the proposed estimation, recall that $a = h_0(u)$, the assumption of compactness is fairly natural; in fact, since Wong and Severini (1991, Theorem 1), it has been used extensively in the nonparametric and semiparametric literature, see for example Carroll, Fan, Gijbels and Wand (1997, Lemma A.1) and Fan and Zhang (2004, Section 5.1) among many others. Alternatively, compactness can be deduced indirectly, using various compact embedding results[5] - see Nickl and Potscher (2007) for some general statistical applications and Freyberger and Masten (2019) for more econometric oriented applications of such results. For example, assume for simplicity that $h : \mathcal{U} \to \mathbb{R}$ and $\mathcal{U}$ is bounded. Let $\mathcal{C}_m(\mathcal{U})$ denote the space of m times continuously differentiable functions h on $\mathcal{U}$, let $\mathcal{C}_{m,\infty} = \{ h \in \mathcal{C}_m(\mathcal{U}) : ||h||_{m,\infty} < \infty \}$, where $||.||_{m,\infty}$ is the Sobolev sup norm $\max_{0 \le \lambda \le m} \sup_{u \in \mathcal{U}} \| d^\lambda h / du^\lambda \|$ and assume that $\mathcal{H} = \mathcal{C}_{m+m_0,\infty}$ for some $m_0 \ge 1$. Then the embedding $\mathcal{C}_{m+m_0,\infty} \hookrightarrow \mathcal{C}_{m,\infty}$ is compact (Freyberger and Masten 2019,

[4]See the Appendix for a formal definition of entropy.

[5]A normed space $(\mathcal{X}, ||.||_\mathcal{X})$ is said to be *embedded* into the normed space $(\mathcal{Y}, ||.||_\mathcal{Y})$ if $\mathcal{X}$ is a linear subspace of $\mathcal{Y}$ and the identity map $id : \mathcal{X} \to \mathcal{Y}$ is continuous; such embedding is denoted as $(\mathcal{X}, ||.||_\mathcal{X}) \hookrightarrow (\mathcal{Y}, ||.||_\mathcal{Y})$. An embedding is said to be *compact* if the image (under the embedding operator) of the unit ball of $\mathcal{X}$ is totally bounded in $\mathcal{Y}$.

Theorem 1), hence we can take $\mathcal{A}$ as $\mathcal{C}_{m,\infty}$. Assumptions A2(ii)-(iii) are standard identification conditions that can be often verified by imposing more primitive conditions on $m_t(h)$ and/or some of the components of the random vector Z_t. In particular A2(iii) implies that when $a(v)$ (the coefficient of the approximation of the true h_0 for v in a neighborhood of $u \in \mathcal{U}$) the local EE is bounded away from 0. A similar assumption can be found for example in Zhang and Gjibels (2003, Assumption N.7) and Fan and Zhang (2004, Assumption A.10). Assumption A3(i) does not require $m_t(h,u)$ to be continuous at all $h \in \mathcal{H}$ for all Z_t, so it applies to possibly non smooth EE such as those defined in (9). Assumption A3(ii) is a high level assumption implying that a uniform law of large numbers applies. Given the continuity assumption on $m_t(h,u)$ in A3(i), A3(ii) will follow by a suitable restriction on $\mathcal{A}$ (and possibly on $\mathcal{U}$). For example, under the additional assumptions that $\mathcal{A}$ and $\mathcal{U}$ are compact sets, a finite envelope condition on $m_t(h)$ is sufficient for A3(ii) to hold - see the proof of Proposition 2 for an example of how to verify A3(ii). A2(ii)-(iii) and A3(ii) are used to establish the consistency of both the LGMM $\widehat{a}$ and the LGEL $\widehat{a}^{\rho}$ estimators, see the proofs of Theorems 1 and 2 in the Supplemental Appendix for more details; we note here that the assumption on the class of functions $\mathcal{M}_2^K$ in A3(ii) is not necessary for the consistency of $\widehat{a}$ unless we consider the two step LGMM estimator defined in Remark 2. Assumptions A2(iv), A4(i)-(ii) are standard in the nonparametric estimation literature, see for example Cai, Fan and Yao (2000, Condition A.1) and Fan and Zhang (2004, Assumption A.9) among many others; note that the differentiability of h_0 assumed in A2(iv) implicitly imposes a restriction on the function space $\mathcal{H}$, such as that it is a Sobolev or Holder function space. Assumption A4(iii) is a smoothness assumption at h_0 with "derivative" G being of full rank, that is typically assumed in the non smooth EE literature, see for example Pakes and Pollard (1989, Theorem 3.3, (ii)) for parametric EE models, and Chen, Linton and van Keilegom (2003, Theorem 2, (2.2)) for semiparametric EE models. Note that this condition is on the expectation of $m_t(h)$, that is the EE needs not to be smooth in h, although its expectation must be locally differentiable in the sense described in Remark 1. Since expectations involve integrals, hence they smooth out functions, it is frequently the case that $E[m_t(h)]$ is differentiable even though $m_t(h)$ is not. This is the case for the varying coefficients median regression model considered in the example of Section 2 and the more general specification of $m_t(h)$ given in (9) below. Assumption A4(iv) requires a central limit theorem to hold for $(Tb)^{1/2}\, v_{Tb}^K(h_0)$, which, as mentioned before, depends on the summability of the β-mixing rate of A1 and on the undersmoothing condition A7(iv) - Assumption E1 in Section 4 provides an example of such rate. We note here that in the proof of Proposition 2, we provide a general method to obtain the required asymptotic normality, which can be used for many other NPEE models. A5 is another high level assumption, as it implies the stochastic equicontinuity of the empirical process $(Tb)^{1/2}\, v_{Tb}^K(h)$. As previously mentioned, stochastic equicontinuity has been used in the context of non smooth parametric EE models, see for example Pakes and Pollard (1989, Theorem 3.3, (iii)), and non smooth semiparametric EE models, see for example Chen et al. (2003, Theorem 2, (2.5)), but, as far as we are aware of, not in the context of the type of non smooth NPEE models considered in this paper. In order to verify A5, one has typically to make further assumptions on both $m_t(h)$ and $\mathcal{H}$. For example, for non smooth varying coefficients models, the following

$$m_t(h) = \Psi_1\left(Y_t - X_t^\top h(U_t)\right)\Psi_2\left(Z_t, h(U_t)\right)X_t, \tag{9}$$

provides a fairly general specification of the possible forms that $m_t(h)$ can take; for example, for $\Psi_1(.) = sign_\theta(.)$ with $0 < \theta < 1$ and $\Psi_2(.) = 1$ one obtains a varying coefficients quantile regression model (an extension of which is considered in Section 4), whereas for $\Psi_2(.) = I\left(X_t^\top h(U_t) \geq 0\right)$ one obtains a varying coefficients extension to the censored quantile regression model proposed by Powell (1986). As another example, for $\Psi_1(.) = |sign_\theta(.)|$ and $\Psi_2(.) = Y_t - X_t^\top h(U_t)$ one obtains a varying coefficients extension to the asymmetric least squares estimator of Newey and Powell (1987). Next, we provide some examples of possible restrictions on $\mathcal{H}$. Recall that the differentiability condition A2(iv) implies that $\mathcal{H}$ must be a space of sufficiently smooth functions such as a Sobolev or a Holder space. For such function spaces various entropy results are available in the probability literature; for example, for $h : \mathcal{U} \to \mathbb{R}$ with $\mathcal{U}$ bounded, assume that $h \in \mathcal{H}^H_{C,\lambda}(\mathcal{U})$, that is for some $\underline{\lambda}$ being the largest integer smaller than λ, h is such that $||h||_{\infty,\lambda} \leq C$ for some finite positive C, with

$$||h||_{\infty,\lambda} = \max_{0\leq\lambda\leq\underline{\lambda}} \sup_{u\in\mathcal{U}} |D^\lambda h(u)| + \max_{\lambda=\underline{\lambda}} \sup_{u\neq u'\in\mathcal{U}} \frac{|D^\lambda h(u) - D^\lambda h(u')|}{|u-u'|^{\lambda-\underline{\lambda}}},$$

