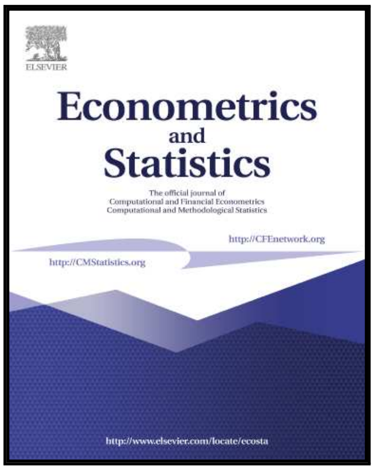
Accepted Manuscript



Generalized Empirical Likelihood M Testing for Semiparametric Models with Time Series Data

Francesco Bravo, Ba M. Chu, David T. Jacho-Chávez

PII: S2452-3062(17)30002-3
DOI: 10.1016/j.ecosta.2016.12.004
Reference: ECOSTA 37

To appear in: *Econometrics and Statistics*

Received date: 19 March 2016
Revised date: 20 December 2016
Accepted date: 20 December 2016

Please cite this article as: Francesco Bravo, Ba M. Chu, David T. Jacho-Chávez, Generalized Empirical Likelihood M Testing for Semiparametric Models with Time Series Data, *Econometrics and Statistics* (2017), doi: 10.1016/j.ecosta.2016.12.004

Generalized Empirical Likelihood M Testing for Semiparametric Models with Time Series Data

Francesco Bravo*
University of York

Ba M. Chu†
Carleton University

David T. Jacho-Chávez‡
Emory University

December 2016

Abstract

The problem of testing for the correct specification of semiparametric models with time series data is considered. Two general classes of M test statistics that are based on the generalized empirical likelihood method are proposed. A test for omitted covariates in a semiparametric time series regression model is then used to showcase the results. Monte Carlo experiments show that the tests have reasonable size and power properties in finite samples. An application to the demand of electricity in Ontario (Canada) illustrates their usefulness in practice.

Keywords: α-Mixing; Instrumental variables; Kernel Smoothing; Stochastic Equicontinuity

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

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

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

1 Introduction

In this paper we consider testing for the correct specification of smooth semiparametric models with time series observations. The tests we propose here are important generalizations of the so-called M tests originally proposed by Newey (1985) (see White 1994 for a review and some applications to parametric models) and commonly used in empirical work. The basic idea behind M testing is to use a set of restrictions, expressed in the form of a set of estimating equations, as indicators of the correct specification of the statistical model under investigation. For example in linear time series regression models the orthogonality between a set of possibly irrelevant regressors and the errors gives rise to a natural estimating equation that can be used to test for the irrelevance of those regressors. White's (1982) information matrix test based on a conditional likelihood, which can be used to test for the correct specification of dynamic parametric models, and Newey's (1985) conditional moment tests based on a set of unconditional estimating equations are other examples.

The M tests we consider are cast into Newey and Smith's (2004) Generalized Empirical Likelihood (GEL henceforth) framework, in which a preliminary estimator for the infinite dimensional parameter and a generic (possibly also semiparametric) estimator for the finite dimensional parameter are available. GEL provides a natural framework for estimating and/or obtaining inferences in statistical models defined by a set of estimating equations. Examples of GEL include Owen's (1988) well-known Empirical Likelihood (EL henceforth) and Kitamura and Stutzer's (1997) Exponential Tilting.

