Local polynomial estimation of nonparametric general estimating equations

Francesco Bravo*
University of York

January 2023

Abstract

This paper investigates estimation of nonparametric general estimating equations. The paper considers estimators based on local polynomial versions of the generalized method of moments and the generalized empirical likelihood approaches, and derives their asymptotic distribution under weak dependency of the observations.

Keywords: α-mixing, Estimating equations, Local polynomial estimation,

1 Introduction

General estimating equations models[1] are very useful generalizations of the standard estimating equations models widely used in statistics, see for example the quasi likelihood approach to generalized linear models of McCullagh and Nelder (1989) and **?**, as they allow for the dimension of the estimating equations to be bigger than the dimension of the unknown parameters, and include, for example, the quadratic inference functions developed by **?**. When the unknown parameters are finite dimensional, their estimation is typically carried out using Hansen's (1982) (see also **?**) generalized method of moments (GMM) approach, or, more recently, Newey and Smith's (2004) generalized empirical likelihood (GEL) approach. Both methods have their own merits: GMM is often easier to compute (especially for linear models), but requires a two-step

*I am grateful to an Associate Editor for useful comments. The usual disclaimer applies.
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] We use the term general rather than generalized to avoid any confusion with the generalized estimating equations (GEE) terminology that has become predominant in the quasi likelihood estimation with longitudinal data literature (see for example **?**).

estimation procedure to achieve the minimum (asymptotic) variance bound, whereas GEL can be computationally more demanding, but it achieves the variance bound automatically (the so-called internal studentization property).

The main contribution of this paper is to extend these two general approaches to nonparametric general estimating equations models, that is general estimating equations models with unknown infinite dimensional parameters. This extension is theoretically interesting but, more importantly, empirically relevant as it is now widely acknowledged that many general estimating equations models are misspecified, with the cause of misspecification often attributed to parametric constraints implied by an underlying theoretical model. For example, in the so-called C-CAPM (consumer based capital assets pricing model), see for example **?**, the general estimating equations are based on a parametric specification of the utility function. More generally, any parametric specification of the so-called pricing kernel in the popular stochastic discount factor model used in the asset pricing literature (see Cochrane (2001) for a review, and Example 2 in Section 2 below) can lead to misspecified general estimating equations. In another important example, the parametric restrictions implied by the rational expectations assumption, see for example **?**, can also lead to misspecified estimating equations. The nonparametric general estimating equations model considered in this paper is an important generalization of a number of papers, including 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) and the nonparametric stochastic discount factor model of Fang, Ren and Yuan (2011) and of Cai, Ren and Sun (2015), among many others, because it allows for the dimension of the estimating equations to be bigger than that of the unknown infinite dimensional parameters. In particular, instrumental variable estimation, which is widely used in econometrics and statistics, see for example **?**, **?** and Example 1 in Section 2 below, is allowed.

This paper contributes to the literature on estimating equations by proposing new local polynomial versions of the GMM (LPGMM henceforth) and GEL (LPGEL henceforth) estimators. These two new estimators are important extensions of the local polynomial estimator introduced by Fan and Gijbels (1996) (see also **?**). The paper establishes the asymptotic normality of both the LPGMM and LPGEL estimators under an α-mixing assumption on the dependence of the observations (see Doukhan (1994) for examples and properties of α mixing processes). This result is important (and useful) for three reasons: first, it allows weakly dependent observations, which is particularly useful in macroeconomics and finance, since macroeconomic and financial data typically exhibit some form of serial dependence. Second, it allows to estimate derivatives, which are often of interest; for example, in finance, the first and second derivatives of the expected value of an option price with respect to the underlying asset price or its volatility are useful indicators of financial risk. Third, the minimum square error (MSE) optimal bandwidth for estimating the unknown infinite dimensional parameters depends, among other things, on an asymptotic bias term involving higher order derivatives (see the theorems in Section 3 for an expression of this bias), which have to be estimated and the local polynomial estimators proposed in this paper automatically give the estimates of such derivatives.

The rest of the paper is structured as follows: next section introduces the model, whereas Section 3 presents the main results. An Appendix contains all the proofs.

The following notation is used throughout the paper: "τ" indicates transpose, "$'$","$''$", "$'^{...j}$" denote, respectively, first, second and the $j-th$ derivative of a function with respect to its unique argument, "$\otimes$" is the Kronecker product, "$diag(\cdot)$" is a diagonal matrix, $\underline{0}$ 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 $\mathcal{Z} \subset \mathbb{R}^{dz}$ and $\mathcal{U} \subset \mathbb{R}$, 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 $\mathcal{H}$ is a pseudo-metric space of functions. The model considered is

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

where $m : \mathcal{Z} \times \mathcal{H} \to \mathbb{R}^l$ 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 two example illustrate.