where $D^\lambda(h) = d^\lambda h/du^\lambda$. For this Holder function space, its entropy $H\left(\delta, \mathcal{H}^H_{C,\lambda}(\mathcal{U}), ||.||_\infty\right)$ is bounded by a constant$\times\delta^{-1/\lambda}$, which implies that the entropy integral $\int_0^\infty \left(H\left(\delta, \mathcal{H}^H_{C,\lambda}(\mathcal{U}), ||.||_\infty\right)\right)^{1/2} d\delta < \infty$. A similar result holds for the Sobolev function space $\mathcal{H}^S_{C,\lambda}(\mathcal{U})$, that is for $||h||_\infty \leq C_1$, $||h||_{p,\lambda} \leq C_2$ for some positive constants C_j, $(j = 1, 2)$, with $||h||_{p,\lambda} = \left(\int |D^\lambda h(u)|^p du\right)^{1/p} (1 \leq p < \infty)$, the bracketing entropy $H_{[]}\left(\delta, \mathcal{H}^S_{C,\lambda}(\mathcal{U}), ||.||_\infty\right)$ is bounded by a constant$\times\delta^{-1/\lambda}$, so that the bracketing entropy integral $\int_0^1 \left(H_{[]}\left(\delta, H^S_{C,\lambda}(\mathcal{U}), ||.||_\infty\right)\right)^{1/2} d\delta < \infty$. Under i.i.d. sampling, the finiteness of these entropy integrals, combined with an appropriate envelope condition on $m_t(h)$ would imply the stochastic equicontinuity of the related empirical processes $(Tb)^{1/2} v^K_{Tb}(h)$. However, the assumed weakly dependent structure of the observations requires some additional assumptions on the dependency structure and on $\mathcal{H}$. For β-mixing sequences, one such restriction relates the β-mixing coefficient to the tail behavior of h. To be specific let $Q_h(u)$ denote the quantile function of h and let $\lfloor v \rfloor$ denote the largest integer smaller or equal to v and define $\beta^{-1}(u) = \inf\{u : \beta(\lfloor v \rfloor) \leq u\}$. Then, Doukhan et al. (1995) showed that a sufficient condition for A5 to hold is that the bracketing entropy integral $\int_0^1 \left(H_{[]}\left(\delta, \mathcal{H}, ||.||_{2,\beta}\right)\right)^{1/2} d\delta$ is finite, where $||h||_{2,\beta} = \left(\int_0^1 \beta^{-1}(u) Q_h^2(u) du\right)^{1/2}$. In Section 4 we show how to verify the finiteness of the latter bracketing entropy integral under an appropriate restriction on the β-mixing coefficient.

Assumption A6 is standard in the GEL literature, see Newey and Smith (2004, Section 2); Assumptions A7(i)-(iii) are also standard in the nonparametric estimation literature, see for example Cai et al. (2000, Condition A.1); the undersmoothing condition A7(iv) is required for A4(iv) and A5 to hold. Note that the asymptotic bias of the kernel estimation is of the standard order $O\left(b^2\right)$, see Proposition 4 in the supplemental Appendix for more details. Finally, Assumption A8 is a standard regularity condition for GMM estimation.

The following two theorems establish the asymptotic normality of the LGMM and LGEL estimators.

Theorem 1 *Under the regularity conditions A1-A5 and A7-A8,*

$$(Tb)^{1/2}\left(\widehat{a}\left(u\right)-h_0\left(u\right)\right)\xrightarrow{d}N\left(0,\Sigma\left(u\right)_W^{-1}G\left(u\right)^{\top}W\left(u\right)\Omega\left(u\right)W\left(u\right)G\left(u\right)\Sigma\left(u\right)_W^{-1}\right);$$

Theorem 2 *Under the regularity conditions A1-A7,*

$$(Tb)^{1/2}\left(\widehat{a}^{\rho}\left(u\right)-h_0\left(u\right)\right)\xrightarrow{d}N\left(0,\Sigma\left(u\right)^{-1}\right).$$

Remark 2 *It is easy to see that the asymptotic covariance of the LGMM estimator is larger than that of the LGEL estimator (that is the difference between the two asymptotic covariances is a positive semidefinite matrix), unless $W\left(u\right)=\Omega\left(u\right)^{-1}$, which requires a 2 step local estimation procedure, where the first step is used to obtain a consistent consistent estimator of $\Omega\left(u\right)^{-1}$, say $\widetilde{\Omega}\left(u\right)^{-1}$, which is then used in the second step to obtain the so-called efficient LGMM (ELGMM) estimator with $\widetilde{\Omega}\left(u\right)^{-1}$ replacing $\widehat{W}\left(u\right)$ in (3). It is important to note that this two step LGMM estimation procedure has two disadvantages compared to LGEL estimation: first, it requires computing the bandwidth b twice, which can be time consuming for large data sets. Second, (and perhaps more importantly) it might result in additional (higher order) bias, a fact noted by Newey and Smith (2004) (for parametric overidentified smooth EE models) and by Bravo (2022) (for nonparametric overidentified smooth EE models).*

We conclude this section by proposing two novel overidentification test statistics that can be used to test the correct specification of (1). Let $\left\{u_j\right\}_{j=1}^{m}$ denote a set of local points such that (2) holds.

Proposition 1 *Under the assumptions of Theorem 2,*
(i) under the alternative local hypothesis

$$H_a: E\left[m_t\left(h_0\left(u_j\right)\right)|U_t=u_j\right]=\delta_{Tb}\left(u_j\right)$$

for some bounded function δ_{Tb} (whose form will depend on the specification of $m_t(h)$, see (12) below for an example) such that $(Tb)^{1/2}\delta_{Tb} \to C > 0$ as $Tb \to \infty$

$$\max_{j=1,\ldots,m} (Tb) \frac{1}{Tb} \sum_{t=1}^{T} m_{t,K}\left(\widehat{a}(u_j)\right)^{\top} \widehat{\Omega}(u_j)^{-1} \frac{1}{Tb} \sum_{s=1}^{T} m_{s,K}\left(\widehat{a}(u_j)\right) \tag{10}$$
$$\xrightarrow{d} \max_{j=1,\ldots,m} \chi_j^2(\kappa_j, l-k),$$
$$\max_{j=1,\ldots,m} (2Tb)\left(\frac{1}{Tb}\sum_{t=1}^{T} \rho\left(s_{t,K}\left(u_j, \widehat{\lambda}, \widehat{a}^{\rho}(u_j)\right)\right) - \frac{\rho_0}{b}\right) \xrightarrow{d} \max_{j=1,\ldots,m} \chi_j^2(\kappa_j, l-k),$$

where $\widehat{\Omega}(u_j) = \sum_{t=1}^{T}\left(m_t\left(\widehat{a}(u_j)\right) K\left(\frac{U_t-u_j}{b}\right)\right)^{\otimes 2}/Tb$, and $\chi_j^2(\kappa_j, l-k)$ are independent noncentral chi-squared variates with $l-k$ degrees of freedom and noncentrality parameter $\kappa_j = \delta(u_j)^{\top}\Omega(u_j)^{-1/2}\left(I - \Omega(u_j)^{-1/2} G(u_j)\Sigma(u_j)^{-1} G(u_j)^{\top}\Omega(u_j)^{-1/2}\right)\Omega(u_j)^{-1/2}\delta(u_j)$;
(ii) under the alternative global hypothesis H_a such that $(Tb)^{1/2}\delta_{Tb} \to \infty$ as $Tb \to \infty$

$$\max_{j=1,\ldots,m} (Tb) \frac{1}{Tb} \sum_{t=1}^{T} m_{t,K}\left(\widehat{a}(u_j)\right)^{\top} \widehat{\Omega}(u_j)^{-1} \frac{1}{Tb} \sum_{s=1}^{T} m_{s,K}\left(\widehat{a}(u_j)\right) \xrightarrow{p} \infty,$$
$$\max_{j=1,\ldots,m} (2Tb)\left(\frac{1}{Tb}\sum_{t=1}^{T} \rho\left(s_{t,K}\left(u_j, \widehat{\lambda}, \widehat{a}^{\rho}(u_j)\right)\right) - \frac{\rho_0}{b}\right) \xrightarrow{p} \infty.$$

4 Instrumental variables varying coefficients quantile regression

In this section, we extend the Example of Section 2 to an instrumental variables varying coefficients quantile regression model and illustrate how some of the regularity conditions A1-A8 can be verified under more primitive conditions. Let

$$Y_t = X_t^{\top} h_0(U_t) + \varepsilon_t,$$

where the unobservable error term ε_t does not satisfy the θ-th conditional quantile restriction $Pr(\varepsilon_t \leq 0|X_t, U_t) = \theta$ $a.s.$ because of the possible endogeneity of some of the covariates X_t, which implies that $E\left(X_t\left(sign_{\theta}\left(Y_t - X_t^{\top}h_0(U_t)\right)\right)|U_t\right) \neq 0$ $a.s.$ Suppose, however, that there exists an R^l- valued $(l > k)$ vector of instruments $\{V_t, t \in \mathbb{Z}\}$ such that $Pr(\varepsilon_t \leq 0|V_t, U_t) = \theta$ $a.s.$ Then the NPEE model is

$$E\left[V_t\left(sign_{\theta}\left(Y_t - X_t^{\top}h_0(U_t)\right)\right)|U_t\right] = 0\, a.s.,$$

and the LGMM and LGEL (quantile) estimators are based on

$$m_{t,K}^{q}(h) = V_t\left(sign_{\theta}\left(Y_t - X_t^{\top}h(u)\right)\right) K\left(\frac{U_t-u}{b}\right).$$

Let

$$G(u) = E[f_{\varepsilon_t|X_t,V_t}(0) V_t X_t^\top | U_t = u] f(u), \tag{11}$$
$$\Omega(u) = E\left[\theta(1-\theta) V_t^{\otimes 2} | U_t = u\right] \int K^2(v)\, dv f(u),$$

where $f_{\varepsilon_t|X_t,V_t}(0)$ is the conditional density of ε_t.[6]

Assume that:

E1 The sequence $\left\{\left(Y_t, X_t^\top, V_t^\top, U_t\right)^\top, t \in \mathbb{Z}\right\}$, is strictly stationary β mixing with mixing coefficient $\beta(t) = O(t^{-c})$, with $c = (2+\delta)(1+\delta)/\delta$ and δ specified in E3(ii) below.