We assume that the observations are α-mixing (see Doukhan 1994 for a review of the statistical properties of α-mixing processes) and develop a rather general theory of M testing that can be applied to various semiparametric possibly nonlinear statistical models (see Gao 2007 for some examples). The results of this paper generalize and extend results of Bertail (2006), Bravo (2009), Hjort, McKeague, and van Keilegom (2009) and many others on EL inferences for semiparametric models with independent and identically distributed observations. The new results are the following: First we use the same kernel based smoothing used by Smith (1997) and Kitamura and Stutzer (1997) and propose two general types of test statistics, one based on an appropriately corrected GEL criterion function and one based on a Lagrange Multiplier (LM henceforth) approach. We show that both the GEL and LM statistics are asymptotically distribution free under the null hypothesis of correct specification and have power against the hypotheses of local and global misspecification. We note that smoothing is necessary to obtain an asymptotic distribution free GEL statistic, but it is also useful to obtain consistent estimators of the long run variances used in the LM statistic. Second, we explicitly consider the case where the estimation of the infinite dimensional parameter might affect the asymptotic properties of the proposed GEL and LM statistics and provide a general formula to characterize it. The characterization is based on the pathwise derivative as in Newey (1994) and relies on a certain linear representation of the infinite dimensional estimator, which is satisfied for example in the important case of nonparametric regression estimators and can also be used when the infinite dimensional parameter can depend on estimated random vectors - the so-called nonparametric generated regressor, see for example Escanciano, Jacho-Chávez, and Lewbel (2014, 2016). Finally we propose a test for omitted covariates in a partially linear regression model in which we allow some of the covariates to be endogenous (that is they are correlated with the unobservable errors) and others not to be directly observable but can be consistently estimated. Examples of unobservable covariates include individuals' expectations and risk terms, which are important in both economics and finance applications.

The rest of the paper is organized as follows: Next section introduces the statistical model and the test statistics and section 3 derives their asymptotic distributions. Section 4 illustrates the results of the paper with the partially linear model with unobservable and possibly endogenous covariates, while Section 5 provides evidence of the small sample performance of the proposed test statistics, and it also contains an empirical application to the electricity demand in Ontario (Canada) for the period 1971-1994. All mathematical proofs are gathered in the Appendix.

2 The Model and Test Statistics

Let $\{z_t, t = 1, 2, ...\}$ be a sequence of $\mathcal{Z}$-valued $\left(\mathcal{Z} \subset \mathbb{R}^d\right)$ weakly dependent random vectors defined on a probability space $(\Omega, \mathcal{B}, P)$ and let $S(z_t, \theta, h)$ denote a statistical model defined on it, where $\theta \in \Theta \subset \mathbb{R}^k$ is a finite dimensional parameter and $h \in \mathcal{H}$ is an infinite dimensional nuisance parameter where $\mathcal{H}$ is a pseudo-metric space. As in Andrews (1994a) h is allowed to depend on z_t and possibly on a finite dimensional parameter $\alpha \in A \subset \mathbb{R}^p$, so that $h_0 =: h_0(z_t, \alpha_0)$. If $S(z_t, \theta, h)$ is correctly specified, then typically there exist measurable functions $m(\cdot) : \mathcal{Z} \times B \times \mathcal{H} \to \mathbb{R}^s$ such that

$$E\left[m(z_t, \beta, h)\right] = 0 \quad \text{iff } \beta = \beta_0 \text{ and } h = h_0, \tag{2.1}$$

where $\beta \in B \subset \mathbb{R}^p$ $(p \geq k)$ might contain θ, and β_0, h_0 are the true unknown parameters. For example if $S(z_t, \theta, h)$ is a correctly specified partially linear time series regression model

$$y_t = x_{1t}'\theta_0 + g_0(x_{2t}) + \varepsilon_t, \quad t = 1, \ldots, T, \tag{2.2}$$

where g_0 is an unknown real valued function and ε_t is an unobservable error term, then $E(\varepsilon_t|\mathcal{I}_t) = 0$ $a.s.$, where $\mathcal{I}_t$ is the σ-field generated by a set of variables that contains, but it is larger than $x_t = [x_{1t}', x_{2t}']'$. Thus the null hypothesis

$$H_0: \; E(\varepsilon_t|\mathcal{I}_t) = 0 \; a.s. \tag{2.3}$$

can be used to test for the correct specification of (2.2). Suppose that a possible source of misspecification consists of an $s \times 1$ vector of additional (omitted) $\mathcal{I}_t$ - measurable covariates