Example 1 *(Instrumental variables smooth coefficients model) Let*

$$Y_t = X_t^\tau h_0\left(U_t\right) + \varepsilon_t,$$

where $E\left(\varepsilon_t|X_t, U_t\right) \neq 0$, *and assume that there exists a vector of instruments* $W_t \in \mathbb{R}^l$ $(l \geq k)$ *such that* $E\left(\varepsilon_t|W_t, U_t\right) = 0$. *The law of iterated expectations implies that*

$$E\left[W_t\left(Y_t - X_t^\tau h\left(U_t\right)\right)|U_t\right] = \underline{0} \; a.s. \; for \; a \; unique \; h = h_0,$$

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

$$m\left(Z_t, h\left(U_t\right)\right) = W_t\left(Y_t - X_t^\tau h\left(U_t\right)\right).$$

Example 2 *(Nonparametric stochastic discount factor model) Let* $R_{j,t}$ $(j = 1, ..., J)$ *denote the (excess) returns of* J *risky assets and* $R_{M,t}$ *denote the (excess) market return. Following Wang (2003) and Wang (2009), let* $h_{t+1} = 1 - h_0\left(U_t\right) R_{M,t+1}$ *denote the so-called nonparametric pricing kernel that satisfies* $E\left[h_{t+1} \otimes R_{t+1}|X_t, U_t\right] = 0$ *a.s., where* $R_t = \left[R_{1,t}, ..., R_{J,t}\right]^\tau$ *and* X_t *denote a* *set of additional conditioning variables. The law of iterated expectations implies that, for any vector of functions* $q : \mathcal{X} \to \mathbb{R}^l$

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

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

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

Throughout the rest of the paper, it is assumed that, at a given point $U_t = u_0$, h_0 can be polynomially approximated by

$$h_0\left(U_t\right) = \sum_{j=0}^{p} \frac{1}{j!} h_0^{\prime ... j}\left(u_0\right)\left(U_t - u_0\right)^j := \sum_{j=1}^{p+1} \frac{1}{(j-1)!} h_j^{\prime ... j}\left(U_t - u_0\right)^{j-1}. \tag{2}$$

Thus, for $U_t \approx u_0$, (2) implies that (1) must also be approximately zero, that is

$$E\left[m\left(Z_t, \underline{h}^\tau B_{p+1}\left[1, ..., \left(\frac{U_t - u_0}{b}\right)^p\right]^\tau\right)|U_t = u_0\right] \approx \underline{0}, \tag{3}$$

where $\underline{h} = \left[h_1^\tau, ..., h_{p+1}^\tau\right]^\tau$ and $B_{p+1} = diag\left[1/0!, ..., b^p/p!\right]$, which forms the basis for the local polynomial estimation proposed in this paper, and represents a generalization of the localized moment restriction used by **?** for the smooth coefficients model they considered.

It is important to note that unless the dimension $\dim(m) = l$ of the estimating equations m is such that $l \geq \sum_{j=1}^{p+1} \dim\left(h_j\right)$, the unknown parameters $h_0^{\prime ... p}\left(u_0\right)$ cannot be consistently estimated, a well-known fact in both the econometric and statistical estimation theory. Therefore, to achieve consistency, the following augmented general estimating equation is considered

$$g_t\left(\underline{h}\right) = \begin{bmatrix} 1 & \frac{U_t - u_0}{b} & ... & \left(\frac{U_t - u_0}{b}\right)^p \end{bmatrix}^\tau \otimes m\left(Z_t, \underline{h}^\tau B_{p+1}\left[1, ..., \left(\frac{U_t - u_0}{b}\right)^p\right]^\tau\right).$$

To incorporate the localized nature of restriction (3), let $b =: b(T)$ denote the size of the local neighborhood where the polynomial approximation (2) is valid - the bandwidth - and let the kernel function $K : \mathcal{U} \to \mathbb{R}$ denote a symmetric probability density function. Then the localized sample analog of (3) is given by

$$\frac{1}{Tb} \sum_{t=1}^{T} g_t\left(\underline{h}\right) K\left(\frac{U_t - u_0}{b}\right) := \widehat{g}^K\left(\underline{h}, u_0\right), \tag{4}$$

where $g^K\left(\underline{h}, u_0\right) = g_t\left(\underline{h}\right) K\left(\left(U_t - u_0\right)/b\right)$ and for a generic function f_t, $\widehat{f_t} := \sum_{t=1}^{T} f_t/Tb$; the LPGMM estimator then is defined as

$$\widehat{\underline{h}} = \arg\min_{\underline{h} \in \mathcal{H}_C} \widehat{g}^K\left(\underline{h}, u_0\right)^\tau \widehat{W}_{p+1}\left(u_0\right) \widehat{g}^K\left(\underline{h}, u_0\right), \tag{5}$$

where the parameter space $\mathcal{H}_C$ is specified in the next section, and $\widehat{W}_{p+1}(u_0)$ is a, possibly random, $\mathbb{R}^{(p+1)l\times(p+1)l}$ valued positive semi definite matrix- see the discussion after Theorem (2) for an example. The LPGEL estimator is defined as the solution to the saddlepoint problem

$$\widehat{\underline{h}}^{\rho} = \arg\min_{\underline{h}\in\mathcal{H}_C} \sup_{\lambda\in\Lambda_T(\underline{h})} \widehat{\rho}\left(v^K\left(\lambda, \underline{h}, u_0\right)\right), \tag{6}$$