E2 (i) $\Pr\left(Y_t - X_t^\top h_0(U_t) \le 0 | V_t, U_t\right) = \theta$ *a.s.* and $E\left(V_t^{\otimes 2} | U_t\right)$ is positive definite *a.s.*, (ii) the parameter space $\mathcal{A}$ is a compact set, (iii) $\mathcal{U}$ is a compact set, (iv) for $j = 1, ...k$ $h_j \in \mathcal{H}_{C,\lambda}^S(\mathcal{U})$ for $\lambda > 2$, (v) Assumption A2(iv) holds.

E3 For all v in a neighborhood of $u \in \mathcal{U}$, (i) $E\left\|V_t(1-\theta) K(v)\right\|^j < \infty$ $(j = 1, 2)$, (ii)

$$E\left(\sup_{a\in\mathcal{A}} \left\|\left[V_t\left(sign_\theta\left(Y_t - X_t^\top a\right)\right)\right]\right\|^{2(1+\delta)} | U_t = v\right) f(v) < \infty,$$

for some $\delta > 0$, (iii) for $t \ge 2$

$$E\left[\left(\left\|V_1\left(sign_\theta\left(Y_1 - X_1^\top h_0\right)\right)\right\|^2 + \left\|V_t\left(sign_\theta\left(Y_t - X_t^\top h_0\right)\right)\right\|^2\right) | U_1 = u_1, U_t = u_2\right] < \infty,$$

for u_1 and $u_2 \in \mathcal{U}$, (iv) for all v in a neighborhood of $u \in \mathcal{U}$, $rank(G(v)) = k$, with $G(v) = E[f_{\varepsilon_t|X_t,V_t}\left(X_t^\top h_0(v)\right) V_t X_t^\top | U_t = v] f(v)$ and $\Omega(v)$ is positive definite with $\Omega(v)$ given in (11), (v) $\Sigma(u)$ is nonsingular.

E4 Assumptions A6-A8 hold.

Assumption E1 assumes a mixing rate that is slightly stronger than the minimal condition on the summability of the mixing coefficients $\sum_{t=1}^{\infty} t^{2/(2-\delta)}\alpha(t) < \infty$ for a central limit theorem for α mixing random variables (and hence β mixing) to hold (see for example Doukhan (1994)), however it is sufficient to verify the stochastic equicontinuity Assumption A5. Assumption E2(i) implies the identification conditions A2(ii)-(iii) by standard arguments, see the proof

[6] Note that by iterated expectations and a standard kernel calculation

$$\frac{\partial}{\partial h^\top} E[\frac{1}{b} m_{t,K}^q(h)] = \frac{\partial}{\partial h^\top} E[\frac{V_t}{b}(\theta - F_{Y_t|X_t,V_t}\left(Y_t - X_t^\top h(U_t)\right) | U_t = u] K \frac{(U_t - u)}{b} |_{h=h_0} =$$

$$\int E\left(f_{\varepsilon_t|X_t,V_t}(0) V_t X_t^\top | U_t = u + vb\right) K(v) f(u + bv)\, dv := G(u).$$

of Proposition 2 in the supplemental Appendix for more details; Assumption E2(ii) can be deduced by the compact embedding $\mathcal{H}_{C,\lambda}^{S}(\mathcal{U}) \hookrightarrow \mathcal{H}_{C,2}^{S}(\mathcal{U}) := \mathcal{A}$; E2(iii) is often assumed in the nonparametric and semiparametric estimation literature, see for example Carroll et al. (1997) and Masry (1996), to obtain the uniform consistency and more generally the convergence rates of the nonparametric estimators. Here it is used (in combination with E2(ii)) to verify the Glivenko-Cantelli assumption A3(ii), but we note that it could be relaxed at the cost of lengthier proofs based on a truncation argument and additional regularity conditions on the kernel used in the estimation, see for example Hansen (2008). Assumption E2(iv) requires that h belongs to a Sobolev function space; it implies the differentiability assumption A2(iv) on h_0 and is used to verify the stochastic equicontinuity A5 for the empirical process $(Tb)^{1/2} v_{Tb}^{K}(h)$ based on $m_{t,K}^{q}(h)$. Assumption E3(i) gives the envelopes for the function classes

$$\mathcal{Q}_1^K = \left\{ V_t\left(sign_\theta\left(Y_t - X_t^\top a\right)\right) K\left(\frac{U_t - v}{b}\right), v \in \mathcal{U},\ a \in \mathcal{A}, \right\} \text{ and}$$
$$\mathcal{Q}_2^K = \left\{ \left(V_t\left(sign_\theta\left(Y_t - X_t^\top a\right)\right) K\left(\frac{U_t - v}{b}\right)\right)^{\otimes 2},\ v \in \mathcal{U}, a \in \mathcal{A}, \right\}.$$

E3(iii) is used to prove the central limit theorem for $(Tb)^{1/2} v_{Tb}^{K}(h_0)$. The rest of the assumptions are standard, see the discussion of the corresponding assumptions in Section 3.

Proposition 2 *Under assumptions E1- E4 the conclusions of Theorems 1 and 2 hold with* $G(u)$ *and* $\Omega(u)$ *given in* (11).

We conclude this section by showing how the overidentifying restriction test statistics of Proposition 1 specializes to this example. In this case the local alternative hypothesis is

$$H_a : E\left[V_t\left(sign_\theta\left(Y_t - X_t^\top\left(h_0(U_t) - \frac{\delta(U_t)}{(Tb)^{1/2}}\right)\right)\right)|U_t = u\right] = 0,$$

for some bounded function δ, which, by iterated expectations and a standard Taylor expansion, can be written as

$$H_a : E\left[V_t\left(sign_\theta\left(Y_t - X_t^\top h_0(U_t)\right)\right)|U_t = u\right] = E\left(f_{\varepsilon_t|X_t,V_t}(0) V_t X_t^\top |U_t = u\right)\frac{\delta(u)}{(Tb)^{1/2}} := \delta_{Tb}(u). \tag{12}$$

5 Simulations results

The results of Theorems 1 and 2 require undersmoothing, hence least squares cross validation or other bandwidth selection methods cannot be used directly to automatically choose the bandwidth b. In this paper we consider a bandwidth selection procedure that is similar to the ad-hoc cross validation method of Otsu, Xu and Matsushita (2015) but is less computationally intensive. Specifically, we consider a two fold cross validation procedure, which consists of computing for a random subset of the sample, the training set S_v with $0 < v < 1$, and a pilot bandwidth b_p

$$\widehat{a}_p = \arg\min_{a\in\mathcal{A}} \frac{1}{T_v b_p} \sum_{t\in S_v} m_t(a)^\top K\left(\frac{U_t - u}{b_p}\right) \widehat{\Omega}(u)^{-1} \frac{1}{T_v b_p} \sum_{s\in S_v} m_s(a) K\left(\frac{U_s - u}{b_p}\right),$$
$$\widehat{a}_p^\rho = \arg\min_{a\in\mathcal{A}} \frac{1}{T_v b_p} \sum_{t\in S_v} \rho\left(s_{t,K}\left(u, \widehat{\lambda}, a\right)\right),$$

and then using the remaining part of the sample, the validation set S_{1-v}, to select the bandwidth as

$$\widehat{b} = \arg\min_{b\in B} \frac{1}{T_{1-v} b} \sum_{t\in S_{1-v}} m_{t,K}(\widehat{a}_p)^\top \widehat{\Omega}(u)^{-1} \frac{1}{T_{1-v} b} \sum_{s\in S_{1-v}} m_{s,K}(\widehat{a}_p)$$
$$\widehat{b}^\rho = \arg\min_{b\in B} \frac{1}{T_{1-v} b} \sum_{t\in S_{1-v}} \rho\left(s_{t,K}\left(u, \widehat{\lambda}, \widehat{a}_p^\rho\right)\right), \tag{13}$$

where B is a grid of possible values of b, and $\widehat{a}_p$ and $\widehat{a}_p^\rho$ are the estimators based on the pilot bandwidth b_p. In the simulations we use $v = 0.8$, which seems a commonly used value in the literature as it corresponds to the so-called Pareto principle[7], with 80% of the observations randomly chosen from the sample for the training set S_v and the remaining 20% for the validation set S_{1-v}. Finally, as in Otsu et al. (2015), $\widehat{a}$ and $\widehat{a}^\rho$ are multiplied by T^{-c}, where $c > 0$ is a value consistent with undersmoothing. In the simulations below we use the value $c = 0.3$, however to assess the sensitivity of the estimators to such choice, we also consider $c = [0.2, 0.4, 0.6]$, see Tables 7-10 in Section B of the supplemental Appendix for the results and some additional comments.