x_{3t}. Then under (2.3), the covariance between ε_t and x_{3t} is zero, but possibly not under the alternative. Noting that (2.2) can be rewritten (see, i.e., Robinson, 1988) as

$$y_t = E(y_t|x_{2t}) - (x_{1t} - E(x_{1t}|x_{2t}))' \theta_0 + \varepsilon_t, \quad t = 1, \ldots, T, \tag{2.4}$$

the function

$$m(z_t, \beta, h_0) = \left[y_t - E(y_t|x_{2t}) - (x_{1t} - E(x_{1t}|x_{2t}))' \theta\right][x_{3t} - E(x_{3t}|x_{2t})], \tag{2.5}$$

with $\beta = \theta$ and $h_0 = \left[E(y_t|x_{2t}), E(x_{1t}|x_{2t})', E(x_{3t}|x_{2t})'\right]'$, provides the basis for an M test for the omission of relevant in (2.4) (and hence of (2.2)), since under (2.3) $E[m(z_t, \beta_0, h_0)] = E[\varepsilon_t (x_{3t} - E(x_{3t}|x_{2t}))] = 0$.

Let $m(z_t, \beta, h) := m_t(\beta, h)$; to handle the dependent structure of $m_t(\cdot)$ we follow the same approach[1] as in Smith (1997) and Kitamura and Stutzer (1997), and consider the following smoothed version of $m_t(\cdot)$

$$m_{ts}(\beta, h) = \frac{1}{s_T} \sum_{s=1-T}^{t-1} \omega\left(\frac{s}{s_T}\right) m_{t-s}(\beta, h), \quad t = 1, \ldots, T, \tag{2.6}$$

where s_T is a bandwidth parameter and $\omega(\cdot)$ is a kernel function. Smith (2011) provides a detailed discussion of different choices of $\omega(\cdot)$ and s_T in the context of finite dimensional parameter estimation, using as optimality criterion the asymptotic mean squared error used for example by Andrews (1991). A close inspection of Andrews's (1991) arguments reveals that his results can be applied to automatically select s_T in the semiparametric models of this paper.[2] To be specific, the optimal bandwidth is

$$s_T^* = \left(q\omega_q^{*2} \xi(q) T / \int \omega^*(x)^2 dx\right)^{1/(2q+1)}, \tag{2.7}$$
$$\xi(q) = \frac{2\text{vec}(S^q)' W \text{vec}(S^q)}{\text{trace}[W(I+K)(S \otimes S)]},$$

where $\omega^*(x) = \int \omega(y-x)\omega(y) dy/\omega_2$ is the induced kernel, $\omega_q^* = \lim_{x\to 0}(1-\omega^*(x))/|x|^q$, $\omega_j = \int \omega(x)^j dx$, S is the spectral density matrix at the zero frequency of $m_t(\beta_0, h_0)$, S^q is the generalized derivative of S defined as $\sum_{s=-\infty}^{\infty} |j|^s E\left[m_t(\beta_0, h_0) m_{t-s}(\beta_0, h_0)'\right]$, W is a nonstochastic $s^2 \times s^2$ weighting matrix, K is the commutation matrix and $\gamma \in (0, \infty) = \lim_{T\to\infty} s_T^{2q+1}/T$. Thus given a kernel $\omega(\cdot)$ in (2.6), the optimal bandwidth s_T^* depends on the induced kernel $\omega^*(x)$ and the unknown quantities S, S^q and W that need to be estimated. The dependence on the induced kernel $\omega^*(\cdot)$ arises implicitly in GEL estimation of the asymptotic

[1]For an asymptotically equivalent approach based on blocking techniques see for example Kitamura (1997).
[2]Details can be obtained from the authors upon request.

