Article:

Bravo, Francesco orcid.org/0000-0002-8034-334X (2022) Second order expansions of estimators in nonparametric moment conditions models with weakly dependent data. Econometric Reviews. pp. 583-606. ISSN 0747-4938

https://doi.org/10.1080/07474938.2021.1991140

Second order expansions of estimators in nonparametric moment conditions models with weakly dependent data

Francesco Bravo

ECONOMETRIC REVIEWS
https://doi.org/10.1080/07474938.2021.1991140

Second order expansions of estimators in nonparametric moment conditions models with weakly dependent data

Francesco Bravo

University of York, York, UK

ABSTRACT

This paper considers estimation of nonparametric moment conditions models with weakly dependent data. The estimator is based on a local linear version of the generalized empirical likelihood approach, and is an alternative to the popular local linear generalized method of moment estimator. The paper derives uniform convergence rates and pointwise asymptotic normality of the resulting local linear generalized empirical likelihood estimator. The paper also develops second order stochastic expansions (under a standard undersmoothing condition) that explain the better finite sample performance of the local linear generalized empirical likelihood estimator compared to that of the efficient local linear generalized method of moments estimator, and can be used to obtain (second order) bias corrected estimators. Monte Carlo simulations and an empirical application illustrate the competitive finite sample properties and the usefulness of the proposed estimators and second order bias corrections.

KEYWORDS

α-mixing; local linear estimation; stochastic discount factor; stochastic expansion

JEL CLASSIFICATION

C14; C32

1. Introduction

Moment conditions models arise naturally in economics, finance and statistics, often in a conditional form. For example, many dynamic stochastic general equilibrium models used in macroeconomics, many assets pricing models under the no arbitrage condition used in finance and the generalized estimating equations (GEE) models for longitudinal data used in statistics, all give rise to a set of (possibly conditional) moment conditions. Estimation of the unknown parameters in such models is typically carried out using Hansen's (1982) generalized method of moments (GMM) approach, see also Qu, Lindsay, and Li (2000) for GEE models, or, alternatively, Newey and Smith's (2004) generalized empirical likelihood (GEL) approach. When the unknown parameters are finite dimensional, the asymptotic properties of the asymptotically equivalent efficient GMM and GEL estimators are well established. In particular, Newey and Smith (2004) showed that the empirical likelihood estimator has the smallest second order bias among the GEL and the efficient GMM estimators, a property which is important given that GMM estimators are typically characterized by poor finite sample properties, see for example, Hansen, Heaton, and Yaron (1996).

In this paper, we consider nonparametric moment conditions models, that is moment conditions models where the unknown parameters are infinite dimensional. These models are a natural extension of the parametric ones and can be applied in a variety of situations where a parametric

CONTACT Francesco Bravo francesco.bravo@york.ac.uk Department of Economics, University of York, York YO10 5DD, UK.

Supplemental data for this article is available online at https://doi.org/10.1080/07474938.2021.1991140

specification might be reductive and/or questionable. For example, the popular stochastic discount factor model used in the asset pricing literature (see Cochrane (2001) for a thorough review, and Example 2 and Section 4 for more details) relies on a parametric specification of a given utility function. Similarly, the quasi-likelihood approach commonly used in the statistics literature to estimate generalized linear models (see for example McCullagh and Nelder (1989) and Liang and Zeger (1986) for GEE models) rely on a parametric specification of the link function. The model we consider is quite general, since it allows for the number of unknown parameters to be less than the number of moment conditions (that is the model is overidentified) and for weakly dependent observations, which is particularly useful in macroeconomics and finance, since macroeconomic and financial data typically exhibit some form of serial dependence. For example, the nonparametric quasi-likelihood model of Severini and Staniswalis (1994), the nonparametric estimating equations model of Cai (2003), the nonparametric moment conditions model of Lewbel (2007), the nonparametric dynamic panel data model of Cai and Li (2008) (see also Bravo (2016)) and the nonparametric stochastic discount factor model of Fang, Ren, and Yuan (2011) and of Cai, Ren, and Sun (2015) can all be considered as special cases of the model considered here.

This paper contributes to the literature on estimation of nonparametric moment conditions models by considering GEL estimation of the unknown infinite dimensional parameters. The estimator is based on the local linear method of Fan and Gijbels (1996), which, when compared to the Nadaraya–Watson kernel (local constant) method, is characterized by good statistical properties including smaller bias, efficiency in a minimax sense and design adaptivity. We call the resulting estimator local linear GEL (LLGEL henceforth). We make a number of contributions:

First, we establish uniform convergence rates and the pointwise asymptotic normality of the LLGEL estimator. As far as we are aware of, this is the first paper that obtains strong uniform convergence rates for LLGEL estimators of nonparametric moment conditions models. For the pointwise asymptotic normality, we show that the LLGEL estimator has the same asymptotic bias as that of the local linear GMM estimator (LLGMM henceforth) of Fang et al. (2011) and Cai, Ren, and Sun (2015), but with an asymptotic covariance that is smaller than that of the LLGMM estimator unless one uses a two-step estimation procedure, which results in the so-called efficient LLGMM estimator, but does require choosing the bandwidth twice – see Remark 1 in Section 4. We use a simulation study (see the motivating Example 4 in Section 2) to show that one of the proposed LLGEL estimators (namely the local linear empirical likelihood (LLEL henceforth) defined in (8)) is characterized by a smaller mean squared error (MSE henceforth) compared to that of the efficient LLGMM estimator across a range of bandwidths, and, more importantly, that the MSE seems to be unaffected by the number of instruments used in the estimation, as opposed to that of the efficient LLGMM estimator, which seems to be growing as the number of instruments grows. To investigate this issue further, we consider the second order asymptotic properties of both the LLGEL and LLGMM estimators under a standard undersmoothing condition. Undersmoothing is often used in nonparametric estimation and inference (see for example Chen (1996), Lewbel (2007), Lewbel (2007), Fang et al. (2011), Chen and Qin (2000) and Otsu, Xu, and Matsushita (2015) among others); it is practically useful, as it removes the need to estimate the asymptotic bias resulting from the local estimation, and theoretically interesting as it produces confidence intervals (regions) with more accurate coverage (see Hall (1992) for a theoretical justification of the merit of undersmoothing over direct bias estimation).

Second, we develop second order stochastic expansions, that as far as we are aware of, are new even with i.i.d. data. It should be noted that Gospodinov and Otsu (2012) also considered second order expansions in the context of moment conditions models with weakly dependent data. However, their analysis focuses on the conditional aspect of the moment conditions and, as such, proposes a local version of the GMM estimator that operates directly on the moment conditions and can be used (among other things) to obtain a second order expansion of the estimator of the unknown parameter of an autoregressive model. This paper focuses on the estimation of unknown infinite dimensional parameters in moment conditions models, rather than on the possibly conditional aspect of the moment conditions themselves and, as such, is very different in scope from that of Gospodinov and Otsu (2012). There is however a theoretically interesting overlap between the two papers, in the sense that the results of this paper could be used to obtain a second order expansion of (the local linear version of) the same local GMM estimator considered by Gospodinov and Otsu (2012) but for a functional version of their autoregressive model, in which the unknown functional parameter depends on the same instrument that is used as the conditioning variable. The stochastic expansions we obtain are useful both theoretically and practically, as they explain why in the motivating Example 4 the LLEL estimator is characterized by a better MSE compared to that of the asymptotically equivalent efficient LLGMM. In particular, they explain why the MSE of the LLEL estimator is unaffected by the number of instruments used in the estimation. This result can be generalized to the important case where the moment conditions are based on instrumental variables, such as those presented in the examples below. In this case, the second order bias – that is the expectation of the components in the vector obtained in the second order stochastic expansion of the local linear estimator with undersmoothing, see (17) for a precise definition – of the LLEL estimator is always bounded and smaller than the corresponding second order bias of the LLGMM estimator (see Corollary 1 for more details). For the general case, the expansions show that the efficient LLGMM estimator has an additional component that is going to be positive if the correlation between the derivative of the moment indicator and the moment indicator is positive, which is often the case with moment conditions models. From a practical point of view, the expansions can be used to construct second order bias corrected LLGEL and LLGMM estimators that have a reduced MSE compared to that of the original ones for a reasonable range of bandwidths.

Finally, we show the usefulness and applicability of the proposed estimators and second order bias corrections with a Monte Carlo study and an empirical application. The Monte Carlo evidence we provide is encouraging and suggests that while LLGEL estimators are typically characterized by a smaller MSE than the efficient LLGMM estimator, their second order bias corrected analogs have smaller MSEs, which confirms the usefulness of the proposed bias correction. In the empirical application we estimate a nonparametric specification of the stochastic discount factor version of Fama and French's (1993) three factors model. The application shows the usefulness of the nonparametric approach used in the paper, as it clearly identifies (conditional) nonlinearities in the risk prices associated with the three factors.

The rest of the paper is structured as follows: in the next section we introduce the model, provide four illustrative examples and describe the estimation methods. Section 3 presents the main results. Sections 4 and 5 contain, respectively, the results of the Monte Carlo simulations and the empirical application. Section 6 contains some concluding remarks. An online Supplemental Appendix contains additional Monte Carlo results and all the proofs.

The following notation is used throughout the paper: "τ" indicates transpose, "$'$", "$''$", etc. indicate, respectively, first, second, etc. derivative of a function with respect to its unique argument, "$\otimes$" is the Kronecker product, "$\|\cdot\|_\infty$" is the sup-norm, "tr" is the trace operator, $\underline{0}$ and O denote, respectively a vector and a matrix of zeros, and for any vector v, $v^{\otimes 2} = vv^\tau$.

2. The model and the estimators

Let $\{Z_t^\tau, U_t\}_{t\in\mathbb{Z}}^\tau$ denote a strictly stationary sequence of random vectors taking values in $\mathcal{Z} \subset \mathbb{R}^{d_Z}$ and $\mathcal{U} \subset \mathbb{R}$, and let $h \in \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \cdots \times \mathcal{H}_k$ denote a k dimensional vector of unknown functions, where $\mathcal{H}$ is a pseudo-metric space of functions. The model we consider is

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

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 specification of (1) is fairly general and can accommodate many models used in empirical research as the following three general examples illustrate.

Example 1. (*Instrumental variables estimation of a generalized nonparametric regression model*) Let

$$Y_t = f(X_t, h_0(U_t)) + \varepsilon_t, \tag{2}$$

where $f : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ is a known function, $X_t^\tau = [X_{1t}^\tau, X_{2t}^\tau]^\tau$ is an $\mathbb{R}^k$ $(k = k_1 + k_2)$ valued vector of covariates with X_{1t} possibly endogenous, and the unobservable error ε_t is such that $E(\varepsilon_t | X_{2t}, U_t) = 0$ *a.s.* Then, the law of iterated expectations implies that, for any vector of functions $q : \mathcal{X}_2 \to \mathbb{R}^l$,

$$E\left[q(X_{2t})\left(Y_t - f(X_t, h(U_t))\right)|U_t\right] = \underline{0} \ \ a.s. \ \text{for a unique} \ h = h_0,$$

which is of the same form as that of (1) with $Z_t = [Y_t, X_t^\tau]^\tau$ and

$$m(Z_t, h(U_t)) = q(X_{2t})\left(Y_t - f(X_t, h(U_t))\right).$$

Example 2. (*Nonparametric stochastic discount factor model*). Let $R_{j,t}$ $(j = 1, \ldots, J)$ denote the (excess) returns of J risky assets and $R_{M,t}$ denote the (excess) market return. Following Wang (2002, 2003) (see also footnote 2 in Section 4), let $h_{t+1} = 1 - h_0(U_t)R_{M,t+1}$ denote a nonparametric pricing kernel that satisfies $E[h_{t+1} \otimes R_{t+1} | X_t, U_t] = 0$ *a.s.*, where $R_t = [R_{1,t}, \ldots, R_{J,t}]^\tau$ and X_t denote a set of additional conditioning variables. The law of iterated expectations implies that, for any vector of functions $q : \mathcal{X} \to \mathbb{R}^l$

$$E\left[q(X_t)(h_{t+1} \otimes R_{t+1})|U_t\right] = \underline{0} \ \ a.s. \ \text{for a unique} \ h = h_0, \tag{3}$$

which is of the same form as that of (1) with $Z_t = [R_t^\tau, X_t^\tau]^\tau$ and

$$m(Z_t, h(U_t)) = q(X_t)h_{t+1} \otimes R_{t+1}.$$

Example 3. (*Dynamic panel data model - small N and large T*). Let

$$Y_{it} = X_{it}^\tau h_0(U_{it}) + \varepsilon_{it} \quad i = 1, \ldots, N,$$

where X_{it} is an $\mathbb{R}^k$ valued vector of covariates that may contain lagged values of Y_{it} and the unobservable errors ε_{it} are such that $E(\varepsilon_{it}|X_{is}, U_{it}) \neq 0$ for $s \leq t$, but there exists an $\mathbb{R}^l$ valued vector of covariates W_{it} such that $E(\varepsilon_{it}|W_{is}, U_{it}) = 0$ for $s \leq t$. Then, the law of iterated expectations implies that

$$E\left[W_{it}\left(Y_{it} - X_{it}^\tau h(U_{it})\right)|U_{it}\right] = \underline{0} \ \ a.s. \ \text{for a unique} \ h = h_0, \tag{4}$$

which is of the same form as that of (1) with $Z_t = [W_{1t}^\tau, \ldots, W_{Nt}^\tau, Y_{1t}, \ldots, Y_{Nt}, X_{1t}^\tau, \ldots, X_{Nt}^\tau]^\tau$, $U_t = [U_{1t, \ldots,} U_{Nt}]^\tau$ and

$$m(Z_t, h(U_t)) = W_{it}\left(Y_{it} - X_{it}^\tau h(U_{it})\right).$$

Interestingly, the moment indicator $W_{it}(Y_{it} - X_{it}^\tau h(U_{it}))$ can be used to obtain a first-step LLGEL estimator analog to the LLGMM estimator considered by Cai, Chen, and Fang (2015) in their two-step estimation procedure for dynamic partially linear varying coefficient model, although it is important to note that their model is based on large N and small T, whereas we consider small N and large T.