We consider a varying coefficients quantile regression model with an endogenous regressor

$$Y_t = h_{10}(U_t) + X_{1t} h_{20}(U_t) + \varepsilon_t,$$
$$X_{1t} = 0.4 X_{2t} + \eta_t,$$

where $h_{10}(U_t) = \sin(\pi U_t/2)$, $h_{20} = \cos(\pi U_t/3)$, $X_{2t} = \rho X_{2t-1} + \zeta_t$ with $\zeta_t \sim N(0,1)$ independent of

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right),$$

and U_t is either $U_t \sim U(0,1)$ or $U_t \sim \Phi\left((a\xi_t + b\xi_{t-1})/\sqrt{a^2+b^2}\right)$, where Φ is the cumulative standard normal distribution, $\xi_t \sim N(0,1)$, $a = 0.9$ and $b = 0.1$, which implies that U_t is

[7] The Pareto principle states that 80% of outputs comes from 20% of inputs; it is often used in fields such as Computer Science, Economics, Engineering and Quality Control.

a 1-dependent process. In the simulations, we use the Epanechnikov kernel, i.e. $K(u) = (3/4)(1-u^2)$ for $|u| \leq 1$, the vector of instruments V_t is specified as either $[1, X_{2t}, X_{2t}^2]^\top$ or $[1, X_{2t}, ..., X_{2t}^6]^\top$, the autoregressive coefficient ρ is either 0.4 or 0.9, the bandwidths $\widehat{b}$ and $\widehat{b}^\rho$ are chosen using the two fold cross validation method (13) with sample sizes $T = 200$ and $T = 400$. We consider three estimators, LEL $\widehat{a}^{el}$ defined in (5), LET $\widehat{a}^{et}$ defined in (6) and ELGMM $\widehat{a}$ defined in (3) with a preliminary consistent $\widetilde{\Omega}(u)^{-1}$ replacing $\widehat{W}(u)$ and use the combined mean squared error (MSE)

$$MSE(\widehat{a}^\circ) = \sum_{j=1}^{2} \frac{1}{T} \sum_{t=1}^{T} \left(\widehat{a}_j^\circ(U_t) - h_j^\circ(U_t) \right)^2,$$

where $\widehat{a}_j^\circ$ is any of the three local estimators LEL, LET and ELGMM for the unknown functional parameters h_{j0}, to evaluate their finite sample performance for $\theta = [0.25, 0.50, 0.75]$.

Tables 1-4 approximately here

The results of Tables 1-4 can be summarized as follows: when the dimension of the instruments V_t ($\dim(V_t)$) is 3 (that is the degree of overidentification is 1) the MSE of the proposed estimators are broadly comparable, regardless of the degree of persistence of the instrument X_{2t} and the error specification and are decreasing with the sample size, which is consistent with the asymptotic theory of Section 3. However, when $\dim(V_t) = 7$, the MSE of the ELGMM estimator is considerably worse than that of both the LEL and LET estimators, especially that of the LEL estimator, which seems to confirm the findings of Newey and Smith (2004) and Bravo (2022) that for ordinary (parametric) and local (nonparametric) GMM estimators their bias (and hence MSE) is increasing with the dimension of the instruments. Figures 1-2 show the estimated varying coefficients for $\theta = [0.25, 0.50, 0.75]$, $U_t \sim \Phi\left((a\xi_t + b\xi_{t-1})/\sqrt{a^2+b^2}\right)$, $\rho = 0.9$ with $T = 200$, when $\dim(V_t) = 3$ and $\dim(V_t) = 7$, respectively. Figure 2 clearly shows a significant bias in the ELGMM estimator.

Figures 1-2 approximately here.

Next, we investigate the finite sample properties of the test statistics (10) of Proposition 1, using the same three estimators. The null hypothesis is

$$E\left[V_t sign_\theta\left(Y_t - h_{10}(u_j) - X_{1t}(h_{20}(u_j)) - \delta(u_j)\right) | U_t = u_j\right] = 0$$

where h_{10} and h_{20} are as above, $\delta(u_j) = \delta \cos(\pi u_j/3)$ with $\delta = [0, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8]$ and $j = 20$ (equally spaced over the U_t's). Tables 5-6 report, respectively, the finite sample sizes (corresponding to $\delta = 0$) for $U_t \sim U(0,1)$ or $U_t \sim \Phi((a\xi_t + b\xi_{t-1})/0.9)$, $\rho = 0.9$ and $T = 200$. The simulated critical values at the $[0.10, 0.05]$ nominal level (based on 10^5 simulations) for the test statistics are $[3.351, 4.803]$ when the degree of overidentification is 1, and $[10.298, 12.507]$

when the degree of overidentification is 5.

Tables 5-6 approximately here.

Tables 5-6 show a similar pattern to that of Tables 1-4, in the sense that, when $\dim(V_t) = 3$, the three test statistics (10) have similar finite sample sizes, however, when $\dim(V_t) = 7$, the finite sample size of the ELGMM based test statistic is considerably worse compared to that of the other two LGEL based statistics. Figures 3-4 show the size adjusted finite sample power of the three test statistics (10) for $\theta = [0.25, 0.50, 0.75]$, $U_t \sim \Phi\left(\left(a\xi_t + b\xi_{t-1}\right)/\sqrt{a^2+b^2}\right)$, $\rho = 0.9$ and $T = 200$, when $\dim(V_t) = 3$ and $\dim(V_t) = 7$, respectively. Again, it is clear that the degree of overidentification has a significant negative effect on the finite sample power of the ELGMM based test statistic.

6 Empirical application

We consider a varying coefficients extension of the log-linearized version of the quantile based consumption capital assets pricing model (C-CAPM) of de Castro, Galvao, Kaplan and Liu (2019) and de Castro and Galvao (2019). To be specific, let $Z_t = \left[C_t, R_t, Inf_t, NR_t, PD_t\right]^\top$, where, respectively, C_t is real total per capita consumption, R_t is the real interest rate - deflated by the consumer price index, Inf_t is the inflation rate, NR_t is the nominal interest rate and PD_t is the price dividend ratio for equities, all at time t. Then

$$m_t^q(h) = \begin{bmatrix} 1 \\ \log \frac{C_{t-2}}{C_{t-3}} \\ NR_{t-2} \\ \log PD_{t-2} \end{bmatrix} \left(sign_\theta \left(\log \frac{C_t}{C_{t-1}} - h_1(u) - h_2(u) \log(1+R_t) \right) \right) K\left(\frac{Inf_{t-2} - u}{b} \right), \tag{14}$$

is the corresponding instrumental variables varying coefficients quantile regression model, which relates consumption growth to the real interest rate using the inflation rate as the varying coefficient and the twice lagged consumption growth, nominal interest rate and log dividend ratio as instruments. We use United Kingdom (UK) data originally from Campbell (2003), which consists of aggregate level quarterly data for the period 1970Q3–1999Q1. Figure 5 shows the consumption growth, inflation rate and $\log(1+R_t)$.

Figure 5 approximately here.

Figures 6 and 7 show, respectively, the LEL and ELGMM estimated quantile varying coefficients at the $\theta = [0.25, 0.50, 0.75]$ quantiles with bandwidths calculated using the two fold procedures described in Section 5.

Figures 6-7 approximately here

The left panel of Figure 6 shows the first varying coefficient plotted versus the consumption growth; as expected, it shows that consumption growth is overall negatively affected by the inflation rate, albeit in a rather nonlinear way, which reflects the high volatility of the inflation rate especially in the Seventies, early Eighties and early Nineties. The right panel of Figure 6 shows the second estimated varying coefficient overlaid on the plot of the consumption growth versus $\log(1 + R_t)$ - the interaction variable in (14). In this case, following the consumption growth negative shock of 1979, it is first increasing in the inflation rate when the real interest rate goes from negative to positive, which can be explained by the fact that the nominal interest rate grew more than the inflation rate (a fact known as the Fisher effect in economic theory), for then showing a more variable pattern reflecting the volatility of the the inflation and real interest rate. Figure 7 shows a similar pattern as that of Figure 6, but with a visible bias in terms of the fitting of the two estimated varying coefficients over the plots of the actual data. Finally, we test for the correct specification of (14) using the test statistics (10). We use $j = 15$, which, with the degree of overidentification equaling 2, gives simulated critical values at the $[0.10, 0.05, 0.01]$ nominal level (based on 10^5 simulations) of $[5.506, 7.310, 11.536]$. The sample values of the LEL and ELGMM statistics are, respectively, 10.992 and 6.865 with corresponding p-values of 0.0166 and 0.08637, which suggests that the null hypothesis of correct specification of (14) cannot be rejected at the 0.05 level for the LEL statistic and at the 0.01 level for the ELGMM statistic. As a matter of comparison, the standard Hansen's (1982) overidentifying test statistic for the correct specification of the parametric quantile C-CAPM model used by de Castro et al. (2019) has a p-value of 0.003.