covariance $\lim_{T\to\infty} \text{var}(\sum_{t=1}^{T} m_t(\beta_0, h_0)/T^{1/2})$ and hence in the asymptotic mean squared calculation used to determine s_T^*. For example, if $\omega(\cdot)$ is the Bartlett kernel $\omega(x) = 1-|x|$ for $|x| \le 1$ and 0 elsewhere, the induced kernel $\omega^*(\cdot)$ is the Parzen kernel $\omega^*(x) = 1 - 6(x/2)^2 + 6|x/2|^3$ for $|x| \le 1$, $2(1-|x/2|)^3$ for $1 < |x| \le 2$ and 0 elsewhere. To estimate S and S^q one can use for example Andrews's (1991) parametric approach or Newey and West's (1994) nonparametric approach to obtain $\widehat{S}$ and $\widehat{S}^q$ and $\widehat{W}$ so that

$$\widehat{\xi}(q) = \frac{2\text{vec}(\widehat{S}^q)'\widehat{W}\text{vec}(\widehat{S}^q)}{\text{trace}[\widehat{W}(I+K)(\widehat{S}\otimes\widehat{S})]}.$$

It is important to note, however, that although s_T^* is consistent, it might not be optimal for the tests considered here. The selection of optimal bandwidths is still an open problem in testing and beyond the scope of the present paper.

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

$$\Gamma(\beta, h, \psi) = \frac{2}{T}\sum_{t=1}^{T} \rho(\omega\psi' m_{ts}(\beta, h)), \tag{2.8}$$

where $\omega = \omega_1/\omega_2$ is a normalization constant and ψ is a vector of unknown auxiliary parameters. (2.8) is the basis for the test statistics that we propose.

Let $\widehat{h}$ denote a preliminary estimator of h_0 (which might also depend on a $T^{1/2}$ - consistent estimator $\widehat{\alpha}$ for α_0), $\widehat{\beta}$ denote a $T^{1/2}$ - consistent estimator for β_0, $\widehat{m}_s(\beta, h) = \sum_{t=1}^{T} m_{ts}(\beta, h)/T$ and $\widehat{V}_{11}(\beta, h) = \sum_{t=1}^{T} m_{ts}(\beta, h)m_{ts}(\beta, h)'/T$. The M statistic for the null hypothesis (2.1) is

$$M_T = \frac{T}{\omega_1^2}\widehat{m}_s(\widehat{\beta}, \widehat{h})'\widehat{K}^{\times}(\widehat{\beta}, \widehat{h}, \widehat{g})^{-1}\widehat{m}_s(\widehat{\beta}, \widehat{h}),$$

while the GEL criterion function and the LM statistics for the same hypothesis are given by

$$D_T = \frac{2}{s_T}\left(\frac{\omega_2}{\omega_1^2}\right)[\Gamma(\widehat{\beta}, \widehat{h}, \widehat{\psi}) - \Gamma(\widehat{\beta}, \widehat{h}, 0)], \tag{2.9}$$
$$LM_T = \frac{T}{s_T^2}\widehat{\psi}'\widehat{V}_{11}(\widehat{\beta}, \widehat{h})\widehat{K}^{\times}(\widehat{\beta}, \widehat{h}, \widehat{g})^{-1}\widehat{V}_{11}(\widehat{\beta}, \widehat{h})\widehat{\psi},$$

where $\widehat{\psi} = \arg\max_{\psi} \Gamma(\widehat{\beta}, \widehat{h}, \psi)$ and $\widehat{K}^{\times}(\widehat{\beta}, \widehat{h}, \widehat{g})$ is an estimator of either the matrix $K(\beta_0, h_0, g_0)$ or the matrix $K^e(\beta_0, h_0, g_0)$ defined, respectively, in (3.2) and (3.3) below.