Throughout the rest of the paper, we assume that, at a given point $U_t = u$, h_0 can be linearly approximated by

$$h_0(U_t) = h_0(u) + h_0'(u)(U_t - u) := h_1 + h_2(U_t - u); \tag{5}$$

thus, for $U_t \approx u$, (1) can be written as

$$E\big[m(Z_t, h_1 + h_2(U_t - u))|U_t = u\big] \approx \underline{0}. \tag{6}$$

It is important to note that unless the dimension $\dim(m)$ of the moment indicator $m\ (\cdot)$ is such that $\dim(m) \geq \dim(h_1) + \dim(h_2)$, $h_j\ (j = 1, 2)$ cannot be consistently estimated using the kernel based sample analog of (6). Therefore we consider the augmented moment indicator

$$g_t(h_1, h_2) = \begin{bmatrix} 1 \\ \dfrac{U_t - u}{b} \end{bmatrix} \otimes m(Z_t, h_1 + h_2(U_t - u))$$

and base the estimation on the localized augmented sample moment indicator

$$\frac{1}{Tb}\sum_{t=1}^{T} g_t(h_1, h_2)K\left(\frac{U_t - u}{b}\right),$$

where $K : \mathcal{U} \to \mathbb{R}$ is a kernel function and $b =: b(T)$ is the bandwidth. To define the LLGEL estimator, let

$$\rho\left(\lambda(u)^\tau g_t(h_1, h_2)K\left(\frac{U_t - u}{b}\right)\right) := \rho(v_t(u, \lambda, h_1, h_2))$$

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

$$\sum_{t=1}^{T} \pi_t(u, \lambda, h_1, h_2)g_t(h_1, h_2)K\left(\frac{U_t - u}{b}\right) = \underline{0},$$

with

$$\pi_t(u, \lambda, h_1, h_2) = \frac{\partial\rho(v_t(u, \lambda, h_1, h_2))/\partial v_t}{\sum_{t=1}^{T}\partial\rho(v_t(u, \lambda, h_1, h_2))/\partial v_t}$$

playing the role of the "implied probabilities" associates with the localized constraint (6). The LLGEL estimator is then defined as

$$\hat{h}_1^\rho, \hat{h}_2^\rho = \mathrm{argmin}_{h_1, h_2 \in \mathcal{H}_C} \frac{1}{Tb}\sum_{t=1}^{T} \rho\big(v_t\big(u, \hat{\lambda}, h_1, h_2\big)\big), \tag{7}$$

where $\mathcal{H}_C$ is defined in Assumption A2(iii),

$$\hat{\lambda} = \mathrm{argmax}_{\lambda \in \Lambda_T(h_1, h_2)} \frac{1}{Tb}\sum_{t=1}^{T} \rho\big(v_t\big(u, \lambda, \bar{h}_1, \bar{h}_2\big)\big),$$

$\Lambda_T(h_1, h_2) = \big\{\lambda(u) : \lambda(u)^\tau g_t(h_1, h_2)K((U_t - u)/b) \in \Lambda_0, t = 1, \ldots, T\big\}$ and $\bar{h}_j$ are fixed values of h_j $(j = 1, 2)$. Examples of (7) include the local linear version of empirical likelihood (LLEL)

$$\hat{h}_1^{el}, \hat{h}_2^{el} = \mathrm{argmin}_{h_1, h_2 \in \mathcal{H}_C} \frac{1}{Tb}\sum_{t=1}^{T} \log\big(1 - v_t\big(u, \hat{\lambda}, h_1, h_2\big)\big) \tag{8}$$

and the local linear version of exponential tilting (LLET)

$$\hat{h}_1^{et}, \hat{h}_2^{et} = \mathrm{argmin}_{h_1, h_2 \in \mathcal{H}_C} -\frac{1}{Tb}\sum_{t=1}^{T} \exp\big(v_t\big(u, \hat{\lambda}, h_1, h_2\big)\big). \tag{9}$$

For completeness, we define the efficient LLGMM estimator as the solution of the minimization problem

Table 1. MSE for the LLGMM $\hat{h}_1$ and LLEL $\hat{h}_1^{el}$ estimators.

		$\hat{h}_1$			$\hat{h}_1^{el}$		
$\dim(q(X_{2t}))$	T	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$[a]	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$[a]
3	250	0.0039	0.0041	0.0045	0.0030	0.0032	0.0035
	1000	0.0012	0.0014	0.0016	0.0007	0.0009	0.0010
9	250	0.0231	0.0245	0.0250	0.0031	0.0033	0.0036
	1000	0.0075	0.0078	0.0080	0.0008	0.0011	0.0011

[a] $\hat{b}_1 = \hat{b}T^{-1/4},\ \hat{b}_2 = \hat{b}T^{-1/3},\ \hat{b}_3 = \hat{b}T^{-2/5}$.

$$\hat{h}_1, \hat{h}_2 = \mathrm{argmin}_{h_1, h_2 \in \mathcal{H}_C} \frac{1}{Tb}\sum_{t=1}^{T} g_t(h_1, h_2)^{\tau} K\left(\frac{U_t - u}{b}\right)\hat{\Omega}(u)^{-1}\frac{1}{Tb}\sum_{t=1}^{T} g_t(h_1, h_2) K\left(\frac{U_t - u}{b}\right), \qquad (10)$$

where $\hat{\Omega}(u)^{-1}$ is a weight matrix based on a preliminary LLGMM estimator, see for example (14).

We conclude this section with an additional example that illustrates both theoretically and numerically the importance of the stochastic expansions developed in Section 3.3.

Example 4. (*Motivating example*). Let

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

where $X_t^{\tau} = [1, X_{1t}]$ and $E(\varepsilon_t | X_{2t}, U_t) = 0$ *a.s.*, which is a simplified version of the model (2), and consider both LLEL and efficient LLGMM estimation of h_0 based on the localized moment conditions

$$E\left[q(X_{2t})\left(Y_t - X_t^{\tau}(h_1 + h_2(U_t - u))\right) | U_t = u\right] \approx \underline{0}$$

for a known vector of functions $q : \mathcal{X}_2 \to \mathbb{R}^l$ $(l \geq 3)$. Table 1 reports the average MSE of the LLEL and efficient LLGMM estimators $\hat{h}_1^{\circ}$,

$$MSE\left(\hat{h}_1^{\circ}\right) = \frac{1}{2T}\sum_{j=1}^{2}\sum_{t=1}^{T}\left(\hat{h}_{1j}^{\circ}(U_t) - h_{j0}(U_t)\right)^2,$$

where $\hat{h}_{1j}^{\circ}$ is either $\hat{h}_{1j}$, or $\hat{h}_{1j}^{el}$, using the Epanechnikov kernel function, that is $K(u) = (3/4)(1 - u^2)$ for $|u| \leq 1$, for three different undersmoothed bandwidths chosen using the same *ad hoc* bandwidth selection procedure described in Section 4 and $h_0(U_t) = [\sin(\pi U_t/2), \sin(6\pi U)]^{\tau}$, $U_t \sim U(0,1)$, $X_{1t} = 0.8X_{2t} + \eta_t$, $X_{2t} = 0.3X_{2t-1} + \zeta_t$ with $\zeta_t \sim N(0,1)$ independent of

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

We consider two sets of instruments, either $q(X_{2t}) = \left[1, X_{2t}, X_{2t}^3\right]^{\tau}$ or $\left[1, X_{2t}, \ldots, X_{2t}^8\right]^{\tau}$ so that the degrees of overidentification are, respectively, 1 and 7, and two different sample sizes, $T = 250$ and $T = 1000$.

The results of Table 1 are based on 1000 replications and can be summarized as follows: when $\dim(q(X_{2t})) = 3$, the MSEs of the LLEL and efficient LLGMM estimators are broadly comparable, however when $q(X_{2t}) = 9$ the MSE of the LLEL estimator is still comparable to that of the case where $\dim(q(X_{2t})) = 3$, whereas the MSE of the efficient LLGMM estimator is significantly higher. The stochastic expansions derived in Section 3.3 clearly explain this difference, because some calculations show that under the specification of the Monte Carlo design, the second order bias of the LLGEL estimator is given by $\Sigma(u)_{11}^{-1} E\ (X_t\varepsilon_t | U_t = u) f(u)/Tb$, where

$$\Sigma(u)_{11} = \left(E\left(X_t q(X_{2t})^\tau | U_t = u\right) \left(E\left(q(X_{2t})^{\otimes 2} \varepsilon_t^2 | U_t = u\right) \int K^2(v) dv \right)^{-1} E\left(q(X_{2t}) X_t^\tau | U_t = u\right) \right) f(u),$$

with the matrix $\Sigma(u)_{11}$ defined as the upper left block of the block diagonal matrix $\Sigma(u)$ defined in (11), whereas that of the efficient LLGMM estimator is given by $(l-k-1)\Sigma(u)_{11}^{-1} E(X_t \varepsilon_t | U_t = u) f(u)/Tb$, which explains the results of Table 1, in the sense that it shows that it is increasing in the degrees of overidentification – see the discussion after Theorem 3 for more details.

3. Asymptotic results

To simplify the notation, let $m(Z_t, \cdot) := m_t(\cdot)$, and define

$$\begin{aligned} G(u) &= f(u) diag\left[1, \int v^2 K(v) dv\right] \otimes E\left[\frac{\partial m_t(h_0)}{\partial h^\tau} \middle| U_t = u \right], \\ \Omega(u) &= f(u) diag\left[\int K^2(v) dv, \int v^2 K^2(v) dv \right] \otimes E\left[m_t(h_0)^{\otimes 2} | U_t = u \right], \\ \Sigma(u) &= G(u)^\tau \Omega(u)^{-1} G(u). \end{aligned} \tag{11}$$

3.1. Uniform convergence rates

Assume that:

A1 The sequence $\{Z_t^\tau, U_t\}_{t \in \mathbb{Z}}^\tau$ is strictly stationary α mixing, with mixing coefficient satisfying $\sum_{t=1}^{\infty} t^\beta \alpha(t)^{\frac{\eta-2}{\eta}} < \infty$ for some $\eta > 2$ and $\beta > 1 - 2/\eta$,
A2 (i) There exists a unique h_0 such that $E\left[m_t(h_0) | U_t = u\right] = \underline{0}$ for all $u \in \mathcal{U}$, (ii) for any $\xi > 0$, there exists a $C(\xi) > 0$, such that

$$\inf_{\substack{h_1, h_2 \in \mathcal{H}_C \\ \|(h_1 \bar{h}_0)^\tau, (h_2 \bar{h}_0')^\tau\|_\infty \geq \xi}} \left[\left\| E\left(g_t(h_1, h_2) K\left(\frac{(U_t - u)}{b} \right) \right) \right\| \right] > C(\xi) \tag{12}$$

for some $u \in \mathcal{U}$, (iii) the parameter space $\mathcal{H}_C \subset \mathbb{R}^{2l}$ and the support $\mathcal{U}$ of u are compact sets, (iv) h_0 is twice continuously differentiable on $\mathcal{U}$,

A3 (i) $\partial m_t(h)/\partial h^\tau$ exists and is continuous for each $h \in \mathcal{H}_C$ *a.s.*, (ii) there exist functions $M_j(Z_t)$ $(j = 1, 2, 3)$ such that $\sup_{h_1, h_2 \in \mathcal{H}_C} \|g_t(h_1, h_2)\| \leq M_1(Z_t)$, $\sup_{h_1, h_2 \in \mathcal{H}_{0C}} \|g_t(h_1, h_2)^{\otimes 2}\| \leq M_2(Z_t)$ and $\sup_{h_1, h_2 \in \mathcal{H}_{0C}} \|\partial g_t(h_1, h_2)/\partial(h_1^\tau, h_2^\tau)^\tau\| \leq M_3(Z_t)$, where $\mathcal{H}_{0C}$ is an open neighborhood of h_0 and h_0', such that $E(M_j(Z_t)^\eta | U = u_t) f(u) < \infty$ uniformly in $u \in \mathcal{U}$, where η is defined in A1,
A4 (i) for $t \geq 2$

$$E\left[\sup_{h_1, h_2 \in \mathcal{H}_C} \left(\|m_1(h_1, h_2)\|^2 + \|m_t(h_1, h_2)\|^2 \right) | U_1 = u, U_t = v \right] < \infty,$$

uniformly in $u, v \in \mathcal{U}$, (ii) the matrices $G(u)$ and $\Omega(u)$ are Lipschitz continuous in $u \in \mathcal{U}$ and are, respectively, of rank $2k$ and positive definite uniformly in $u \in \mathcal{U}$,

A5 $\rho(v_t)$ is twice continuously differentiable in v_t in a neighborhood of 0, with $\rho_j = -1$ $(j = 1, 2)$ and $\rho_j = \partial^j \gamma(v_t)/\partial v_t^j|_{v_t=0}$,

A6 (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 Lipschitz continuous and strictly positive at $U_t = u$, (iii) the joint density $f_{1,t}$ of U_1 and U_t for $t \geq 2$ is Lipschitz continuous at $u \in \mathcal{U}$.