7 Conclusions

In this paper we consider local versions of GMM (LGMM) and GEL (LGEL) estimators for non smooth overidentified NPEE models with weakly dependent observations. We show that the proposed estimators are asymptotically normal under a β mixing and a set of relatively high level assumptions, and provide an example where we show how these high level assumptions can be verified under more primitive conditions. We also propose new LGEL and ELGMM based tests for the correct specification of NPEE models that are characterized by a nonstandard but asymptotically pivotal distribution and can detect local (at the nonparametric rate) alternatives and are consistent. Monte Carlo simulations show that LGEL estimators perform better than ELGMM estimators and related specification tests, in particular for models with larger degrees of overidentification. An empirical application, where a varying coefficients specification of a quantile C-CAPM is estimated, illustrates the applicability and usefulness of the local estimation method proposed in this paper.

8 Aknowledgements

I am grateful to a Co-Editor and three referees for constructive criticism and a number of useful suggestions that improved considerably the original draft. The usual disclaimer applies.

References

Andrews, D. and Pollard, D.: 1994, An introduction to functional central limit theorems for dependent stochastic processes, *International Statistical Review* **62**, 119–132.

Arcones, M. and Yu, B.: 1994, Central limit theorems for empirical and U processes of stationary mixing sequences, *Journal of Theoretical Probability* **7**, 47–71.

Bravo, F.: 2016, Local information theoretic methods for smooth coefficients dynamic panel data models, *Journal of Time Series Analysis* **37**, 690–708.

Bravo, F.: 2022, Second order expansions of estimators in nonparametric moment conditions models with weakly dependent data, *Econometric Reviews* **41**, 583–606.

Cai, Z.: 2003, Nonparametric estimating equations for time series data, *Statistics and Probability Letters* **62**, 379–390.

Cai, Z., Fan, J. and Yao, Q.: 2000, Functional-coefficient regression models for nonlinear time series, *Journal of the American Statistical Association* **95**, 941–956.

Cai, Z. and Li, Q.: 2008, Nonparametric estimation of varying coefficient dynamic panel data models, *Econometric Theory* **24**, 1321–1342.

Cai, Z., Ren, Y. and Sun, L.: 2015, Pricing kernel estimation: a local estimating equation approach, *Econometric Theory* **31**, 560–580.

Cai, Z. and Xu, X.: 2008, Nonparametric quantile estimation for dynamicsmooth coefficient models, *Journal of the American Statistical Association* **103**, 1595–1608.

Campbell, J.: 2003, Consumption-based asset pricing models, *in* G. Constantinides, M. Harris and R. Stulz (eds), *Handbook of the Economics of Finance: Financial Markets and Asset Pricing*, pp. 803–887.

Carrasco, M. and Chen, X.: 2002, Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory* **18**, 17–39.

Carroll, R., Fan, J., Gijbels, I. and Wand, M.: 1997, Generalized partially linear single index models, *Journal of the American Statistical Association* **92**, 477–489.

Chen, S. and Hall, P.: 1993, Smoothed empirical likelihood confidence intervals for quantiles, *Annals of Statistics* **21**, 1166–1181.

Chen, X., Linton, O. and van Keilegom, I.: 2003, Estimation of semiparametric models when the criterion fuction is not smooth, *Econometrica* **71**, 1591–1608.

Cuzik, J.: 1988, Rank regression, *Annals of Statistics* **16**, 1369–1389.

Davydov, Y.: 1973, Mixing properties of markov chains, *Theory of Probability and its Applications* **XVIII**, 312–328.

de Castro, L. and Galvao, A.: 2019, Dynamic quantile models of rational behavior, *Econometrica* **87**, 1893–1939.

de Castro, L., Galvao, A., Kaplan, D. and Liu, X.: 2019, Smoothed GMM for quantile models, *Journal of Econometrics* **213**, 121–144.

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

Doukhan, P., Massart, P. and Rio, E.: 1995, Invariance principles for absolutely regular processes, *Annales de l' Institut Henri Poincare* **31**, 393–427.

Fan, J., Hu, T. and Truong, Y.: 1994, Robust non-parametric function estimation, *Scandinavian Journal of Statistics* **21**, 433–446.

Fan, J. and Zhang, J.: 2004, Sieve empirical likelihood ratio tests for nonparametric functions, *Annals of Statistics* **32**, 1858–1907.

Fang, Y., Ren, Y. and Yuan, Y.: 2011, Nonparametric estimation and testing of stochastic discount factor, *Finance Research Letters* **8**, 196–205.

Freyberger, J. and Masten, M.: 2019, A practical guide to compact infinite dimensional parameter spaces, *Econometric Reviews* **38**, 979–1006.

Hansen, B.: 2008, Uniform convergence rates for kernel estimation with dependent data, *Econometric Theory* **24**, 726–748.

Hansen, L.: 1982, Large sample properties of generalized method of moments estimators, *Econometrica* **50**, 1029–1054.

Horowitz, J.: 1998, Bootstrap methods for median regression models, *Econometrica* **66**, 1327–1351.

Koenker, R. and Bassett, G.: 1978, Regression quantiles, *Econometrica* **46**, 33–50.

Lewbel, A.: 2007, A local generalized method of moments estimator, *Economics Letter* **94**, 124–128.

Masry, E.: 1996, Multivariate local polynomial regression for time series: uniform strong consistency and rates, *Journal of Time Series Analysis* **17**, 571–599.

Mokkadem, A.: 1990, Proprietes de melanges des modeles autoregressifs polynomiaux, *Annales de l'Institut Henri Poincare* **26**, 219–260.

Newey, W. and McFadden, D.: 1994, Large sample estimation and hypothesis testing, *in* R. Engle and D. McFadden (eds), *Handbook of Econometrics, Vol IV*, North Holland.

Newey, W. and Powell, J.: 1987, Asymmetric least squares estimation and testing, *Econometrica* **55**, 819–847.

Newey, W. and Smith, R.: 2004, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* **72**, 219–256.

Nickl, R. and Potscher, D.: 2007, Bracketing metric entropy rates and empirical central limit theoremsfor fuction classes of Besov and Sobolev-type, *Journal of Theoretical Probability* **20**, 177–199.

Otsu, T.: 2008, Conditional empirical likelihood estimation and inference for quantile regression models, *Journal of Econometrics* **142**, 508–538.

Otsu, T., Xu, K.-L. and Matsushita, Y.: 2015, Empirical likelihood for regression discontinuity design, *Journal of Econometrics* **186**, 94–112.

Pakes, A. and Pollard, D.: 1989, Simulation and the asymptotics of optmization estimators, *Econometrica* .

Patton, A.: 2012, A review of copula models for economic time series, *Journal of Multivariate Analysis* **110**, 4–18.

Pepe, M.: 1997, A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing, *Biometrika* **84**, 595–608.

Powell, J.: 1986, Censored regression quantiles, *Journal of Econometrics* **32**, 143–155.

Qu, A., Lindsay, B. G. and Li, B.: 2000, Improving generalised estimating equations using quadratic inference functions, *Biometrika* **87**(4), 823–836.

Severini, T. and Staniswalis, J.: 1994, Quasi-likelihood estimation in semiparametric models, *Journal of The American Statistical Association* **89**, 501–511.

Van der Vaart, A. and Wellner, J.: 1996, *Weak Convergence and Empirical Processes*, Springer, New York.

Volkonskii, V. and Rozanov, Y.: 1959, Some limit theorems for random functions I, *Theory of Probability and its Applications* **4**, 1978–197.

Whang, J.: 2006, Smoothed empirical likelihood methods for quantiles regression models, *Econometric Theory* **22**, 173–205.

Wong, W. and Severini, T.: 1991, On maximum likelihood estimation in infinite dimensional spaces, *Annals of Statistics* **19**, 603–632.

Yu, K. and Jones, M.: 1998, Local linear quantile regression, *Journal of the American Statistical Association* **93**, 228–237.

Zhang, J. and Gjibels, I.: 2003, Sieve empirical likelihood and extensions of the generalized least squares, *Scandinavian Journal of Statistics* **30**, 1–24.

Appendix

This Appendix presents *formal definitions* of some of the most technical concepts used in the paper. They are based on modern empirical process theory, see for example Van der Vaart and Wellner (1996).
Let $(\mathcal{F}, ||.||_{\mathcal{F}})$ denote a metric space of real valued functions $f : \mathcal{X} \to \mathbb{R}$.