where $\widehat{\rho}\left(v^K\left(\lambda, \underline{h}, u_0\right)\right) =: \sum_{t=1}^{T} \rho\left(v_t^K\left(\lambda, \underline{h}, u_0\right)\right)/Tb$, $v_t^K\left(\lambda, \underline{h}, u_0\right) = \lambda^{\tau} g_t^K\left(\underline{h}, u_0\right)$, $\lambda =: \lambda(u_0)$, ρ is a a concave function on its domain, an open set Λ_0 containing 0, $\Lambda_T\left(\underline{h}\right) = \left\{\lambda : v_t^K\left(\lambda, \underline{h}, u_0\right) \in \Lambda_0\right\}$ for $t = 1, ..., T$, and $\overline{\underline{h}}$ are fixed values of $\underline{h}$. For example, the local polynomial version of the empirical likelihood (LPEL) estimator is

$$\widehat{\underline{h}}^{el} = \arg\min_{\underline{h}\in\mathcal{H}_C} \sup_{\lambda\in\Lambda_T(\underline{h})} \frac{1}{Tb}\sum_{t=1}^{T} \log\left(1 - v_t^K\left(\lambda, \underline{h}, u_0\right)\right). \tag{7}$$

Remark 1 *In practice the LPGEL estimator can be computed iteratively as follows:*

Step 1. For an arbitrary fixed $\overline{\underline{h}}$, *compute* $\widehat{\lambda}^{(1)} = \arg\max_{\lambda\in\Lambda_T(\underline{h})} \widehat{\rho}\left(v^K\left(\lambda, \overline{\underline{h}}, u_0\right)\right)$.

Step 2. Compute $\widehat{\underline{h}}^{\rho(1)} = \arg\min_{\underline{h}\in\mathcal{H}_C} \widehat{\rho}\left(v^K\left(\widehat{\lambda}^{(1)}, \underline{h}, u_0\right)\right)$.

Step 3. Iterate Steps 1 and 2 until a specified degree of convergence is achieved for both $\widehat{\lambda}$ *and* $\widehat{\underline{h}}^{\rho}$. *The convergence of both estimators is guaranteed by the concavity of* ρ *in* λ *and Assumption A2 below on the parameter space* $\mathcal{H}_C$.

3 Asymptotic results

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

$$\Omega_{p+1}(u_0) = \int v^{i+j} K^2(v)\, dv \otimes E\left(m_t\left(h_0\left(U_t\right)\right)^{\otimes 2} | U_t = u_0\right) f(u_0),$$
$$G_{p+1}(u_0) = \int v^{j} K(v)\, dv \otimes E\left(\partial m_t\left(h_0\left(U_t\right)\right)/\partial h^{\tau} | U_t = u_0\right) f(u_0)$$

for $(0 \le i, j \le p)$, where f is the marginal density of U_t at u_0, and assume that:

A1 The process $\{Z_t^{\tau}, U_t\}_{t\in\mathbb{Z}}^{\tau}$ is strictly stationary α mixing, with mixing coefficient $\alpha(t) = O\left(t^{-a}\right)$ with $a = (2+\delta)(1+\delta)/\delta$ for some $\delta > 0$,

A2 (i) There exists a unique h_0 such that $E\left[m_t\left(h_0\left(U_t\right)\right) | U_t = u_0\right] = \underline{0}$, (ii) $\left[h_0^{\tau}, ..., \left(h_0^{\prime\ldots p}\right)^{\tau}\right]^{\tau} \in int\left(\mathcal{H}_C\right)$, where $\mathcal{H}_C$ is a compact subset of $\mathbb{R}^{(p+1)k}$, (iii) h_0 is $(p+1)$ continuously differentiable at u_0,

A3 (i) $\partial m_t(h)/\partial h^{\tau}$ exists and is continuous at u_0 for each $h \in \mathcal{H}_C$ *a.s.*, (ii) the functions $E\left[m_t\left(h_0\right) | U_t = u_0\right]$, $E\left[m_t\left(h_0\right)^{\otimes 2} | U_t = u_0\right]$ and $E\left[\partial m_t\left(h_0\right)/\partial h^{\tau} | U_t = u_0\right]$ are continuous,

(iii) $E\left(\|g_t(\underline{h})\|^{2(1+\delta)} U_t = u\right)$, $E\left(\left\|g_t(\underline{h})^{\otimes 2}\right\| U_t = u\right)$ and $E\left(\|\partial g_t(\underline{h})/\partial \underline{h}^\tau\| U_t = u\right) < \infty$ uniformly in $\underline{h} \in \mathcal{H}_C$ and for all u in a neighborhood of u_0, (iv) for $t \geq 2$

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

for all u and v in a neighborhood of u_0, (v) $G_{p+1}(u_0)$ has rank $(p+1)k$, $\Omega_{p+1}(u_0)$ is positive definite and the matrices $\Sigma_{(p+1)W}(u_0) = G_{p+1}(u_0)^\tau W_{p+1}(u_0) G_{p+1}(u_0)$ and $\Sigma_{(p+1)\Omega^{-1}}(u_0) = G_{p+1}(u_0)^\tau \Omega_{p+1}(u_0)^{-1} G_{p+1}(u_0)$ are nonsingular,