Assumption A1 excludes deterministic and stochastic trends and specifies the dependence structure of the sequence $\{Z_t^\tau, U_t\}_{t\in\mathbb{Z}}^\tau$ as α mixing with size $O(t^{-2(\eta-2)/2-\epsilon})$ for some $\epsilon > 0$ (see Doukhan (1994) for examples and properties of α mixing processes). Assumption A2(i) is a standard identification condition that can be often verified by imposing more primitive conditions on m and/or some of the components of the random vector Z_t. For example, in the generalized nonparametric regression model (2), if $f(X_t, h_0(U_t)) = f(X_t^\tau h_0(U_t))$ and f is a monotonic function, then A2(i) is implied by the more primitive assumption $\Pr(X_t^\tau(h(U_t) - h_0(U_t)) \neq 0|U_t = u) > 0$ for all $h \neq h_0 \in \mathcal{H}_C$, which in turn holds if $E(X_t^{\otimes 2}|U_t = u)$ is positive definite uniformly in $u \in \mathcal{U}$. Similarly, for the dynamic panel data model (4) A2(i) is implied by the condition $rank(E(W_{it}X_{it}^\tau|U_{it} = u)) = k$ uniformly in $u \in \mathcal{U}$. Assumption A2(ii) is necessary because of the local linear nature of the proposed estimation method; it ensures that the augmented moment indicator $g_t(h_1, h_2)$ has two separated minima h_0 and h_0'. Note however that A2(ii) would not be required if the moment indicator is linear in the unknown infinite dimensional parameter, in which case a rank condition of the derivative of the moment indicator is sufficient. The compactness of the parameter space $\mathcal{H}_C$ in the first part of Assumption A2(iii) is as in Lewbel (2007), whereas the compactness of the support $\mathcal{U}$ in the second part of Assumption A2(iii) is often assumed to obtain uniform convergence rates for dependent data, see for example Liebscher (1996) and Masry (1996), but could be relaxed as in Hansen (2008). The conditional moments assumptions in A3(ii) are implied by the existence of the corresponding unconditional moments and are the weakest possible to obtain uniform consistency results. Assumption 4(i) is standard in nonparametric estimation with dependent data, see for example Cai, Fan, and Yao (2000); Assumption A4(ii) is crucial to obtain the uniform convergence rates of Theorem 1. Assumptions 5 and 6 are standard, respectively, in the GEL literature, see Newey and Smith (2004) and in the nonparametric estimation literature, see for example Cai (2003). Finally, we note that the uniformity in $u \in \mathcal{U}$ in Assumptions A3(ii) and A4 can be weakened to the requirement that they hold in an open neighborhood of $u \in \mathcal{U}$ if only pointwise results (such as the asymptotic normality in Section 3.2 and the stochastic expansions in Section 3.3) are required.

Theorem 1. *Under A1–A6, for $b \to 0$, $Tb/\log(T) \to \infty$ as $T \to \infty$ and with the additional summability condition $\sum_{T=1}^{\infty} \varphi(T) < \infty$ for the α mixing coefficient $\alpha(t)$ with $\varphi(T)$ defined in Proposition 1 (in the Supplemental Appendix),*

$$\left\| \begin{matrix} \hat{h}_1^\rho(u) - h_0(u) \\ b\left(\hat{h}_2^\rho(u) - h_0'(u)\right) \end{matrix} \right\|_\infty = O_{a.s.}\left(\left(\frac{\log T}{Tb}\right)^{1/2} + b^2\right). \tag{13}$$

The uniform convergence rates (13) are standard in the nonparametric estimation literature (see for example Liebscher (1996) and Masry (1996)), but are new in the context of nonparametric moment conditions models (note also for i.i.d. data) and are sharper than those obtained for example by Carroll et al. (1997) for the nonparametric component of their generalized partial linear single index model.

3.2. Asymptotic normality

Assume further that:

A7 (i) $h_0, h_0' \in int(\mathcal{H}_C)$, (ii) the matrix $\Sigma(u)$ is nonsingular in a open neighborhood of $u \in \mathcal{U}$.

The following theorem establishes the asymptotic distribution of the LLGEL estimator

Theorem 2. *Under A1–A7 as* $(Tb)^{1/2} \to \infty$

$$(Tb)^{1/2}\begin{bmatrix} \hat{h}_1^{\rho}(u) - h_0(u) - B(u)/2 \\ b\left(\hat{h}_2^{\rho}(u) - h_0'(u)\right) \end{bmatrix} \xrightarrow{d} N\left(\underline{0}, \Sigma^{-1}(u)\right),$$

where $B(u) = b^2 h_0''(u)\int v^2 K(v)dv$.

Theorem 2 shows that the order of the asymptotic bias of the LLGEL estimator is $O(b^2)$ as in standard nonparametric estimation under the smoothness Assumptions A2(iv) and A6(ii). The asymptotic covariance of the LLGEL estimator is the same as that of the efficient LLGMM estimator, which is well known to be the smallest (in the matrix sense) covariance possible for LLGMM estimators based on any positive semidefinite (possibly random) weight matrix $\hat{W}$. Thus, $\Sigma(u)^{-1}$ is the "optimal" covariance matrix but it is important to say that $\Sigma(u)^{-1}$ does not represent the (semiparametric) efficiency bound. Note however that, as opposed to the LLGEL estimator, the efficient LLGMM estimator is based on the preliminary weight matrix

$$\hat{\Omega}(u)^{-1} = \frac{1}{\hat{f}(u)}\left[diag\left[\int K^2(v)dv, \int v^2 K^2(v)dv\right]\frac{1}{Tb}\sum_{t=1}^{T} m_t\left(\bar{h}_1, \bar{h}_2\right)^{\otimes 2}\omega_t(u)\right]^{-1}, \tag{14}$$

where $\omega_t(u) = L((U_t - u)/b_L)/\sum_{j=1}^{T} L((U_j - u)/b_L), \hat{f}(u) = \sum_{t=1}^{T} L((U_j - u)/b_L)/Tb_L$, the kernel function $L : \mathcal{U} \to \mathbb{R}$ can be different from the kernel function K used previously with bandwidth $b_L =: b_L(T)$, and $\bar{h}_j$ $(j = 1, 2)$ is a preliminary estimator, which requires choosing another bandwidth. Note also that Theorem 2 implies that the asymptotic integrated mean squared error $(AIMSE)$ for $\hat{h}_1$ is

$$AIMSE\left(\hat{h}_1\right) = \frac{b^4}{4}\|B\|^2 + \frac{\int K^2(v)dv}{Tb}tr\left(\Sigma_m^{-1}\right), \tag{15}$$

where

$$\Sigma_m = \int \left[E\left(\partial m_t(h_0)/\partial h^{\tau}\right)|U_t = u\right]^{\tau} E\left[m_t(h_0)^{\otimes 2}|U_t = u\right]^{-1} \\ \times E\left[\partial m_t(h_0)/\partial h^{\tau}|U_t = u\right]f(u)du,$$

which implies that the optimal bandwidth b^* minimizing (15) is

$$b^* = \left(\frac{1}{T}\right)^{\frac{1}{5}}\left(\int K^2(v)dvtr\left(\Sigma_m^{-1}\right)\|B\|^{-2}\right)^{\frac{1}{5}},$$

and the optimal convergence rate is the standard nonparametric one $T^{-4/5}$. This result is useful because it implies that data driven methods, such as least squares cross-validation (see for example Li and Racine (2004) for some optimality properties of such procedure) could be used to automatically select b_*. Pointwise $(1-\alpha)\%$ confidence intervals for h_{j0} $(j = 1, \ldots, k)$ at a given point u can be constructed as

$$\hat{h}_{j1}^{\circ}(u) - \hat{B}_j(u)/2 \pm z_{1-\alpha/2}\frac{\hat{\Sigma}(u)_{11jj}^{-1}}{(T\hat{b})^{1/2}},$$

where $\hat{h}_{j1}^{\circ}$ can be either $\hat{h}_{j1}$ or $\hat{h}_{j1}^{\rho}$, $\hat{B}_j(\cdot)$ and $\hat{\Sigma}(\cdot)_{11jj}^{-1}$ are consistent estimators[1] of the jth component of the asymptotic bias $B(\cdot)$ and of the jjth component in the upper left block of the diagonal matrix $\Sigma(u)^{-1}$, $\hat{b}$ is the bandwidth selected by a data driven method and $z_{1-\alpha/2}$ is the critical value of the standard normal. Alternatively, under the additional undersmoothing condition $Tb^5 \to 0$, confidence intervals can be based on

$$\hat{h}_{j1}^{\circ}(u) \pm z_{1-\alpha/2} \frac{\hat{\Sigma}(u)_{11jj}^{-1}}{(T\hat{b})^{1/2}},$$

which, as mentioned in Section 1, is characterized by better (second order) coverage accuracy than the bias corrected one, and this is the approach we follow in the rest of the paper.

3.3. Second order stochastic expansions

To obtain second order stochastic expansions of both the efficient LLGMM and LLGEL estimators, let

$$\begin{aligned} H(u) &= \Sigma^{-1}(u) G(u)^{\tau} \Omega(u)^{-1}, \\ P(u) &= \Omega(u)^{-1}(I - G(u)H(u)), \end{aligned} \tag{16}$$

(note that both $H(u)$ and $P(u)$ are block diagonal), and assume that:

A5′ $\rho(v_t)$ is three times continuously differentiable in v_t with Lipschitz third derivative in a neighborhood of 0,

A8 (i) $\partial^2 m_t(h)/\partial h^{\tau} \partial h_j$ $(j = 1, ..., k)$ exists and is continuous for each $h \in \mathcal{H}_C$ *a.s.*, (ii) there exist functions $M_j(Z_t)$ $(j = 4, 5, 6)$ such that in an open neighborhood $\mathcal{H}_{0C}$ of h_0 and h_0' $\sup_{h_1, h_2 \in \mathcal{H}_{0C}} \|g_t(h_1, h_2)^{\otimes 2} g(h_1, h_2)\| \leq M_4(Z_t)$, $\sup_{h_1, h_2 \in \mathcal{H}_{0C}} \|(\partial g_t(h_1, h_2)/\partial(h_1^{\tau}, h_2^{\tau})^{\tau}) \times g(Z_t, h_1, h_2)\| \leq M_5(Z_t)$, $\sup_{h_1, h_2 \in \mathcal{H}_{0C}} \|\partial^2 g_t(h_1, h_2)/\partial(h_1^{\tau}, h_2^{\tau})^{\tau} \partial h_j\| \leq M_6(Z_t)$, where $E[M_j(Z_t)|U = u] < \infty$ $(j = 4, 5, 6)$ in an open neighborhood of $u \in \mathcal{U}$, (iii) A4(i) holds an open neighborhood of $u \in \mathcal{U}$, and for $t \geq 2$

$$E\left[\sup_{h_1, h_2 \in \mathcal{H}_{0C}} \left(\left\| \frac{\partial m_1(h_1, h_2)}{\partial(h_1^{\tau}, h_2^{\tau})^{\tau}} \right\| + \|m_t(h_1, h_2)\| \right) | U_1 = u, U_t = v \right] < \infty,$$

an open neighborhood of $u, v \in \mathcal{U}$, (iv) the vectors

$$E\left[\frac{\partial m_t(h_0)}{\partial h^{\tau}} m_t(h_0) | U_t = u\right], \quad E\left[\sum_{j=1}^{k} \frac{\partial^2 m_t(h_0)}{\partial h^{\tau} \partial h_j} e_j | U_t = u\right],$$
$$E\left[m_t(h_0)^{\otimes 2} m_t(h_0) | U_t = u\right],$$

are Lipschitz continuous in $u \in \mathcal{U}$, (iv) $\rho(v_t)$ is three times continuously differentiable in v_t with third order derivative Lipschitz continuous in a neighborhood of 0.