Definition 1 (Covering number) *The covering number $N(\delta, \mathcal{F}, ||.||_{\mathcal{F}})$ is the minimal number N for which there exist δ balls $\{g : ||g - f_j||_{\mathcal{F}} \leq \delta\, ||.||_{\mathcal{F}}, j = 1, ..., N\}$ to cover $\mathcal{F}$.*

Definition 2 (Covering number with bracketing) *The covering number with brackets $N_{[]}(\delta, \mathcal{F}, ||.||_{\mathcal{F}})$ is the minimal number of N for which there exist δ brackets $\{[l_j, u_j] : ||l_j - u_j||_{\mathcal{F}} \leq \delta, ||l_j||_{\mathcal{F}} < \infty, ||u_j||_{\mathcal{F}} < \infty, j = 1, ...N\}$ to cover $\mathcal{F}$.*

Definition 3 (Entropy) *(i) The (metric) entropy of $\mathcal{F}$ is $H(\delta, \mathcal{F}, ||.||_{\mathcal{F}}) = \log N(\delta, \mathcal{F}, ||.||_{\mathcal{F}})$; (ii) the entropy with bracketing of $\mathcal{F}$ is $H_{[]}(\delta, \mathcal{F}, ||.||_{\mathcal{F}}) = \log N_{[]}(\delta, \mathcal{F}, ||.||_{\mathcal{F}})$.*

Definition 4 (Glivenko-Cantelli class of functions) *If either (i) $\sup_{f \in \mathcal{F}} N(\delta, \mathcal{F}, L_1(Q)) < \infty$ and $E||f||_{\mathcal{F}} < \infty$, or (ii) $N_{[]}(\delta, \mathcal{F}, L_1(P)) < \infty$, then $\mathcal{F}$ is a Glivenko-Cantelli class. As a result, for $\widehat{f} = \sum_{t=1}^{T} f_t / T$,*

$$||\widehat{f} - E(f_t)||_{\mathcal{F}} = o_p(1).$$

Definition 5 (Stochastic equicontinuity) *Let* $v_T(f) = \sum_t^T (f_t - E(t_t))/T^{1/2}$*; if either (i)* $\int_0^1 \sup_Q (H(\delta||F||_{Q,2}, \mathcal{F}, L_2(Q)))^{1/2} d\delta < \infty$*, or (ii)* $\int_0^1 \left(H_{[]}(\delta, \mathcal{F}, L_2(P))\right)^{1/2} d\delta < \infty$*, then for* $f, g \in \mathcal{F}$ *and* $\delta_T \to 0$

$$\sup_{\rho(f,g)\leq\delta_T} |v_T(f) - v_T(g)| = o_p(1)$$

for the semi-metric $\rho(f,g) = Var(f-g)$.

Figures and Tables

Table 1. Combined MSE of the local estimators with $U_t \sim U(0,1)$, $\dim(V_t) = 3$.

		$\widehat{a}^{el}$	$\widehat{a}^{et}$	$\widehat{a}$
$\rho = 0.4$				
	$\theta = 0.25$	0.038	0.039	0.041
$T = 200$	$\theta = 0.50$	0.033	0.035	0.038
	$\theta = 0.75$	0.036	0.037	0.041
	$\theta = 0.25$	0.020	0.022	0.024
$T = 400$	$\theta = 0.50$	0.018	0.019	0.026
	$\theta = 0.75$	0.021	0.022	0.026
$\rho = 0.9$				
	$\theta = 0.25$	0.040	0.040	0.044
$T = 200$	$\theta = 0.50$	0.035	0.037	0.040
	$\theta = 0.75$	0.037	0.038	0.043
	$\theta = 0.25$	0.023	0.024	0.026
$T = 400$	$\theta = 0.50$	0.020	0.022	0.028
	$\theta = 0.75$	0.021	0.023	0.027

Table 2. Combined MSE of the local estimators with $U_t \sim \dfrac{\Phi\left(0.9\xi_t + 0.1\xi_{t-1}\right)}{0.9}$, $\dim\left(V_t\right) = 3$.

		$\widehat{a}^{el}$	$\widehat{a}^{et}$	$\widehat{a}$
$\rho = 0.4$				
	$\theta = 0.25$	0.039	0.038	0.042
$T = 200$	$\theta = 0.50$	0.032	0.036	0.039
	$\theta = 0.75$	0.037	0.038	0.043
	$\theta = 0.25$	0.024	0.026	0.028
$T = 400$	$\theta = 0.50$	0.020	0.022	0.027
	$\theta = 0.75$	0.023	0.024	0.027
$\rho = 0.9$				
	$\theta = 0.25$	0.041	0.042	0.046
$T = 200$	$\theta = 0.50$	0.036	0.038	0.045
	$\theta = 0.075$	0.036	0.037	0.047
	$\theta = 0.25$	0.025	0.026	0.029
$T = 400$	$\theta = 0.50$	0.021	0.023	0.029
	$\theta = 0.75$	0.023	0.027	0.030

Table 3. Combined MSE of the local estimators with $U_t \sim U\left(0,1\right)$, $\dim\left(V_t\right) = 7$.

		$\widehat{a}^{el}$	$\widehat{a}^{et}$	$\widehat{a}$
$\rho = 0.4$				
	$\theta = 0.25$	0.039	0.064	0.124
$T = 200$	$\theta = 0.50$	0.035	0.066	0.128
	$\theta = 0.75$	0.037	0.066	0.126
	$\theta = 0.25$	0.022	0.037	0.068
$T = 400$	$\theta = 0.50$	0.020	0.037	0.066
	$\theta = 0.75$	0.023	0.038	0.067
$\rho = 0.9$				
	$\theta = 0.25$	0.042	0.047	0.134
$T = 200$	$\theta = 0.50$	0.038	0.042	0.130
	$\theta = 0.75$	0.039	0.044	0.133
	$\theta = 0.25$	0.026	0.040	0.071
$T = 400$	$\theta = 0.50$	0.024	0.028	0.069
	$\theta = 0.75$	0.024	0.027	0.072

Table 4. Combined MSE of the local estimators with $U_t \sim \dfrac{\Phi\left(0.9\xi_t + 0.1\xi_{t-1}\right)}{0.9}$, $\dim\left(V_t\right) = 7$.

		$\widehat{a}^{el}$	$\widehat{a}^{et}$	$\widehat{a}$
$\rho = 0.4$				
$T = 200$	$\theta = 0.25$	0.041	0.045	0.131
	$\theta = 0.50$	0.037	0.047	0.128
	$\theta = 0.75$	0.039	0.048	0.130
$T = 400$	$\theta = 0.25$	0.027	0.030	0.073
	$\theta = 0.50$	0.024	0.028	0.075
	$\theta = 0.75$	0.025	0.029	0.076
$\rho = 0.9$				
$T = 200$	$\theta = 0.25$	0.041	0.042	0.136
	$\theta = 0.50$	0.036	0.038	0.131
	$\theta = 0.75$	0.036	0.037	0.133
$T = 400$	$\theta = 0.25$	0.025	0.026	0.079
	$\theta = 0.50$	0.021	0.023	0.076
	$\theta = 0.75$	0.023	0.027	0.077

Table 5. Finite sample size of the test statistics (10) for $U_t \sim U\left[0,1\right]$.

		$\widehat{a}^{el}$		$\widehat{a}^{et}$		$\widehat{a}$	
				$\dim\left(V_t\right) = 3$			
$\rho = 0.4$	$\theta = 0.25$	0.119^a	0.058^b	0.120^a	0.060^b	0.128^a	0.067^b
	$\theta = 0.50$	0.117^a	0.058^b	0.121^a	0.061^b	0.130^a	0.066^b
	$\theta = 0.75$	0.118^a	0.059^b	0.119^a	0.062^b	0.131^a	0.068^b
$\rho = 0.9$	$\theta = 0.25$	0.120^a	0.059^b	0.121^a	0.061^b	0.130^a	0.072^b
	$\theta = 0.50$	0.121^a	0.058^b	0.122^a	0.062^b	0.132^a	0.073^b
	$\theta = 0.75$	0.119^a	0.060^b	0.121^a	0.063^b	0.131^a	0.073^b
				$\dim\left(V_t\right) = 7$			
$\rho = 0.4$	$\theta = 0.25$	0.122^a	0.059^b	0.123^a	0.062^b	0.138^a	0.074^b
	$\theta = 0.50$	0.119^a	0.059^b	0.123^a	0.063^b	0.140^a	0.073^b
	$\theta = 0.75$	0.120^a	0.058^b	0.122^a	0.064^b	0.142^a	0.074^b
$\rho = 0.9$	$\theta = 0.25$	0.123^a	0.060^b	0.124^a	0.062^b	0.140^a	0.063^b
	$\theta = 0.50$	0.122^a	0.061^b	0.124^a	0.061^b	0.144^a	0.062^b
	$\theta = 0.75$	0.122^a	0.060^b	0.123^a	0.061^b	0.143^a	0.063^b

a 0.10 nominal level, b 0.05 nominal level.