A4 (i) the kernel function $K : \mathcal{U} \to \mathbb{R}$ has a compact support, say $[-1, 1]$, (ii) the marginal density f of U_t is twice continuously differentiable at $U_t = u_0$ and strictly positive at $U_t = u_0$, (iii) the joint density $f_{1,t}$ of U_1 and U_t for $t \geq 2$ is continuous at u_0, (v) for δ given in A1, the bandwidth b satisfies $b \to 0$ and $Tb^{1+2/(1+\delta)} \to \infty$ as $T \to \infty$,

A5 $\rho\left(v_t^K(\lambda, \underline{h}, u_0)\right)$ is twice continuously differentiable in $v_t(\cdot)$ 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 for all u in a neighborhood of u_0, (i) $\widehat{W}_{p+1}(u) \xrightarrow{p} W_{p+1}(u)$, where $W_{p+1}(u)$ is a positive definite matrix, (ii)

$$\begin{aligned}
&\text{(a)} \quad \left\|\left(\widehat{h}_1 - h_0\right)^\tau, ..., \left(\widehat{h}_{p+1} - h_0^{\prime...p}\right)^\tau\right\| \xrightarrow{p} 0,\\
&\text{(b)} \quad \left\|\widehat{\lambda}\right\| \xrightarrow{p} 0, \ \left\|\left(\widehat{h}_1^\rho - h_0\right)^\tau, ..., \left(\widehat{h}_{p+1}^\rho - h_0^{\prime...p}\right)^\tau\right\| \xrightarrow{p} 0,
\end{aligned}$$

(iii) $\max_{t\leq T}\left|\widehat{\lambda}^\tau g_t^K\left(\underline{\widehat{h}}^\rho, u_0\right)\right| = o_p(1)$.

Assumption A1 specifies the dependence structure of the process $\{Z_t^\tau, U_t\}_{t\in\mathbb{Z}}^\tau$ as α mixing with a rate of decay on the mixing coefficient $\alpha(t)$ that is standard in the literature on nonparametric models for time series, see for example Cai, Fan and Yao (2000). Assumption A2(i) is a standard identification assumption 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, for the instrumental variables smooth coefficients model (1), A2(i) is implied by the condition $rank\left(E\left(W_t X_t^\tau|U_t = u_0\right)\right) = k$. The compactness assumption A2(ii) is as in Lewbel (2007), but can be replaced by other assumptions to control for the complexity of the pseudo metric space $\mathcal{H}$ - typically expressed in terms of covering or bracketing numbers (see Van der Vaart and Wellner (1996) for a definition), such as when $\mathcal{H}$ is a Holder or a Sobolev space. Assumptions A3-A5 are standard, respectively, in nonparametric estimation and GEL estimation, see, for example, Masry (1996) and Newey and Smith (2004), respectively. Assumption A6(ii)(a) is a high level assumption, which can be shown using the uniform law of large numbers combined with the slightly stronger (than A2(i)) identification condition that for any $\xi > 0$, there exists a $c(\xi) > 0$, such that for all u in a neighborhood of u_0

$$\inf_{\substack{\underline{h} \in \mathcal{H}_C \\ \left\| (h_1 - h_0)^\tau, \ldots, \left(h_{p+1} - h_0^{\prime \ldots p}\right)^\tau \right\| \geq \xi}} E\left[\left\|\widehat{g}^K\left(\underline{h}, u\right)\right\|\right] > c(\xi). \tag{8}$$

Assumption A6(ii)(b) is also a high level one, and can be verified using the same arguments used by Newey and Smith (2004), which consist of first assuming the existence of a consistent LPGEL estimator for $\underline{h}$, say $\underline{\widetilde{h}}^\rho$, such that $\left\|\widehat{g}^K\left(\underline{\widetilde{h}}^\rho, u\right)\right\| = O_p\left(B_{p+1}/(Tb)^{1/2} + b^{p+j}\right)$ ($j = 1$ for j odd, or $j = 2$ for j even), from which the consistency of the corresponding estimator $\widetilde{\lambda} := \arg\max_{\lambda \in \Lambda_T(\underline{h})} \widehat{\rho}\left(v^K\left(\lambda, \underline{\widetilde{h}}^\rho, u\right)\right)$ follows by a Taylor expansion combined with the uniform law of large numbers; next, the consistency of $\underline{\widehat{h}}^\rho$ can by shown again by the uniform law of large numbers combined with (8). Finally, a saddlepoint argument can be used to show that $\left\|\widehat{g}^K\left(\underline{\widehat{h}}^\rho, u\right)\right\| = O_p\left(B_{p+1}/(Tb)^{1/2} + b^{p+j}\right)$, which in turn implies the consistency of $\widehat{\lambda}$.

The following theorems establish the asymptotic distributions of the LPGMM and LLGEL estimators.