The following theorem characterizes the difference between the efficient LLGMM and the LLGEL estimators in term of the components appearing in their second order expansions, which corresponds to their second order biases. To be specific, let

[1]Consistent estimation of $\Sigma^{-1}(\cdot)$ is discussed in Section 3.3. Consistent estimation of $B(\cdot)$ can be carried out using the same sieve method suggested by Cai, Ren and Sun (2015), which involves fitting a polynomial of a sufficiently high degree, say $\sum_{i=0}^{m(T)} \alpha_i u^i$, to $f(u)$ so that an estimator of $f''(u)$ can be obtained as $\hat{f''}(u) = \sum_{i=2}^{m(T)} (i!/(i-2)!) \hat{\alpha}_i u^{i-2}$.

$$
\begin{aligned}
B_{LLGMM}(u) &= bias\left(\left[\hat{h}_1(u)^\tau, \hat{h}_2(u)^\tau\right]^\tau\right) = E\left[\frac{\psi(u)}{Tb}\right] + o\left(\frac{1}{Tb}\right), \\
B_{LLGEL}(u) &= bias\left(\left[\hat{h}_1^\rho(u)^\tau, \hat{h}_2^\rho(u)^\tau\right]^\tau\right) = E\left[\frac{\psi_\rho(u)}{Tb}\right] + o\left(\frac{1}{Tb}\right),
\end{aligned}
\tag{17}
$$

where the random vectors $\psi(u)$ and $\psi_\rho(u)$ are given in Propositions 7 and 8 (in the Supplemental Appendix), respectively.

Theorem 3. *Under the assumptions of Theorem 2, A8 and $Tb^5 \to 0$, the second order biases $B_{LLGMM}(u)$ and $B_{LLGEL}(u)$ of the efficient LLGMM and LLGEL estimators are*

$$
\begin{aligned}
B_{LLGMM}(u) &= \frac{H(u)}{Tb}\left(B_{\partial hHh}(u) - \frac{B_{G\Sigma}(u)}{2} + B_{h^3P}(u)\right) - \Sigma(u)^{-1}B_{\partial hPh}(u), \\
B_{LLGEL}(u) &= \frac{1}{Tb}\left[\left(1 + \frac{\rho_3}{2}\right)H(u)B_{h^3P}(u) + H(u)\left(B_{\partial hHh}(u) - \frac{B_{G\Sigma}(u)}{2}\right)\right],
\end{aligned}
\tag{18}
$$

where

$$
H(u)B_{\partial hHh}(u) = f(u)\begin{bmatrix} H_{11}(u)E\left(\dfrac{\partial m_t(h_0)}{\partial h^\tau}H_{11}(u)m_t(h_0)|U_t = u\right)\int K(v)^2dv \\ H_{22}(u)E\left(\dfrac{\partial m_t(h_0)}{\partial h^\tau}H_{22}(u)m_t(h_0)|U_t = u\right)\int v^2K(v)^2dv \end{bmatrix},
$$

$$
H(u)B_{G\Sigma}(u) = f(u)\begin{bmatrix} H_{11}(u)T_1(u) \\ H_{22}(u)T_2(u)\int v^2K(v)dv \end{bmatrix},
$$

$$
H(u)B_{h^3P}(u) = f(u)\begin{bmatrix} H_{11}(u)E\left[m_t(h_0)^{\otimes 2}\sum_{j=1}^{2}P_{jj}(u)m_t(h_0)|U_t = u\right]\int v^{2(j-1)}K(v)^3dv \\ \underline{0} \end{bmatrix},
$$

$$
\Sigma(u)B_{\partial hPh}(u) = f(u)\begin{bmatrix} \Sigma(u)_{11}^{-1}E\left[\left(\frac{\partial m_t(h_0)}{\partial h^\tau}\right)^\tau\sum_{j=1}^{2}P_{jj}(u)m_t(h_0)|U_t = u\right]\int v^{2(j-1)}K(v)^2dv \\ \underline{0} \end{bmatrix},
$$

where

$$
\begin{aligned}
T_1(u) &= \left[tr\left(\Sigma(u)_{11}^{-1}E\left[\tfrac{\partial^2 m_{1t}(h_0)}{(\partial h)^{\otimes 2}}|U_t = u\right]\right), \ldots, tr\left(\Sigma(u)_{11}^{-1}E\left[\tfrac{\partial^2 m_{lt}(h_0)}{(\partial h)^{\otimes 2}}|U_t = u\right]\right)\right]^\tau, \\
T_2(u) &= \left[tr\left(\Sigma(u)_{22}^{-1}E\left[\tfrac{\partial^2 m_{1t}(h_0)}{(\partial h)^{\otimes 2}}|U_t = u\right]\right), \ldots, tr\left(\Sigma(u)_{22}^{-1}E\left[\tfrac{\partial^2 m_{lt}(h_0)}{(\partial h)^{\otimes 2}}|U_t = u\right]\right)\right]^\tau,
\end{aligned}
$$

and $H_{jj}(u)$ and $P_{jj}(u)$ $(j = 1, 2)$ are the upper left and lower right blocks of the (block diagonal) matrices (16) and similarly for $\Sigma(u)_{jj}^{-1}$ $(j = 1, 2)$.

Theorem 3 shows a number of interesting features: first, the efficient LLGMM estimator has an additional bias term compared to the LLGEL estimator, which might explain the better finite sample performance of the two LLGEL estimators considered in the next section. Secondly, the LLEL estimator (or any other LLGEL estimator with $\rho_3 = -2$) has the smallest second order bias, in terms of its components. Note however that if $E(m_t(h_0)^{\otimes 2}m_{jt}(h_0)|U_t = u) = 0$ $(j = 1, \ldots, l)$, then all LLGEL estimators have the same second order bias, namely $B_{LLGEL}(u) = H(u)(B_{\partial hHh}(u) - B_{G\Sigma}(u)/2)/Tb$. Thirdly, in the important case of moment conditions models based on instrumental variables, that is when $m_t(h_0) = q(X_t)e_t(h_0)$, where $q(X_t)$ is a vector of known instruments and $e_t(h_0)$ is a scalar valued "residual function" assumed to be twice continuously differentiable, Corollary 1 shows that the second order bias of the LLEL estimator of h_0 is bounded (in the sense that it does not depend on $\dim(q(X_t))$) as opposed to that of the efficient LLGMM estimator that depends on the degree of overidentification. Furthermore, the difference between the second order biases of the efficient LLGMM and of the LLEL estimators increases linearly in the degree of overidentification.

Corollary 1. *Assume that there exist two functions* $M_j(Z_t)$ $(j = 7, 8)$ *such that* $\|q(X_t)\partial^2 e_t(h_0)/\partial h^\tau \partial h_l\| \leq M_7(Z_t)$ $(l = 1, ..., k)$ *and* $\|(\partial e_t(h_0)/\partial h) q(X_t)' q(X_t) e_t(h_0)\| \leq M_8(Z_t)$ *with* $E(M_j(Z_i)^2 | U_t = u) < \infty$ *in an open neighborhood of* $u \in \mathcal{U}$; *then under the assumptions of Theorem 3, for two positive constants* C_1 *and* C_2

$$\|B_{LLEL}(u)\| \leq \frac{C_1}{Tb} \|\Sigma(u)_{11}^{-1}\|,$$
$$e_j^\tau (B_{LLGMM}(u) - B_{LLEL}(u)) \geq \frac{C_2}{Tb}(l - k) \quad j = 1, ..., l,$$

where e_j *is the jth basis vector and*

$$\Sigma(u)_{11} = \frac{1}{f(u)} E\left(\left(q(X_t)\frac{\partial e_t(h_0)}{\partial h}\right)^\tau \middle| U_t = u\right) \left[E\left(q(X_t)^{\otimes 2} e_t(h_0)^2\right) | U_t = u \int K^2(v) dv\right]^{-1}$$
$$\times E\left(\left(q(X_t)\frac{\partial e_t(h_0)}{\partial h}\right) \middle| U_t = u\right).$$

Finally, in the general case where one is interested only in the local estimator of $h_0(u)$, the second order bias expressions for the efficient LLGMM $\hat{h}_1(u)$ and LLGEL $\hat{h}_1^\rho(u)$ estimators are

$$\begin{aligned} B_{LLGMM}(u) = & \frac{f(u)}{Tb}\left\{ H_{11}(u) E\left(\frac{\partial m_t(h_0)}{\partial h^\tau} H_1(u) m_t(h_0) \middle| U_t = u\right) \int K(v)^2 dv \right. \\ & - \frac{H_{11}(u)}{2}\left[tr\left(\Sigma(u)_{11}^{-1} E\left[\frac{\partial^2 m_{1t}(h_0)}{(\partial h)^{\otimes 2}} \middle| U_t = u\right]\right), ..., \right. \\ & \left. tr\left(\Sigma(u)_{11}^{-1} E\left[\frac{\partial^2 m_{lt}(h_0)}{(\partial h)^{\otimes 2}} \middle| U_t = u\right]\right)\right]^\tau \\ & + H_{11}(u) E[m_t(h_0)^{\otimes 2} \sum_{j=1}^{2} P_{jj}(u) m_t(h_0) | U_t = u] \int v^{2(j-1)} K(v)^3 dv \\ & \left. + \Sigma(u)_{11}^{-1} E\left[\left(\frac{\partial m_t(h_0)}{\partial h^\tau}\right)^\tau \sum_{j=1}^{2} P_{jj}(u) m_t(h_0) | U_t = u\right] \int v^{2(j-1)} K(v)^2 dv \right\}, \end{aligned} \tag{19}$$

$$\begin{aligned} B_{LLGEL}(u) = & \frac{f(u) H_{11}(u)}{Tb}\left\{\left(1 + \frac{\rho_3}{2}\right) E\left[m_t(h_0)^{\otimes 2} \sum_{j=1}^{2} P_{jj}(u) m_t(h_0) \middle| U_t = u\right] \right. \\ & \times \int v^{2(j-1)} K(v)^3 dv + E\left(\frac{\partial m_t(h_0)}{\partial h^\tau} H_{11}(u) m_t(h_0) \middle| U_t = u\right) \int K(v)^2 dv \\ & - \frac{1}{2}\left[tr\left(\Sigma(u)_{11}^{-1} E\left[\frac{\partial^2 m_{1t}(h_0)}{(\partial h)^{\otimes 2}} \middle| U_t = u\right]\right), ..., \right. \\ & \left.\left. tr\left(\Sigma(u)_{11}^{-1} E\left[\frac{\partial^2 m_{lt}(h_0)}{(\partial h)^{\otimes 2}} \middle| U_t = u\right]\right)^\tau \right\}. \right. \end{aligned} \tag{20}$$