Table 6. Finite sample size of the test statistics (10) for $U_t \sim \dfrac{\Phi\left(0.9\xi_t + 0.1\xi_{t-1}\right)}{0.9}$.

		$\widehat{a}^{el}$		$\widehat{a}^{et}$		$\widehat{a}$	
				$\dim\left(V_t\right) = 3$			
$\rho = 0.4$	$\theta = 0.25$	0.121^a	0.059^b	0.122^a	0.061^b	0.130^a	0.070^b
	$\theta = 0.50$	0.119^a	0.059^b	0.124^a	0.060^b	0.132^a	0.069^b
	$\theta = 0.75$	0.120^a	0.060^b	0.122^a	0.062^b	0.132^a	0.069^b
$\rho = 0.9$	$\theta = 0.25$	0.122^a	0.060^b	0.123^a	0.063^b	0.132^a	0.074^b
	$\theta = 0.50$	0.122^a	0.061^b	0.124^a	0.062^b	0.133^a	0.074^b
	$\theta = 0.75$	0.121^a	0.062^b	0.123^a	0.062^b	0.134^a	0.075^b
				$\dim\left(V_t\right) = 7$			
$\rho = 0.4$	$\theta = 0.25$	0.123^a	0.060^b	0.124^a	0.062^b	0.141^a	0.076^b
	$\theta = 0.50$	0.124^a	0.059^b	0.126^a	0.063^b	0.144^a	0.075^b
	$\theta = 0.75$	0.122^a	0.059^b	0.125^a	0.062^b	0.145^a	0.077^b
$\rho = 0.9$	$\theta = 0.25$	0.124^a	0.061^b	0.125^a	0.063^b	0.144^a	0.080^b
	$\theta = 0.50$	0.125^a	0.062^b	0.125^a	0.064^b	0.146^a	0.079^b
	$\theta = 0.75$	0.126^a	0.061^b	0.127^a	0.064^b	0.148^a	0.079^b

a 0.10 nominal level, b 0.05 nominal level.

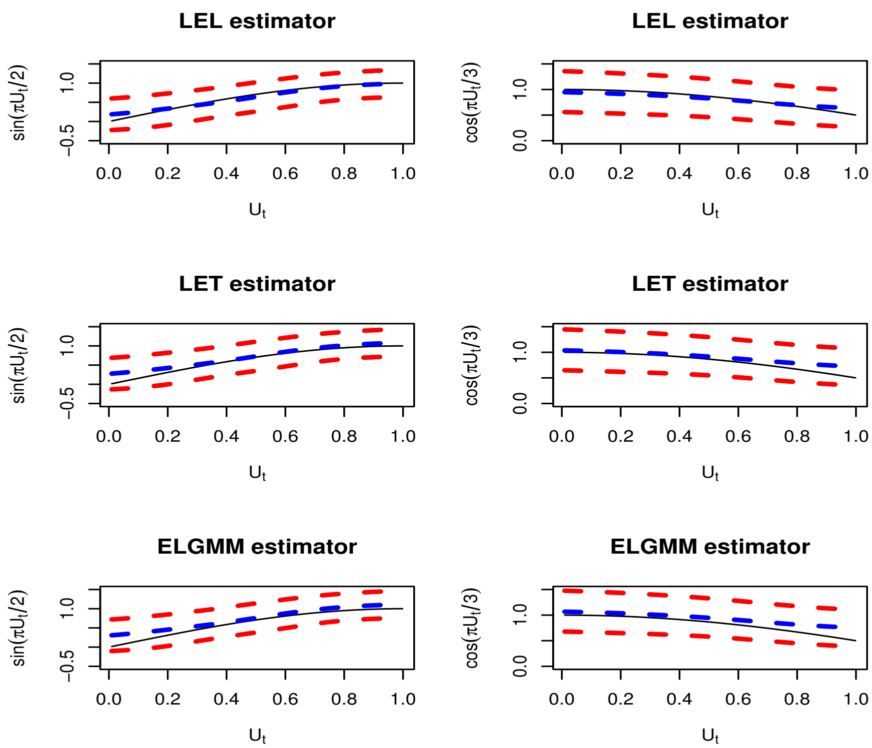

Figure 1: Estimated $[0.25, 0.50, 0.75]$ varying coefficients quantiles with $\dim(V_t) = 3$. The solid line is the true varying coefficient.

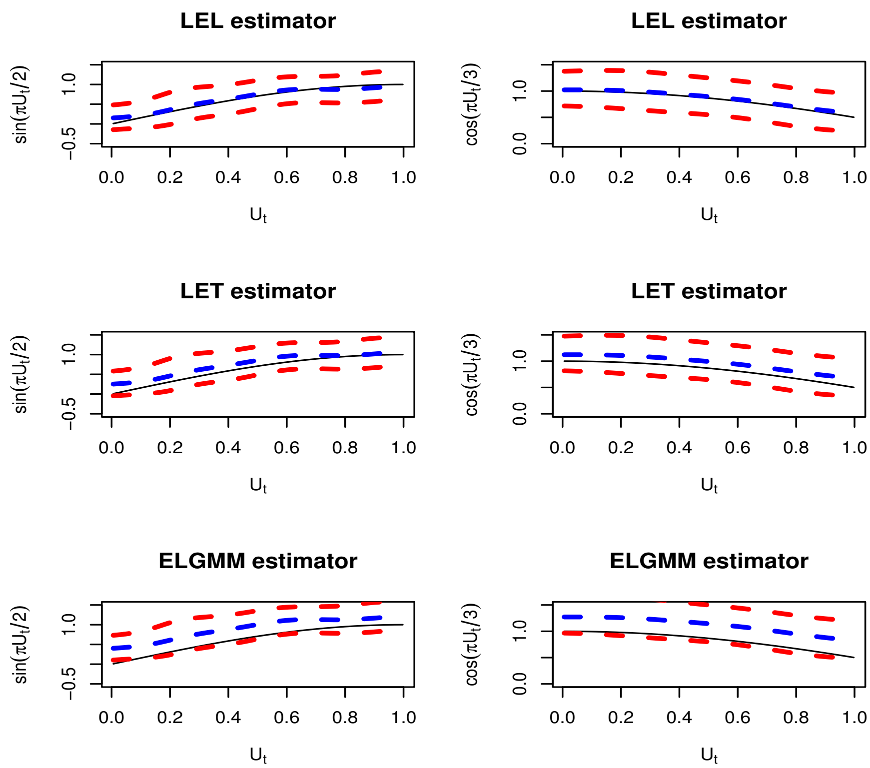

Figure 2: Estimated $[0.25, 0.50, 075]$ varying coefficients quantiles with $\dim(V_t) = 7$. The solid line is the true varying coefficient.

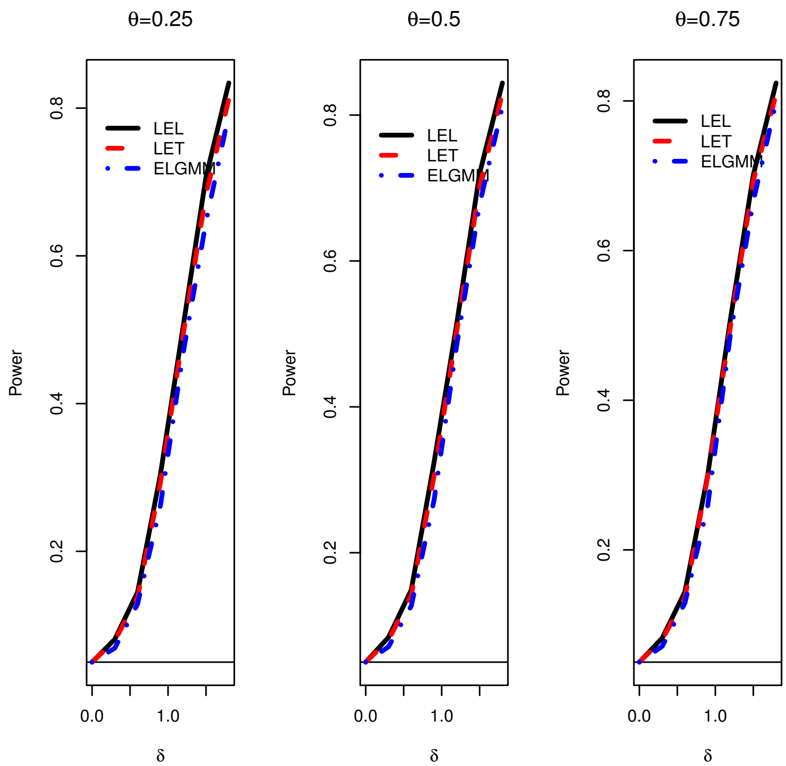

Figure 3: Finite sample power of (10) with $\dim(V_t) = 3$.