Theorem 1 *Under A1-A4 and A6(ii)(a), as* $(Tb)^{1/2} \to \infty$,

$$(Tb)^{1/2} B_{p+1} \left[\begin{array}{c} \widehat{h}_1 - h_0 \\ \vdots \\ \widehat{h}_{p+1} - h_0^{\prime \ldots p} \end{array} - b(u_0)\right] \xrightarrow{d}$$
$$N\left(\underline{0}, \Sigma_{p+1W}^{-1}(u_0) G_{p+1}(u_0)^\tau W_{p+1}(u_0) \Omega_{p+1}(u_0) W_{p+1}(u_0) G_{p+1}(u_0) \Sigma_{p+1W}^{-1}(u_0)\right),$$

where

$$b(u_0) = -\left[\begin{array}{ccc} \int v^{p+1} K(v)\, dv & \cdots & \int v^{2p+1} K(v)\, dv \end{array}\right]^\tau \otimes$$
$$f(u_0) E\left[\frac{\partial m_t(h_0(U_t))}{\partial h^\tau} | U_t = u_0\right] \frac{b^{p+1} h_0^{\prime \ldots p+1}}{(p+1)!}$$

for p odd and

$$b(u_0) = -\left[\begin{array}{ccc} \int v^{p+2} K(v)\, dv & \cdots & \int v^{2p+2} K(v)\, dv \end{array}\right] \otimes$$
$$f(u_0) E\left(\frac{\partial m_t(h_0(U_t))}{\partial h^\tau} | U_t = u_0\right) b^{p+2} \left(\frac{h_0^{\prime \ldots p+1} f'(u_0)}{(p+1)! f(u_0)} + \frac{h_0^{\prime \ldots p+2}}{(p+2)!}\right)$$

for p even.

Theorem 2 *Under A1-A5, A6(ii)b and (iii), as* $(Tb)^{1/2} \to \infty$,

$$(Tb)^{1/2} B_{p+1} \left[\begin{array}{c} \widehat{h}_1^\rho - h_0 \\ \vdots \\ \widehat{h}_{p+1}^\rho - h_0^{\prime \ldots p} \end{array} - b(u_0)\right] \xrightarrow{d} N\left(\underline{0}, \Sigma_{p+1\Omega^{-1}}^{-1}(u_0)\right),$$

where $b(u_0)$ is as defined in Theorem 1.

Theorems 1 and 2 show that both the LPGMM and LPGEL estimators are characterized by the same asymptotic bias. On the other hand, it is easy to show that the asymptotic covariance matrix of the LPGEL estimator is the smallest (in the matrix sense) for any LPGMM estimators based on $\widehat{W}_{p+1}(u_0) \neq \Omega_{p+1}(u_0)^{-1} + o_p(1)$. Thus $\Sigma^{-1}_{p+1\Omega^{-1}}(u_0)$ represents the lower bound (minimum asymptotic covariance matrix) for the class of LPGMM estimators indexed by $W_{p+1}(u_0)$, which can be achieved if the classical two-step GMM estimation procedure is used, where the first-step is used to obtain preliminary consistent estimators, say $\widetilde{h}_1$,..., $\widetilde{h}_{p+1}$ to be used in the second-step estimation to compute $\widehat{W}_{p+1}(u_0) = \widetilde{\Omega}_{p+1}(u_0)^{-1}$, to achieve the lower bound $\Sigma^{-1}_{p+1\Omega^{-1}}(u_0)$.

Theorems 1 and 2 also show that the proposed estimators have a non negligible asymptotic bias $b(u_0)$, as it is typical of any kernel based estimator. To make them operational for inferential purposes, $b(u_0)$ can be either estimated, using the local polynomial approach of this paper, or, alternatively, under the additional undersmoothing condition $Tb^5 \to 0$, can be eliminated altogether.

References

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.

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

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

Fan, J. and Gijbels, I.: 1996, *Local Polynomial Modeling and its Applications*, Chapman and Hall.

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

Hall, P. and Heyde, C.: 1980, *Martingale limit theory and its application*, Academic Press, New York.

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

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.

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

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

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.

Wang, K.: 2003, Asset pricing with conditioning information: A new test, *Journal of Finance* **58**, 161–196.

Wang, K.: 2009, Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio, *Journal of Empirical Finance* **9**, 133–169.

4 Appendix

Throughout this Appendix let $\underline{h}_0 = \left[h_0^{\tau}, ..., h_0^{'...p\tau}\right]^{\tau}$, $\partial^{p+1}(\cdot) = \partial(\cdot)/\underline{h}$, and "CMT", "CLT", "LLN" and "$w.p. \to 1$" denote Continuous Mapping Theorem, Central Limit Theorem, (possibly uniform) law of large numbers, and with probability approaching 1.

Proof of Theorem 1. By A2(ii), the first order conditions $\underline{0} = \partial^{p+1}\widehat{g}^K\left(\widehat{\underline{h}}, u_0\right)^{\tau}\widehat{W}_{p+1}(u_0)\,\widehat{g}^K\left(\widehat{\underline{h}}, u_0\right)$ are satisfied $w.p. \to 1$. Then, by a mean value expansion

$$\underline{0} = \partial^{p+1}\widehat{g}^K\left(\widehat{\underline{h}}, u_0\right)^{\tau}\widehat{W}_{p+1}(u_0)\left[\widehat{g}^K\left(\underline{h}_0, u_0\right) + \partial^{p+1}\widehat{g}^K\left(\overline{\underline{h}}, u_0\right)^{\tau}(Tb)^{1/2}B_{p+1}\left(\widehat{\underline{h}} - \underline{h}_0\right)^{\tau}\right],$$

where $\overline{\underline{h}}$ is on the line segments between $\underline{h}_0$ and $\widehat{\underline{h}}$. By A2(ii) the class of functions $\mathcal{G}_{\partial}^{K} = \left\{\partial^{p+1}g^K\left(\underline{h}\right), \underline{h} \in \mathcal{H}_C, u\right\}$, for all u in a neighborhood of u_0, is Euclidean (see **?** for a definition) since it's the pointwise multiplication of the two Euclidean classes of functions $\mathcal{G}_{\partial} =$