Apart from being theoretically interesting, the expressions given in (18) (and therefore in (19) and (20)) are practically useful because they can be used to obtain second order bias corrected LLGMM and LLGEL estimators. To describe them, let

$$\hat{G}(u) = \hat{f}(u) diag\left[\frac{1}{T\hat{b}}\sum_{t=1}^{T}\frac{\partial m_t(\hat{h}_1)}{\partial h^\tau}, \frac{1}{T\hat{b}}\sum_{t=1}^{T}\frac{\partial m_t(\hat{h}_1)}{\partial h^\tau}\int v^2 K(v)dv\right]\otimes \omega_t(u),$$

$$\hat{\Omega}(u) = \hat{f}(u) diag\left[\int K^2(v)dv, \int v^2 K^2(v)dv\right]\frac{1}{T\hat{b}}\sum_{t=1}^{T} m_t(\hat{h}_1)^{\otimes 2}\omega_t(u),$$

with $\hat{\Sigma}(u)$, $\hat{H}(u)$ and $\hat{P}(u)$ defined accordingly, and let

$$\hat{H}(u)\hat{B}_{\partial hHh}(u) = \hat{f}(u)\begin{bmatrix} \hat{H}_{11}(u)\frac{1}{T\hat{b}}\sum_{t=1}^{T}\frac{\partial m_t(\hat{h}_1)}{\partial h^\tau}\hat{H}_{11}(u) m_t(\hat{h}_1)\omega_t(u)\int K(v)^2 dv \\ \hat{H}_{22}(u)\frac{1}{T\hat{b}}\sum_{t=1}^{T}\frac{\partial m_t(,\hat{h}_1)}{\partial h^\tau}\hat{H}_{22}(u) m_t(\hat{h}_1)\omega_t(u)\int v^2 K(v)^2 dv \end{bmatrix},$$

$$\hat{H}(u)\hat{B}_{G\Sigma}(u) = \hat{f}(u)\begin{bmatrix} \hat{H}_{11}(u)\hat{T}_1(u)\int K(v)^2 dv \\ \hat{H}_{22}(u)\hat{T}_2(u)\int v^2 K(v)dv \end{bmatrix},$$

$$\hat{T}_j(u) = \Bigg[tr(\hat{\Sigma}(u)_{jj}^{-1}\frac{1}{T\hat{b}}\sum_{t=1}^{T}\frac{\partial^2 m_{1t}(\hat{h}_1)}{(\partial h)^{\otimes 2}}\omega_t(u)), \ldots,$$
$$tr(\hat{\Sigma}(u)_{jj}^{-1}\frac{1}{T\hat{b}}\sum_{t=1}^{T}\frac{\partial^2 m_{1t}(\hat{h}_1)}{(\partial h)^{\otimes 2}}\omega_t(u))\Bigg]^\tau,$$

$$\hat{H}(u)\hat{B}_{h^3P}(u) = \hat{f}(u)\begin{bmatrix} \hat{H}_{11}(u)\frac{1}{T\hat{b}}\sum_{t=1}^{T} m_t(\hat{h}_1)^{\otimes 2}\sum_{j=1}^{2}\hat{P}_{jj}(u) m_t(\hat{h}_1)\omega_t(u)\int v^{2(j-1)}K(v)^3 dv \\ \underline{0} \end{bmatrix},$$

$$\hat{\Sigma}(u)\hat{B}_{\partial hPh}(u) = \hat{f}(u)\begin{bmatrix} \hat{\Sigma}(u)_{jj}^{-1}\frac{1}{T\hat{b}}\sum_{t=1}^{T}\left(\frac{\partial m_t(\hat{h}_1)}{\partial h^\tau}\right)^\tau\sum_{j=1}^{2}\hat{P}_{jj}(u) m_t(\hat{h}_1)\omega_t(u)\int v^{2(j-1)}K(v)^2 dv \\ \underline{0} \end{bmatrix}.$$

Then the estimators of the second order bias terms are

$$\begin{aligned} \hat{B}_{LLGMM}(u) &= \frac{1}{T\hat{b}}\left[\hat{H}(u)\left(\hat{B}_{\partial hHh}(u) - \frac{\hat{B}_{G\Sigma}(u)}{2} + \hat{B}_{h^3P}(u)\right) - \hat{\Sigma}(u)\hat{B}_{\partial hPh}(u)\right], \\ \hat{B}_{LLGEL}(u) &= \frac{1}{T\hat{b}}\left[\left(1+\frac{\rho_3}{2}\right)\hat{H}(u)\hat{B}_{h^3P}(u) + \hat{H}(u)\left(\hat{B}_{\partial hHh}(u) - \frac{\hat{B}_{G\Sigma}(u)}{2}\right)\right], \end{aligned} \tag{21}$$

and the bias corrected local linear estimators are

$$\hat{h}_c^\circ(u) = \hat{h}^\circ(u) - \hat{B}_\times(u),$$

where $\hat{h}_c^\circ(u) = \left[\hat{h}_{1c}^\circ(u)^\tau \quad \hat{h}_{2c}^\circ(u)^\tau\right]^\tau$, $\hat{h}^\circ(u)$ is either $\hat{h}(u)$ or $\hat{h}^\rho(u)$ and $\hat{B}_\times(u)$ is either $\hat{B}_{LLGMM}(u)$ or $\hat{B}_{LLGEL}(u)$. The following corollary to Theorem 3 shows that the bias corrected local linear estimators are second order correct.

Corollary 2. *Under the same assumptions of Theorem 3,*

$$bias\left((Tb)^{1/2}\hat{h}_c^{\circ}(u)\right) = o\left(\frac{1}{(Tb)^{1/2}}\right).$$

4. Monte Carlo simulations

In this section, we illustrate the finite sample properties of the LLGMM and LLGEL estimators and their bias corrected versions. The estimators we consider are the efficient LLGMM defined in (10), the LLEL defined in (8) and the LLET defined in (9). We focus on the finite sample properties of $\hat{h}_1$ and $\hat{h}_1^{\rho}$. The discussion after Theorem 2 shows that the optimal (minimizing the asymptotic integrated mean squared error) bandwidth b_* is of order $O(T^{-1/5})$ and standard data driven methods could be used to select it. On the other hand, the results of Theorem 3 require undersmoothing, hence least squares cross-validation or other bandwidth selection methods cannot be used directly to automatically choose the bandwidth. One possibility is to follow the *ad hoc* procedure suggested by Otsu et al.'s (2015) and use least squares cross-validation to estimate the bandwidth and then multiply the resulting bandwidth by a power of the sample size that is consistent with undersmoothing. In this paper, we consider another method, that is, similar to the *ad hoc* cross-validation method of Otsu et al. (2015) but is less computationally intensive. Specifically, we consider a twofold 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

$$\hat{h}_p = \mathrm{argmin}_{h_1,h_2\in\mathcal{H}_C}\frac{1}{T_v b_p}\sum_{t\in S_v} g_t(h_1,h_2)^{\tau}K\left(\frac{U_t-u}{b_p}\right)\hat{\Omega}(u)^{-1}\frac{1}{T_v b_p}\sum_{t\in S_v} g_t(h_1,h_2)K\left(\frac{U_t-u}{b_p}\right),$$
$$\hat{h}_p^{\rho} = \mathrm{argmin}_{h_1,h_2\in\mathcal{H}_C}\frac{1}{T_v b_p}\sum_{t\in S_v}\rho\left(v_t\left(u,\hat{\lambda},h_1,h_2\right)\right),$$

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

$$\hat{b} = \mathrm{argmin}_{b\in B}\frac{1}{T_{1-v}b}\sum_{t\in S_{1-v}} g_t\left(\hat{h}_{1p},\hat{h}_{2p}\right)^{\tau}K\left(\frac{U_t-u}{b}\right)\hat{\Omega}(u)^{-1}\frac{1}{T_{1-v}b}\sum_{t\in S_{1-v}} g_t\left(\hat{h}_{1p},\hat{h}_{2p}\right)K\left(\frac{U_t-u}{b}\right)$$
$$\hat{b}^{\rho} = \mathrm{argmin}_{b\in B}\frac{1}{T_{1-v}b}\sum_{i\in S_{1-v}}\rho\left(v_t\left(u,\hat{\lambda},\hat{h}_{1p}^{\rho},\hat{h}_{2p}^{\rho}\right)\right), \tag{22}$$

where B is a grid of possible values of b, and $\hat{h}_{jp}$ and $\hat{h}_{jp}^{\rho}$ $(j = 1, 2)$ are the estimators based on the pilot bandwidths b_p and $\hat{h}_p^{\rho}$. Finally, as in Otsu et al. (2015), $\hat{b}$ and $\hat{b}^{\rho}$ are multiplied by T^{-c}, where $c > 0$ is a value consistent with undersmoothing.

Remark 1. It should be noted that the LLGEL estimator has the practical advantage over the efficient LLGMM estimator of needing only one bandwidth, since the efficient LLGMM estimator requires an additional bandwidth to compute the preliminary consistent estimators $\tilde{h}_j$ $(j = 1, 2)$ for $\hat{\Omega}(u)^{-1}$. In the simulations below, we use the same *ad hoc* method described above (with $c = 0.25$), but it should also be noted that the choice of the additional bandwidth does not seem to have any important bearings on the performance of the efficient LLGMM estimator in terms of its finite sample MSE. See the additional simulations results reported in Figure 1(A and B) in the Supplemental Appendix.

In the simulations below, the kernel function used is the Epanechnikov one, that is, $K(u) = (3/4)(1-u^2)$ for $|u| \le 1$, which results in $\int v^2K(v)dv = 1/5$, $\int K(v)^2dv = 3/5$, $\int v^2K(v)^2dv =$

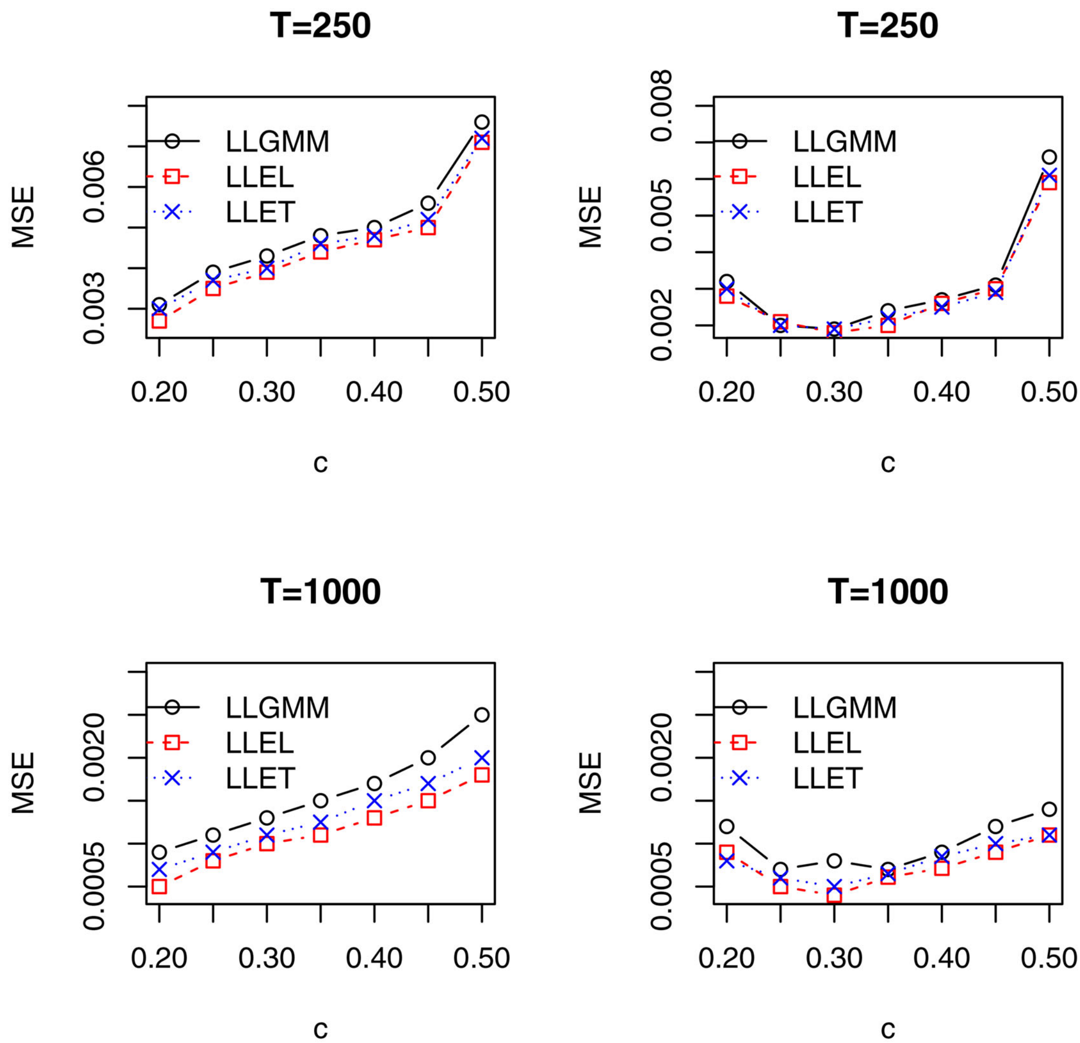

Figure 1. MSE of the three local linear estimators for $\dim(q_t(X_t)) = 3$ and $U_t \sim U(0,1)$, with the original estimators on the left and their bias corrected versions on the right.

$4/35$, $\int K(v)^3 dv = 24/35$ and $\int v^2 K(v)^3 dv = 8/105$. The terms $\hat{G}(u), \hat{\Omega}(u)$ and all the others appearing in the bias correction terms (21) are computed using again the Epanechnikov kernel with bandwidth selected by least squares cross-validation.