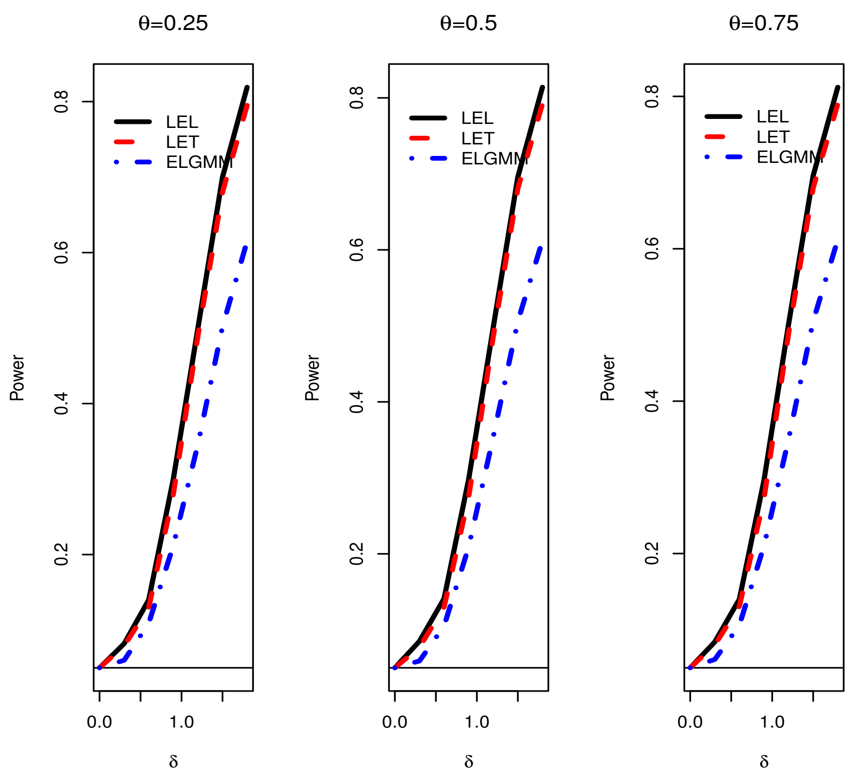

Figure 4: Finite sample power of (10) with $\dim(V_t) = 7$.

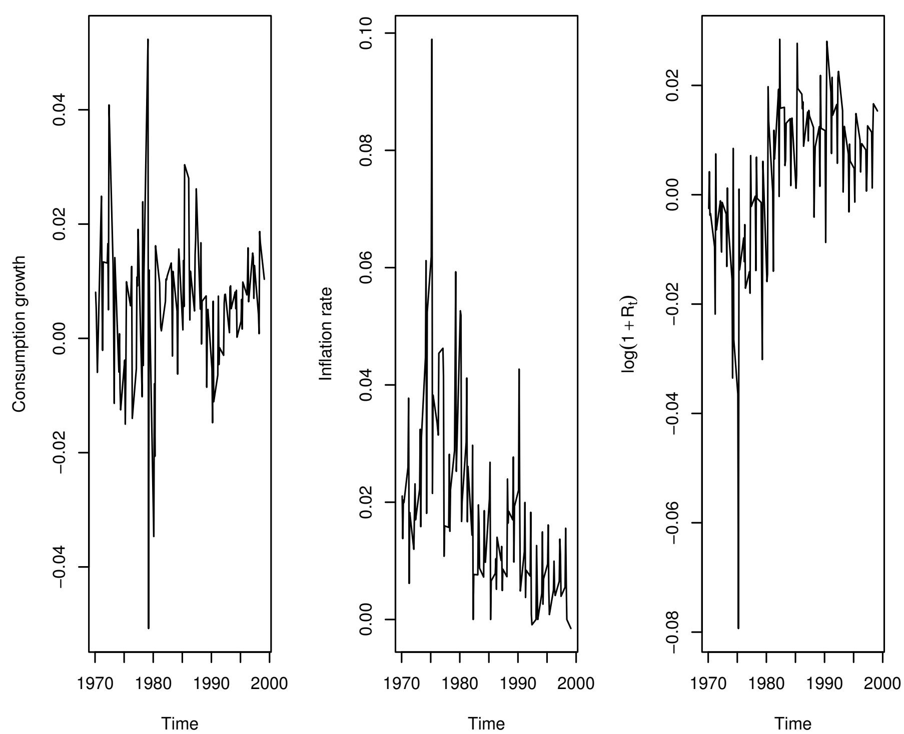

Figure 5: Plots of consumption growth, inflation and $log\,(1 + R_t)$

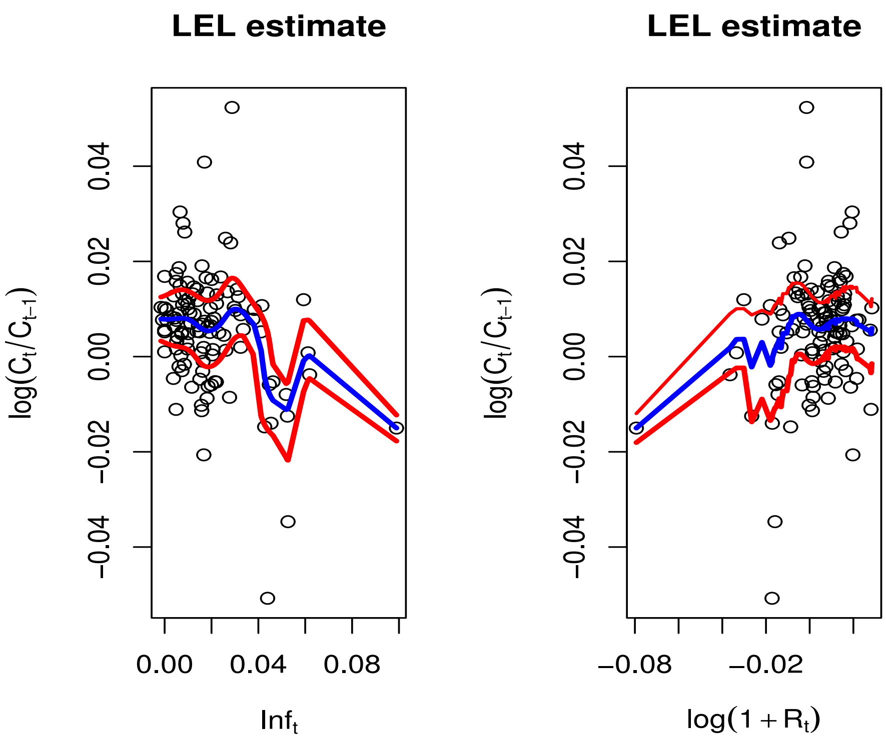

Figure 6: Estimated LEL $[0.25, 0.50, 0.75]$ varying coefficient quantiles.

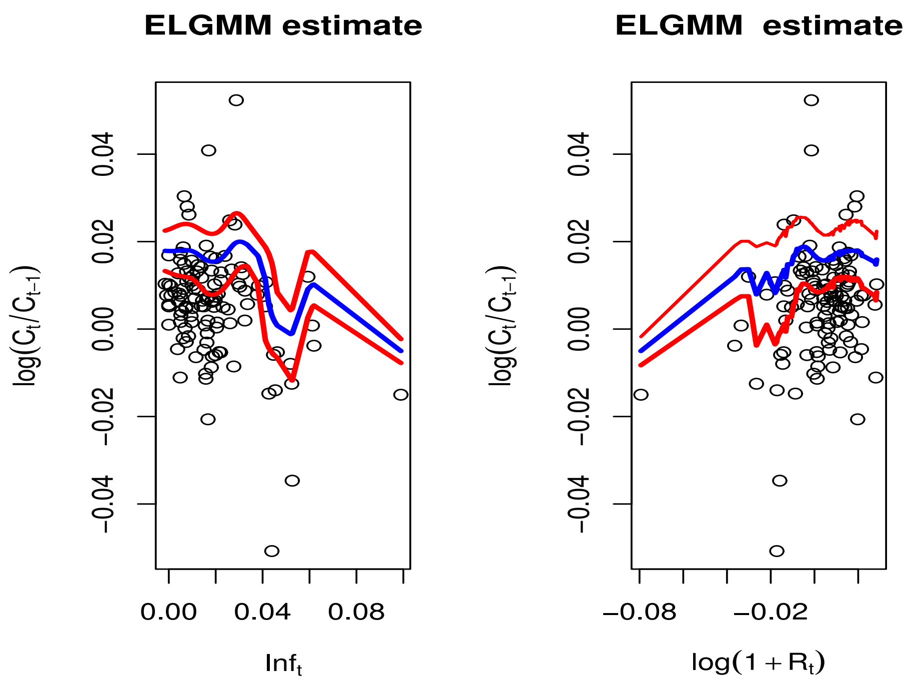

Figure 7: Estimated ELGMM $[0.25, 0.50, 0.75]$ varying coefficient quantiles.