$\{\partial^{p+1} g(\underline{h}), \underline{h} \in \mathcal{H}_C\}$ and $\mathcal{G}^K = \{K(v-u)/b, u\}$, hence by the envelope assumption A3(iii), the LLN implies that $\left\|\partial^{p+1}\widehat{g}^K\left(\widehat{\underline{h}}, u_0\right) - G_{p+1}(u_0)\right\| = o_p(1)$, hence

$$(Tb)^{1/2} B_{p+1}\left[\left(\widehat{\underline{h}} - \underline{h}_0\right)^\tau\right]^\tau = \widehat{\Sigma}^{-1}_{p+1\widehat{W}}(u_0)\,\partial^{p+1}\widehat{g}^K\left(\widehat{h}, u_0\right)^\tau \widehat{W}_{p+1}(u_0)\,(Tb)^{1/2}\,\widehat{g}^K(\underline{h}_0, u_0)$$

where $\widehat{\Sigma}_{p+1\widehat{W}}(u_0) = \partial^{p+1}\widehat{g}^K\left(\widehat{\underline{h}}, u_0\right)^\tau \widehat{W}_{p+1}(u_0)\,\partial^{p+1}\widehat{g}^K\left(\overline{\underline{h}}, u_0\right)$, hence by A3(v) and CMT

$$(Tb)^{1/2} B_{p+1}\left[\left(\widehat{\underline{h}} - \underline{h}_0\right)^\tau\right]^\tau = \Sigma^{-1}_{p+1W}(u_0)\,G_{p+1}(u_0)^\tau\,W(u(_0))\,(Tb)^{1/2}\,\widehat{g}^K(\underline{h}_0, u_0) + o_p(1).$$

Next, we prove the asymptotic normality of $(Tb)^{1/2}\left(\widehat{g}^K(\underline{h}_0, u_0) - E\left(\widehat{g}^K(\underline{h}_0, u_0)\right)\right)$; the proof is similar to that of Cai et al. (2000), so we just sketch it. First notice that

$$\begin{aligned} Var\left((Tb)^{1/2}\,\widehat{g}^K(\underline{h}_0, u_0)\right) &= Var\left(\frac{1}{b^{1/2}} g_t^K(\underline{h}_0, u_0)\right) + \\ \frac{2}{T}\sum_{t=1}^{T-1}(T-t)\,Cov\left(\frac{g_1^K}{b}(\underline{h}_0, u_0), g_{t+1}^K(\underline{h}_0, u_0)\right) &:= A_{1T} + A_{2T}, \end{aligned}$$

and

$$\|A_{2T}\| \le \sum_{t=1}^{d_T}\left\|Cov\left(\frac{g_1^K}{b}(\underline{h}_0, u_0), g_{t+1}^K(\underline{h}_0, u_0)\right)\right\| + \sum_{t=d_T+1}^{T}\left\|Cov\left(\frac{g_1^K}{b}(\underline{h}_0, u_0), g_{t+1}^K(\underline{h}_0, u_0)\right)\right\| := A_{21T} + A_{22T},$$

for a sequence of positive integers $d_T \to \infty$ as $T \to \infty$, such that $d_T b \to 0$. Then by A3(iii) $A_{21T} \le O(bd_T) \to 0$, while $A_{22T} \le Cb\sum_{t=d_T}^{T}\alpha(t)^{\frac{\delta}{2+\delta}}\,b^{-2(1+\delta)/(2+\delta)} = o(1)$ by Davydov's inequality (Hall and Heyde 1980)[Corollary A.2] and $d_T^{2+\delta} b = O(1)$, hence by a standard kernel calculation

$$A_{1T} = E\left[\frac{1}{b}\left(g_t^K(\underline{h}_0, u_0)\right)^{\otimes 2}\right] - \left[E\frac{1}{b}\left(g_t^K(\underline{h}_0, u_0)\right)\right]^{\otimes 2} = \Omega_{p+1}(u_0) + o(1).$$

Let

$$\frac{1}{(Tb)^{1/2}}\sum_{t=1}^{T}\theta^\tau\left(g_t^K(\underline{h}_0, u_0) - E\left(g_t^K(\underline{h}_0, u_0)\right)\right) := (Tb)^{-1/2}\sum_{t=1}^{T} D_t$$

for any $\theta \in \mathbb{R}^{(p+1)l}$ such that $\|\theta\| = 1$, and note that $Var\left((Tb)^{-1/2}\sum_{t=1}^{T} D_t\right) = \omega(u_0) + o(1)$, where $\omega(u_0) = \theta^\tau \Omega_{p+1}(u_0)\,\theta$. Partition $\{1, ..., T\}$ in $2q_T + 1$ subsets, where $q_T = \lfloor T/(r_T + s_T)\rfloor$ with $r_T = \left\lfloor (Tb)^{1/2}\right\rfloor$ and $s_n = \left\lfloor (Tb)^{1/2}/\log T\right\rfloor$, where $\lfloor\cdot\rfloor$ is the integer part function, and define