We consider Examples 1 and 2 (Example 3 can be found in the Supplemental Appendix). For Example 1, the model is specified as

$$\begin{aligned} Y_t &= \exp\left(h_{10}(U_t) + X_{1t} h_{20}(U_t)\right) + \varepsilon_t, \\ X_{1t} &= 0.4 X_{2t} + \eta_t, \end{aligned}$$

where $h_{10}(U_t) = \sin(\pi U_t/2)$, $h_{20} = \cos(\pi U_t)$, $X_{2t} = \rho X_{2t-1} + \zeta_t$ 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 $U_t \sim \Phi((a\xi_t + b\xi_{t-1})/\sqrt{a^2+b^2})$, where Φ is the cumulative

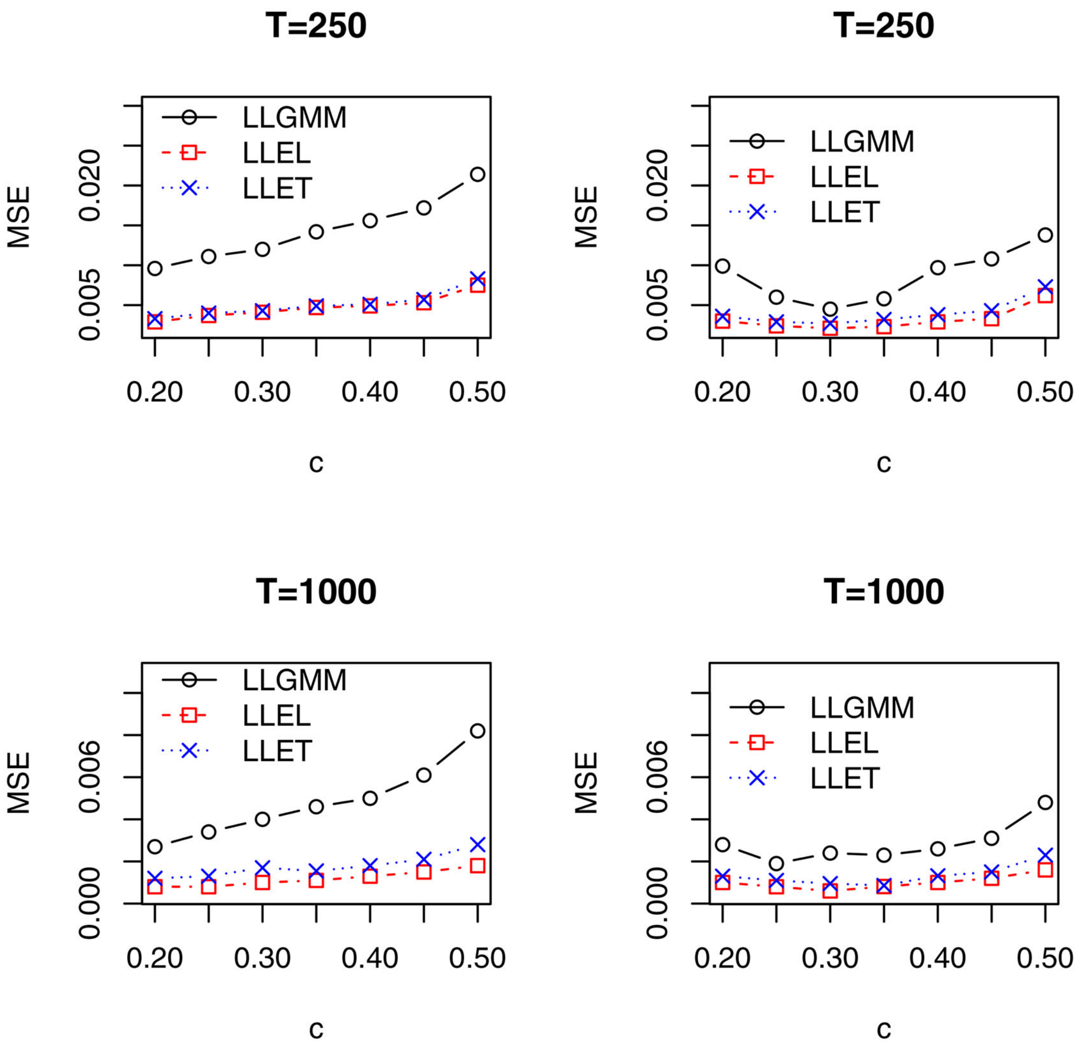

Figure 2. MSE of the three local linear estimators for $\dim(q_t(X_t)) = 9$ and $U_t \sim U(0,1)$, with the original estimators on the left and their bias corrected versions on the right.

standard normal distribution and $\xi_t \sim N(0,1)$. Note that in the latter case U_t is a 1-dependent process. In the simulations, the vector of instruments $q(X_{2t})$ is specified as either $\left[1, X_{2t}, X_{2t}^2\right]^\tau$ or $\left[1, X_{2t}, \ldots, X_{2t}^6\right]^\tau$, with $\rho = 0.4$ and $a = 0.9$, $b = 0.1$, and seven alternative bandwidths chosen by the *ad hoc* cross validated method described above with $c = (0.20, 0.25, \ldots, 0.50)$.

Figures 1–4 report the combined MSE's of $\hat{h}_1^\circ$

$$MSE\left(\hat{h}_1^\circ\right) = \frac{1}{2T}\sum_{j=1}^{2}\sum_{t=1}^{T}\left(\hat{h}_{1j}^\circ(U_t) - h_{j0}(U_t)\right)^2,$$

where $\hat{h}_{1j}^\circ$ is either the LLGMM $\left[\hat{h}_{11}^\tau, \hat{h}_{12}^\tau\right]^\tau$ or the LLEL $\left[\hat{h}_{11}^{el\tau}, \hat{h}_{12}^{el\tau}\right]^\tau$ or the LLET $\left[\hat{h}_{11}^{et\tau}, \hat{h}_{12}^{et\tau}\right]^\tau$ estimators for $[h_{10}^\tau, h_{20}^\tau]^\tau$, or their second order bias corrected versions $\hat{h}_{1c}$, $\hat{h}_{1c}^{el}$ and $\hat{h}_{1c}^{et}$ for two sample sizes, $T = 250$ and $T = 1000$.

Figures 1–4 show a number of interesting features: first, as expected, the MSEs of the three local linear estimators are minimized at the optimal rate $(c = 0.2)$, increase with the degree of

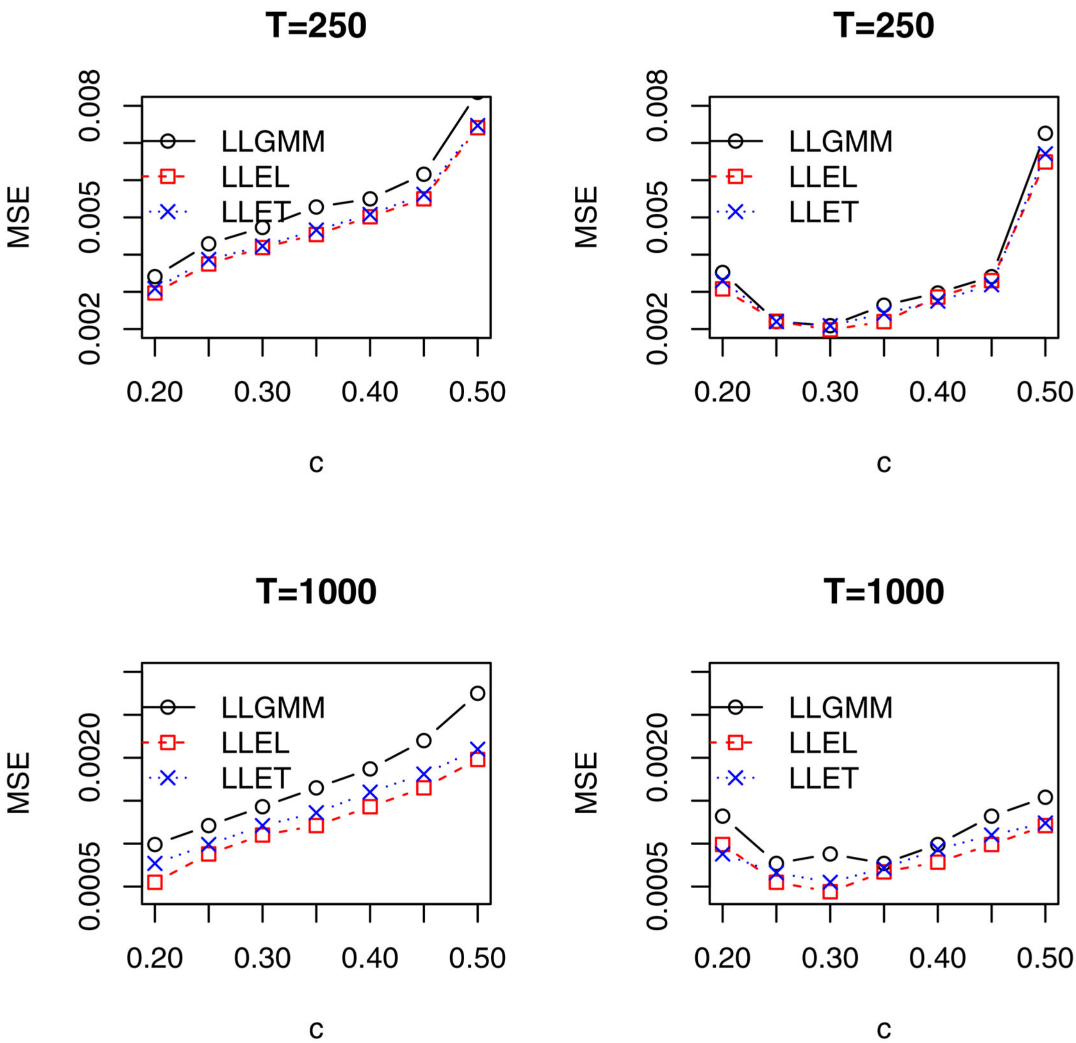

Figure 3. MSE of the three local linear estimators for $\dim(q_t(X_t)) = 3$ and $U_t \sim \Phi((0.9\xi_t + 0.1\xi_{t-1})/0.9)$, with the original estimators on the left and their bias corrected versions on the right.

undersmoothing and decrease as the sample size increases. Furthermore, among the three local linear estimators, the LLEL has the smallest MSE, which is consistent with the first result of Corollary 1. Second, the proposed second order bias corrections are effective in reducing the MSEs of the original local linear estimators for a reasonable set of undersmoothed bandwidths – note that with the optimal bandwidth the corrections are not useful at all, which, again, is consistent with the theoretical results of the previous section. Third, the dimension of the vector of instruments $q(X_{2t})$ impacts negatively the MSE of the LLGMM estimator and less so that of the LLEL and LLET estimators, which is again consistent with the results of Corollary 1.

Next, we consider the same example considered by Cai, Ren, and Sun (2015), which corresponds to Example 2 with $q(X_t) = 1$ and $J = 25$. To be specific, let $U_t = \rho U_{t-1} + 0.01\varepsilon_t$ and $R_{M,t+1} = r(U_t) + 0.05\eta_t$, where ε_t and η_t are independent standard normals, so that under the

Table 2. MSE for the LLGMM $\hat{h}_1$, LLEL $\hat{h}_1^{el}$ and LLET $\hat{h}_1^{et}$ estimators and their bias corrected versions $\hat{h}_{1c}$, $\hat{h}_{1c}^{el}$ and $\hat{h}_{1c}^{et}$..

T	$\hat{h}_1$	$\hat{h}_{1c}$	$\hat{h}_1^{el}$	$\hat{h}_{1c}^{el}$	$\hat{h}_1^{et}$	$\hat{h}_{1c}^{et}$
250	0.038[a]	0.011[a]	0.033[a]	0.009[a]	0.035[a]	0.010[a]
	0.041[b]	0.012[b]	0.034[b]	0.009[b]	0.032[b]	0.009[b]
	0.042[c]	0.014[c]	0.037[c]	0.010[c]	0.034[c]	0.011[c]
1000	0.009[a,c]	0.005[a]	0.004[a]	0.003[a]	0.006[a]	0.005[a]
	0.008[b]	0.003[b]	0.004[b]	0.002[b]	0.006[b]	0.004[b]
	0.010[c]	0.004[c]	0.006[c]	0.003[c]	0.008[c]	0.006[c]

[a] $\hat{b}T^{-1/4}$.
[b] $\hat{b}T^{-1/3}$.
[c] $\hat{b}T^{-2/5}$.

mean covariance efficiency assumption[2]

$$h_0(U_t) = \frac{0.01r(U_t)}{(0.05)^2 + (0.01)^2 r(U_t)^2} \tag{23}$$

and $R_{j,t+1} = e_{j,t}/(1 - h_0(U_t)R_{M,t+1})$ with $e_{j,t} = 0.05e_{j,t-1} + v_{j,t}$, $r(U_t) = 0.01(1 + U_t^2)$ and $v_{j,t}$ a standard normal.

Table 2 reports the MSE of the same three local linear estimators as those considered in Figures 1–4 for three different choices of bandwidths consistent with undersmoothing.

The result of Table 2 provide further evidence of the good finite sample properties of the LLGEL estimators and the effectiveness of the proposed second order bias corrections.

5. Empirical application

We illustrate the applicability of the proposed local estimation method by considering a varying coefficient specification of Fama and French's (1993) three factor model. To be specific, the stochastic discount factor as given in Example 2 of Section 2 is

$$h_{t+1} = 1 - h_0(U_t)^\tau F_{t+1},$$

where $h_0(U_t) = [h_{10}(U_t), h_{20}(U_t), h_{30}(U_t)]^\tau$ is a vector of unknown (time varying state dependent) risk prices associated to the observed risk factors $F_t = [F_{1t}, F_{2t}, F_{3t}]^\tau$, where F_{jt} $(j = 1, 2, 3)$, represent, respectively, the market excess, the size premium (small minus big) and the book to market value. The conditioning variable U_t is chosen to be either the ten year Treasury yield U_t^{10} or the BAA corporate bond yield relative to the constant maturity 10-year Treasury yield U_t^{DEF}. The latter serves as a proxy for the default risk, and it was used in the consumption-based CAPM model of Jagannathan and Wang (1996).

The three local linear estimators (LLGMM, LLEL and LLET) of the stochastic discount factor are compared with their parametric analogs (efficient GMM, EL and ET) based on the linear affine state dependent specification $h_{t+1} = 1 - (\beta_0 + \beta_{0u}U_t)^\tau F_{t+1}$, where $\beta_0 = [\beta_{10}, \beta_{20}, \beta_{30}]^\tau$ and $\beta_{0u} = [\beta_{10u}, \beta_{20u}, \beta_{30u}]^\tau$ are the unknown (time invariant) risk prices. To be specific the parametric efficient GMM, EL and ET estimators are based on the moment indicator $m_{t+1}(\beta, \beta_u) = (1 - (\beta + \beta_u U_t)^\tau F_{t+1}) \otimes R_{t+1}$, and are defined as

[2]Let $R_{p,t}$ denote the excess return of a benchmark portfolio p. If the latter is (conditionally) mean variance efficient, then
$E(R_{j,t+1}|I_t) = E(R_{p,t+1}|I_t)\frac{E(R_{j,t+1}R_{p,t+1}|I_t)}{E(R_{p,t+1}^2|I_t)}$,
where I_t is the information set, which can be rewritten as
$E(R_{j,t+1}|I_t) - E(R_{p,t+1}|I_t)\frac{E(R_{j,t+1}R_{p,t+1}|I_t)}{E(R_{p,t+1}^2|I_t)} = E(h_{t+1}R_{j,t+1}|I_t)$
and $h_{t+1} = 1 - h_{0t}(I_t)R_{p,t+1}$
$h_{0t}(I_t) = E(R_{p,t+1}|I_t)/E(R_{p,t+1}^2|I_t)$;
setting $I_t = U_t$ gives (23). See Wang (2003) for more details on the mean variance efficiency assumption.

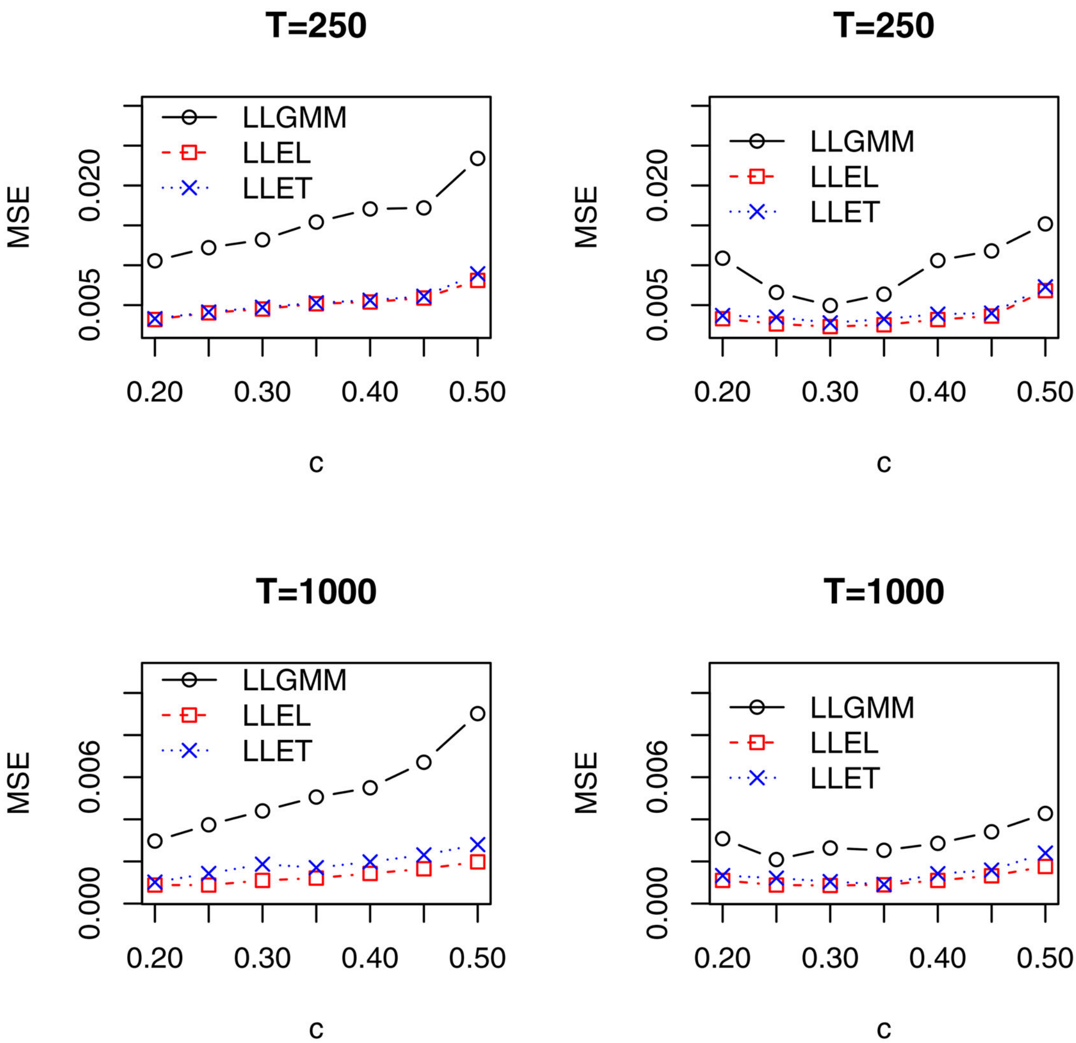

Figure 4. MSE of the three local linear estimators for $\dim(q_t(X_t)) = 9$ and $U_t \sim \Phi((0.9\xi_t + 0.1\xi_{t-1})/0.9)$, with the original estimators on the left and their bias corrected versions on the right.

$$
\begin{aligned}
\hat{\beta}_{GMM} &= \text{argmin}_{\beta,\beta_u\in B}\frac{1}{T}\sum_{t=1}^{T-1} m_{t+1}(\beta,\beta_u)^{\tau}\hat{\Omega}\big(\tilde{\beta},\tilde{\beta}_u\big)^{-1} m_{t+1}(\beta,\beta_u),\\
\hat{\beta}_{EL} &= \text{argmin}_{\beta,\beta_u\in B}\frac{1}{Q}\sum_{t=1}^{Q}\log\Big(1-\hat{\lambda}^{\tau} m^{B}_{t+1}(\beta,\beta_u)\Big),\\
\hat{\beta}_{ET} &= \text{argmin}_{\beta,\beta_u\in B}\frac{1}{Q}\sum_{t=1}^{Q}\exp\Big(\hat{\lambda}^{\tau} m^{B}_{t+1}(\beta,\beta_u)\Big),
\end{aligned}
\tag{24}
$$

where $\hat{\Omega}(\tilde{\beta},\tilde{\beta}_u)$ is the Newey and West (1987) estimator of the long run covariance matrix $\lim_{T\to\infty} Var(\sum_{t=1}^{T} m_{t+1}(\beta_0,\beta_{0u})/T^{1/2})$ with $\tilde{\beta}$ and $\tilde{\beta}_u$ preliminary consistent estimators of β_0 and β_{0u}, $m^{B}_{t+1}(\beta,\beta_u) = \sum_{k=1}^{M} m_{t+k}(\beta,\beta_u)/M$ is an overlapping block of observations with block length $M\to\infty$ as $T\to\infty$ at an appropriate rate and $Q = \lfloor T-M\rfloor + 1$ with $\lfloor\cdot\rfloor$ being the integer part function. More details on the blocking technique and the asymptotics of the resulting estimators can be found in Kitamura and Stutzer (1997) and Bravo (2009).

Table 3. Average MSE of the pricing errors of local linear and parametric estimators.

	LLGMM	LLEL	LLET	GMM	EL	ET
U_t^{10}	0.0027[a]	0.0023[a]	0.0025[a]	0.0119	0.0092	0.0098
	0.0032[b]	0.0028[b]	0.0030[b]			
U_t^{DEF}	0.0026[a]	0.0022[a]	0.0023[a]			
	0.0030[b]	0.0027[b]	0.0027[b]	0.0099	0.0081	0.0083

[a] $\hat{b}^{o} = \hat{b}T^{-1/5}$.
[b] $\hat{b}^{u} = \hat{b}T^{-1/4}$.

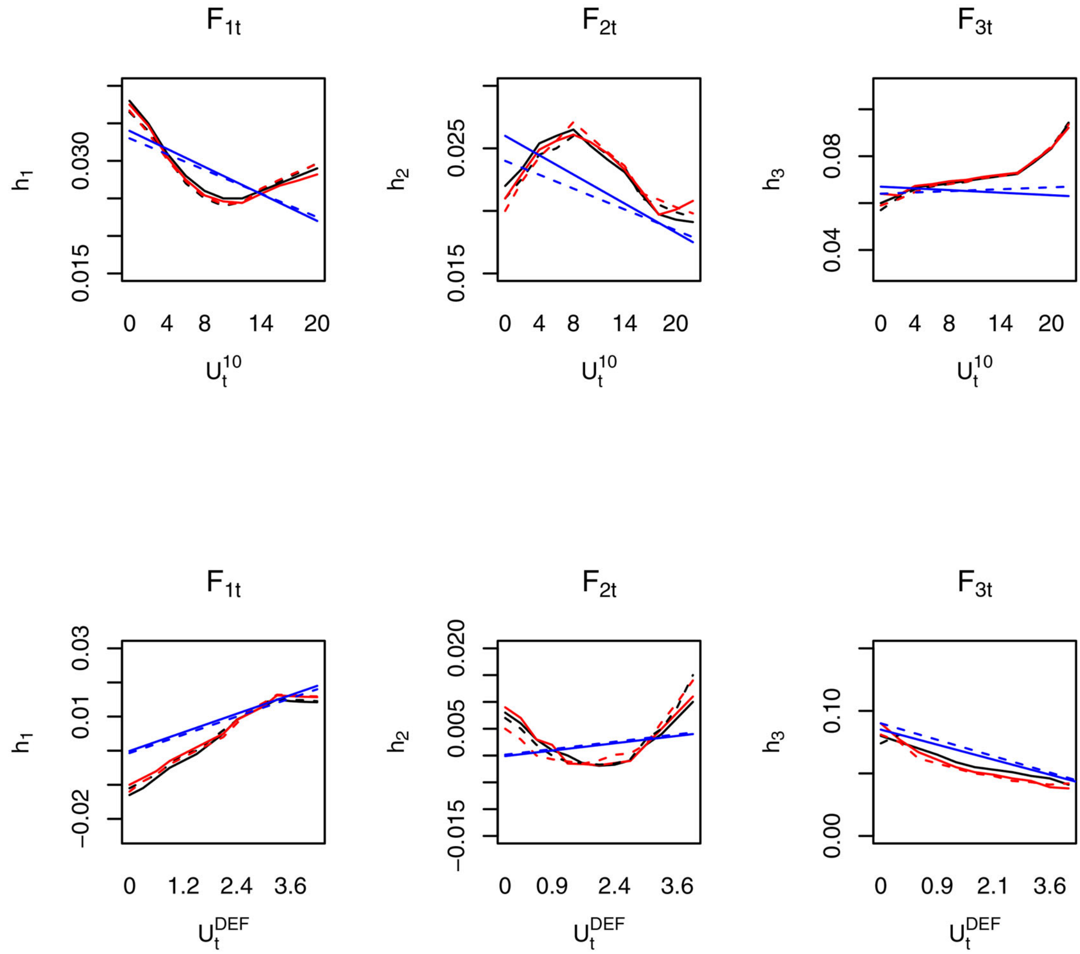

Figure 5. LLGMM (solid black line), LLEL (solid red line), their bias corrected versions (dashed black and dashed red lines) and parametric GMM and EL (solid and dashed blue lines) estimates of risk prices.