$$A_{31jT} = \sum_{t=j(r_T+s_T)+1}^{j(r_T+s_T)+r_T} D_t, \quad A_{32jT} = \sum_{t=j(r_T+s_T)+r_T+1}^{(j+1)(r_T+s_T)} D_t, \quad A_{33T} = \sum_{t=q_T(r_T+s_T)+1}^{T} D_t,$$

so that

$$\frac{1}{(Tb)^{1/2}}\sum_{t=1}^{T}D_t=\frac{1}{(Tb)^{1/2}}\left(\sum_{j=1}^{q_T}A_{31jT}+\sum_{j=1}^{q_T}A_{32jT}+A_{33T}\right):=A_{41T}+A_{42T}+A_{43T}.$$

Then, it is possible to show (see Cai et al. (2000) for details) that $E\left(A_{4lT}^2/T\right)\to 0$ $(l=2,3)$, $\left|E\left[\exp\left(\iota\tau A_{41T}\right)\right]-\prod_{j=1}^{q_T}E\left[\exp\left(\iota\tau A_{31jT}\right)\right]\right|\to 0$, $\sum_{j=1}^{q_T-1}E\left(A_{31jT}^2/T\right)\to\omega\left(u_0\right)$ and $\sum_{j=1}^{q_T-1}E\left(A_{31jT}^2\right)I\left(\left|A_{31jT}\right|\geq\varepsilon\omega\left(u_0\right)(Tb)^{1/2}\right)/(Tb)^{1/2}\to 0$ for any $\varepsilon>0$. Thus, by the Lindeberg-Feller CLT, $(Tb)^{-1/2}A_{41T}\xrightarrow{d}N\left(0,\omega\left(u_0\right)\right)$ and the conclusion follows by the Cramer-Wold device. Finally, we find an explicit expression for the asymptotic bias $E\left(\widehat{g}^K\left(\underline{h}_0,u_0\right)/b\right)$. For p odd, in a neighborhood of $\left|U_t-u_0\right|<b$, (2) and A2(iii) imply

$$h_0\left(U_t\right)=\sum_{j=1}^{p+1}\frac{1}{(j-1)!}h_j^{\prime\ldots j}\left(U_t-u_0\right)^j+o_p\left(b^{p+1}\right),$$

hence

$$\begin{aligned}
&E\left[\frac{g_t^K\left(\underline{h}_0,u_0\right)}{b}\right]=\int\left[\begin{array}{cccc}1 & \frac{U_t-u}{b} & \cdots & \left(\frac{U_t-u}{b}\right)^p\end{array}\right]^\tau\otimes\\
&E\left[m_t\left(h_0\left(U_t\right)\right)-\frac{\partial m_t\left(h_0\left(U_t\right)\right)}{\partial h^\tau}\frac{b^{p+1}h_0^{\prime\ldots p+1}}{(p+1)!}\left(\frac{U_t-u_0}{b}\right)^{(p+1)}|U_t\right]\times\\
&K\left(\frac{U_t-u_0}{b}\right)f\left(U_t\right)dU_t\\
&=-\left[\begin{array}{cccc}\int v^{p+1}K(v)\,dv & \int v^{p+2}K(v)\,dv & \cdots & \int v^{2p+1}K(v)\,dv\end{array}\right]^\tau\otimes\\
&E\left[\frac{\partial m_t\left(h_0\left(U_t\right)\right)}{\partial h^\tau}|U_t=u_0+vb\right]\left(f\left(u_0\right)+O(b)\right)\frac{b^{p+1}h_0^{\prime\ldots p+1}}{(p+1)!}
\end{aligned}$$

and the conclusion follows by the symmetry of $K(\cdot)$. For p even

$$E\left[\frac{g_t^K(\underline{h}_0,u_0)}{b}\right]=\int\left[\begin{array}{cccc}1 & \frac{U_t-u_0}{b} & \cdots & \left(\frac{U_t-u_0}{b}\right)^p\end{array}\right]^{\tau}\otimes$$

$$E\left[m_t(h_0(U_t))-\frac{\partial m_t(h_0(U_t))}{\partial h^{\tau}}\left(\frac{b^{p+1}h_0^{\prime\ldots p}}{(p+1)!}\left(\frac{U_t-u_0}{b}\right)^{(p+1)}+\right.\right.$$

$$\left.\left.\frac{b^{p+2}h_0^{\prime\ldots p}}{(p+2)!}\left(\frac{U_t-u_0}{b}\right)^{(p+2)}|U_t\right]K\left(\frac{U_t-u_0}{b}\right)f(U_t)\,dU_t$$

$$=\int\left[\begin{array}{cccc}1 & v & \cdots & v^p\end{array}\right]^{\tau}\otimes E\left[-\frac{\partial m_t(h_0(U_t))}{\partial h^{\tau}}\left(\frac{b^{p+1}h_0^{\prime\ldots(p+1)}}{(p+1)!}v^{(p+1)}+\right.\right.$$

$$\left.\left.\frac{b^{p+2}h_0^{\prime\ldots(p+2)}}{(p+2)!}v^{(p+2)}\right)|U_t=u_0+vb\right]K(v)\left(f(u_0)+f'(u_0)\,vb+\frac{f''(u_0)}{2}(vb)^2\right)dv$$

$$=-b^{p+2}\left[\begin{array}{cccc}\int v^{p+2}K(v)\,dv & \cdots & \int v^{2p+1}K(v)\,dv & \int v^{2p+2}K(v)\,dv\end{array}\right]\otimes E\left[\frac{\partial m_t(h_0(U_t))}{\partial h^{\tau}}|U_t=u_0\right]$$

$$\left(\frac{h_0^{\prime\ldots p+1}f'(u_0)}{(p+1)!f(u_0)}+\frac{h_0^{\prime\ldots p+2}}{(p+2)!}\right)\left(f(u_0)+O\left(b^2\right)\right)$$

and the conclusion follows again by the symmetry of $K(\cdot)$ ■

Proof of Theorem 2 . By A2(ii) the first order conditions

$$\underline{0}=\left[\partial\widehat{\rho}\left(v^K\left(\widehat{\lambda},\widehat{\underline{h}},u_0\right)\right)/\partial\lambda^{\tau},\partial^{p+1}\widehat{\rho}\left(v^K\left(\widehat{\lambda},\widehat{\underline{h}},u_0\right)\right)\right]^{\tau}$$

are satisfied $w.p.\to 1$. Then by a mean value expansion

$$\underline{0}=\left[\begin{array}{c}\frac{\partial\widehat{\rho}\left(v^K(0,\underline{h}_0,u_0)\right)}{\partial\lambda}\\ \partial^{p+1}\rho\left(v^K(0,\underline{h}_0,u_0)\right)\end{array}\right]+\left[\begin{array}{cc}\frac{\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)}{(\partial\lambda)^{\otimes 2}} & \frac{\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)}{\partial\lambda\partial h^{\tau}}\\ \frac{\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)}{\partial h\partial\lambda^{\tau}} & \frac{\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)}{(\partial h)^{\otimes 2}}\end{array}\right](Tb)^{1/2}B_{p+1}\left(\widehat{\underline{h}}^{\rho}-\underline{h}_0\right),$$

where $\overline{\lambda}$ is on the line segment between 0 and $\widehat{\lambda}$ and $\overline{\underline{h}}^{\rho}$ is on line segments between $\underline{h}_0$ and $\widehat{\underline{h}}^{\rho}$. Since

$$\frac{\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)}{(\partial\lambda)^{\otimes 2}}=\frac{1}{Tb}\sum_{t=1}^{T}\rho_2\left(v_t^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)\left(g_t^K\left(\overline{h}^{\rho},u_0\right)\right)^{\otimes 2},$$

Assumption A6(iii) and CMT imply that $\max_{t\le T}\left|\rho_j\left(v_t^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)+1\right|=o_p(1)$ $(j=1,2)$, hence the triangle inequality and the same arguments used in the proof of Theorem 1 for the Euclidean class of functions $\mathcal{G}_{g^2}=\left\{g^K(h)^{\otimes 2},h\in\mathcal{H}_C,u\right\}$ show that

$$\left\|\frac{\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)\right)}{(\partial\lambda)^{\otimes 2}}+\Omega_{p+1}(u_0)\right\|\le\max_{t\le T}\left|\rho_2 v_t^K\left(\overline{\lambda},\overline{\underline{h}}^{\rho},u_0\right)+1\right|\times$$

$$\left\|\frac{1}{Tb}\sum_{t=1}^{T}\left(g_t^K\left(\overline{\underline{h}}^{\rho},u_0\right)\right)^{\otimes 2}\right\|+\left\|\frac{1}{Tb}\sum_{t=1}^{T}\left(g_t^K\left(\overline{\underline{h}}^{\rho},u_0\right)\right)^{\otimes 2}+\Omega_{p+1}(u_0)\right\|\xrightarrow{p}0$$

and similarly for $\left\|\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^\rho,u_0\right)\right)/\partial\lambda\partial h^\tau+G_{p+1}\left(u_0\right)\right\|=o_p\left(1\right)$ and $\left\|\partial^2\widehat{\rho}\left(v^K\left(\overline{\lambda},\overline{\underline{h}}^\rho,u_0\right)\right)^{\otimes 2}\right\|=o_p\left(1\right)$. Thus by CMT

$$(Tb)^{1/2}\left[I,B_{p+1}\right]\left[\widehat{\lambda}^\tau,\left(\widehat{h}^\rho-\underline{h}_0\right)^\tau\right]^\tau=-\left[\begin{array}{ll}\Omega_{p+1}\left(u_0\right) & G_{p+1}\left(u_0\right)\\ G_{p+1}\left(u_0\right)^\tau & O\end{array}\right]^{-1}\left[\begin{array}{c}(Tb)^{1/2}\,\widehat{g}^K\left(\underline{h}_0,u_0\right)\\ \underline{0}\end{array}\right]+o_p\left(1\right)$$

and the conclusion follows as in the proof of Theorem 1. ■