We use the monthly returns on the 25 size-sorted portfolios for US equities from Kenneth French's data library[3] as risky assets $R_{i,t}$; the excess returns are computed over the one-month Treasury bill yield obtained from the Center for Research in Securities Prices (CRSP). The 10-year Treasury yield data is obtained from the Federal Reserve Statistical Release H.15. The corporate bond spread is obtained from the Federal Reserve Economic Data. The sample period is

[3]Available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 4. Parametric estimates of risk prices with 95% confidence intervals.

		GMM	EL	ET
U_t^{10}	$\hat{\beta}_1$	0.034	0.032	0.032
		[0.014, 0.054]	[0.011, 0.054]	[0.013, 0.049]
	$\hat{\beta}_{1u}$	0.006	0.005	0.004
		[−0.010, 0.022]	[−0.005, 0.015]	[−0.004, 0.016]
	$\hat{\beta}_2$	0.026	0.024	0.023
		[0.001, 0.051]	[0.004, 0.044]	[0.003, 0.039]
	$\hat{\beta}_{2u}$	0.012	0.010	0.011
		[−0.003, 0.027]	[0.002, 0.018]	[0.003, 0.019]
	$\hat{\beta}_3$	0.058	0.053	0.055
		[0.032, 0.082]	[0.026, 0.084]	[0.037, 0.073]
	$\hat{\beta}_{3u}$	0.001	0.0015	0.0016
		[−0.008, 0.012]	[0.006, 0.0024]	[0.007, 0.0025]
U_t^{DEF}	$\hat{\beta}_1$	0.001	0.001	0.001
		[−0.005, 0.007]	[−0.004, 0.006]	[−0.004, 0.006]
	$\hat{\beta}_{1u}$	−0.007	−0.007	−0.006
		[−0.017, 0.003]	[−0.012, −0.002]	[−0.013, −0.03]
	$\hat{\beta}_2$	0.000	−0.001	−0.001
		[−0.012, 0.012]	[−0.010, 0.006]	[−0.011, 0.007]
	$\hat{\beta}_{2u}$	−0.001	−0.001	−0.001
		[−0.006, −0.004]	[−0.005, −0.003]	[−0.005, −0.003]
	$\hat{\beta}_3$	0.088	0.089	0.086
		[0.065, 0.111]	[0.069, 0.109]	[0.065, 0.107]
	$\hat{\beta}_{3u}$	0.003	0.003	0.002
		[−0.007, 0.013]	[−0.005, 0.008]	[−0.006, 0.010]

Table 5. Average local linear estimates of risk prices with 95% confidence intervals.

	LGMM		LLEL	
U_t^{10}	$\hat{h}_1$	$\hat{h}_{1c}$	$\hat{h}_1$	$\hat{h}_{1c}$
	0.029	0.030	0.028	0.031
	[−0.004, 0.058]	[0.008, 0.051]	[0.001, 0.0575]	[0.011, 0.056]
	$\hat{h}_2$	$\hat{h}_{2c}$	$\hat{h}_2$	$\hat{h}_{2c}$
	0.023	0.022	0.022	0.021
	[−0.016. 0.058]	[−0.007, 0.0514]	[−0.008, 0.050]	[−0.001, 0.043]
	$\hat{h}_3$	$\hat{h}_{3c}$	$\hat{h}_3$	$\hat{h}_{3c}$
	0.066	0.061	0.064	0.060
	[0.027, 0.105]	[0.029, 0.092]	[0.032, 0.095]	[0.034, 0.085]
U_t^{DEF}	$\hat{h}_1$	$\hat{h}_{1c}$	$\hat{h}_1$	$\hat{h}_{1c}$
	0.030	0.028	0.029	0.030
	[−0.005, 0.075]	[0.010, 0.066]	[0.001, 0.074]	[0.014, 0.072]
	$\hat{h}_2$	$\hat{h}_{2c}$	$\hat{h}_2$	$\hat{h}_{2c}$
	0.019	0.022	0.021	0.023
	[−0.020, 0.07]	[−0.009, 0.066]	[−0.010, 0.06]	[−0.001, 0.055]
	$\hat{h}_3$	$\hat{h}_{3c}$	$\hat{h}_3$	$\hat{h}_{3c}$
	0.068	0.063	0.064	0.062
	[−0.020, 0.075]	[0.037, 0.119]	[0.037, 0.123]	[0.044, 0.110]

$1964:01-2018:12$ minus the $2008:01-2009:04$ financial crisis, for a total of $T=641$ observations.

Table 3 reports the average (over the 25 portfolios) MSEs of the pricing errors of the *LLGMM*, *LLEL* and *LLET* estimators with the two different conditioning variables U_t^{10} and U_t^{DEF}, and two bandwidths (the optimal one and an undersmoothed one with $c=0.25$), and those of their parametric analogs. The individual pricing errors can be find in the Supplemental Appendix (Tables 2A–C).

Table 3 shows that the MSEs of the local linear estimators are typically smaller than those based on their parametric analogs. This result is consistent with that obtained by Cai, Ren, and Sun (2015) (for the LLGMM estimator). Table 3 also shows that among the three local linear estimators, the LLEL one has typically the smallest MSE, which confirms the findings of the previous section.

Figure 5 shows both the nonparametric and the parametric GMM and EL estimates of the risk prices associated with the three risk factors. The ET estimates are very similar to those based on EL, hence are not shown in the figure.

Figure 5 shows that the risk prices exhibit some clear nonlinear pattern: the risk price associated to the market excess factor F_{1t} is concave in the 10-year Treasury yield conditioning variable and the risk price associated to the size premium factor F_{2t} is concave in the default conditioning variable. There is further evidence of nonlinearity in the risk prices associated to the book to market value in the 10-year Treasury yields and (to a lesser degree) in the risk price associated to the market excess in the default conditioning variable.

Finally, Tables 4 and 5 report, respectively, the point estimates and 95% confidence intervals of the parametric risk prices $\hat{\beta}_j$ and $\hat{\beta}_{ju}(j = 1, 2, 3)$ and the averages $\hat{h}_j^o = \sum_{t=1}^T \hat{h}_j^o(U_t)/T$, $\hat{h}_{jc}^o = \sum_{t=1}^T \hat{h}_{jc}^o(U_t)/T$ $(j = 1, 2, 3)$, where $\hat{h}_j^o$ is either $\hat{h}_j$, or $\hat{h}_j^{el}$ and 95% confidence intervals of the nonparametric risk prices.

Tables 4 and 5 show that the combined parametric and averaged nonparametric estimates are fairly close numerically (especially those based on the bias corrected nonparametric estimators), albeit the confidence intervals based on the parametric estimators are narrower, which is to be expected given their faster convergence rate. Taken together, Tables 3–5 and Figure 5 show the advantages of using local linear estimators (and their bias corrected versions) over traditional parametric estimators for stochastic discount factor models, both in terms of MSE of the pricing errors and in terms of capturing the variability of the risk prices.

6. Conclusions

In this paper, we consider a local linear version of the GEL approach that can be used to estimate the unknown infinite dimensional parameter in nonparametric moment conditions models. We derive a new strong uniform convergence rate and the asymptotic normality of the proposed estimator. We also obtain second order stochastic expansions for both the efficient local linear GMM and local linear GEL, which are both theoretically and practically useful, as they explain why local linear GEL estimators are typically characterized by better finite sample properties than those based on the efficient local linear GMM across a range of different undersmoothed bandwidths, and can be used to obtain analytical expressions of the second order bias of the local linear estimators and correct for it. Monte Carlo simulations show that the local GEL estimators perform well in finite samples and that the proposed bias corrected version effectively reduce the MSE. An empirical application, where a varying coefficient version of Fama and French's (1993) three factor model is estimated, illustrates the applicability and usefulness of the local estimation method proposed in this paper.

Acknowledgements

I am grateful to the Editor, an Associate Editor and two Referees for useful comments and suggestions that improved considerably the paper. The usual disclaimer applies.

ORCID

Francesco Bravo http://orcid.org/0000-0002-8034-334X

References

Bravo, F. (2009). Blockwise generalised empirical likelihood inference for nonlinear dynamic moment conditions models. *Econometrics Journal* 12(2):208–231. doi:10.1111/j.1368-423X.2009.00286.x

Bravo, F. (2016). Local information theoretic methods for smooth coefficients dynamic panel data models. *Journal of Time Series Analysis* 37(5):690–708. doi:10.1111/jtsa.12190

Cai, Z. (2003). Nonparametric estimating equations for time series data. *Statistics & Probability Letters* 62(4): 379–390. doi:10.1016/S0167-7152(03)00042-7

Cai, Z., Chen, L., Fang, Y. (2015). Semiparametric estimation of partially varying coefficients dynamic panel data models. *Econometric Reviews* 34(6–10):695–719. doi:10.1080/07474938.2014.956569

Cai, Z., Fan, J., Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95(451):941–956. doi:10.1080/01621459.2000.10474284

Cai, Z., Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24(5):1321–1342. doi:10.1017/S0266466608080523

Cai, Z., Ren, Y., Sun, L. (2015). Pricing kernel estimation: a local estimating equation approach. *Econometric Theory* 31(3):560–580. doi:10.1017/S0266466614000589

Carroll, R., Fan, J., Gijbels, I. Wand, M. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* 92:477–489. doi:10.1080/01621459.1997.10474001

Chen, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika* 83(2):329–341. doi:10.1093/biomet/83.2.329

Chen, X., Qin, S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* 87(4): 946–953. doi:10.1093/biomet/87.4.946

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

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

Fama, E., French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1):3–56. (doi:10.1016/0304-405X(93)90023-5

Fan, J., Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.

Fang, Y., Ren, Y., Yuan, Y. (2011). Nonparametric estimation and testing of stochastic discount factor. *Finance Research Letters* 8(4):196–205. doi:10.1016/j.frl.2011.04.001

Gospodinov, N., Otsu, T. (2012). Local GMM estimation of time series models with conditional moment restrictions. *Journal of Econometrics* 170(2):476–490. doi:10.1016/j.jeconom.2012.05.017

Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Annals of Statistics* 20:695–711.

Hansen, B. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24(3):726–748. doi:10.1017/S0266466608080304

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50(4): 1029–1054. doi:10.2307/1912775

Hansen, L., Heaton, J., Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* 14(3):262–280. doi:10.2307/1392442

Jagannathan, R., Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance* 51(1):3–53. doi:10.1111/j.1540-6261.1996.tb05201.x

Kitamura, Y., Stutzer, M. (1997). An information theoretic alternative to generalized method of moments estimation. *Econometrica* 65(4):861–874. doi:10.2307/2171942

Lewbel, A. (2007). A local generalized method of moments estimator. *Economics Letters* 94(1):124–128. doi:10.1016/j.econlet.2006.08.011

Li, Q., Racine, J. (2004). Cross validated local linear nonparametric regression. *Statistica Sinica* 14:485–512.

Liang, K., Zeger, S. (1986). Longitudinal data analysis using generalised linear models. *Biometrika* 73(1):13–22. doi:10.1093/biomet/73.1.13

Liebscher, E. (1996). Strong convergence of sums of alpha-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* 65(1):69–80. doi:10.1016/S0304-4149(96)00096-8

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17(6):571–599. doi:10.1111/j.1467-9892.1996.tb00294.x

McCullagh, P., Nelder, J. (1989). *Generalized Linear Models*. London: Chapman and Hall.

Newey, W. K., West, K. D. (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708. doi:10.2307/1913610

Newey, W., Smith, R. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72(1):219–256. doi:10.1111/j.1468-0262.2004.00482.x

Otsu, T., Xu, K.-L., Matsushita, Y. (2015). Empirical likelihood for regression discontinuity design. *Journal of Econometrics* 186(1):94–112. doi:10.1016/j.jeconom.2014.04.023

Qu, A., Lindsay, B. G., Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* 87(4):823–836. doi:10.1093/biomet/87.4.823

Severini, T., Staniswalis, J. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association* 89(426):501–511. doi:10.1080/01621459.1994.10476774

Wang, K. (2003). Asset pricing with conditioning information: A new test. *The Journal of Finance* 58(1):161–196. doi:10.1111/1540-6261.00521

Wang, K. (2002). Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio. *Journal of Empirical Finance* 9(2):133–169. doi:10.1016/S0927-5398(01)00049-4