3 Asymptotic Results

We begin this section by introducing some further notation: let $\|\cdot\|$ denote the standard Euclidean norm and $\|\cdot\|_{\mathcal{H}}$ denote a function norm, such as the sup norm. Let $B_\delta = \{\beta \in B :$

$\|\beta - \beta_0\| \le \delta\}$, $\mathcal{H}_\delta = \{h \in \mathcal{H} : \|h - h_0\|_{\mathcal{H}} \le \delta\}$ (possibly uniformly in $\alpha \in A$), and let $\partial_\cdot$ denote the derivative operator with respect to $\cdot$, which corresponds to an ordinary partial derivative with respect to β, and to the pathwise derivative in the direction of $h - h_0$, that is

$$\frac{\partial m_t(\beta, h_0)}{\partial h}\left[h - h_0\right] := \frac{\partial m_t(\beta, (1-\tau)\, h_0 + \tau h)}{\partial \tau}|_{\tau=0}$$

(see Newey, 1994 for some examples).

Assume that:

Assumption 1 *$\{z_t, t = 1, 2, \ldots\}$ is a sequence of $\mathcal{Z}$-valued $\left(\mathcal{Z} \subset \mathbb{R}^d\right)$ stationary α-mixing random vectors with the mixing coefficient $\alpha\,(t) = o\left(t^{-2(2+\gamma)}\right)$ for some $\gamma > 0$.*

Assumption 2 *(a) $s_T \to \infty$ as $T \to \infty$, and $s_T = O(T^{1/2-\eta})$ for $1/6 < \eta < 1/2$;*
(b) $\omega\,(\cdot) : \mathbb{R} \to [-\overline{\omega}, \overline{\omega}]$ for some $\overline{\omega} < \infty$, $\omega\,(0) \ne 0$, $\omega_1 \ne 0$, $\omega\,(x)$ is continuous at 0 and almost everywhere, $|\,(2\pi)^{-1}\int_{-\infty}^{\infty} \exp\left(-\iota x u\right)\omega\,(x)\,dx| \ge 0$ $\forall u \in \mathbb{R}$, where $\iota = \sqrt{-1}$, and $\int_{-\infty}^{0} \sup_{y\le x} |\omega\,(y)\,dx| + \int_{0}^{\infty} \sup_{y\ge x} |\omega\,(y)\,dx| < \infty$ (c) $|\omega^*\,(x)| \le C_1/\,|x|^{-b}$ for $b > 1 + 1/q$ for some finite $C_1 > 0$ and $q \in (0,\infty)$ such that $\omega_q^* \in (0,\infty)$, $|\omega^*\,(x) - \omega^*\,(y)| \le C_2\,|x - y|$ $\forall x, y \in \mathbb{R}$ for some finite $C_2 > 0$.

Assumption 3 *(a) $E[\sup_{\beta\in B_\delta, h\in\mathcal{H}_\delta} \|m_t(\beta, h)\|^{\alpha_1}] < \infty$, $E\left[\sup_{\beta\in B_\delta, h\in\mathcal{H}_\delta} \|\partial_\beta m_t(\beta, h)\|^{\alpha_1}\right] < \infty$, $E[\|m_t(\beta_0, h_0)\|^{\alpha_2}] < \infty$, $E[\sup_{\beta\in B_\delta, h\in\mathcal{H}_\delta} \|\partial_h m_t(\beta, h)\|^{\alpha_2}] < \infty$ for some $\alpha_1 > 2$, $\alpha_2 > 4$; (b) $E\left[\sup_{\beta\in B_\delta, h\in\mathcal{H}_\delta} \left\|\partial_\beta m\,(\beta, h)\, m\,(\beta, h)'\right\|\right] < \infty$, $E\left[\sup_{\beta\in B_\delta, h\in\mathcal{H}_\delta} \left\|\partial^2_{\beta\beta} m\,(\beta, h)\, m\,(\beta, h)'\right\|\right] < \infty$, $E\left[\sup_{\beta\in B_\delta, h\in\mathcal{H}_\delta} \left\|\partial_\beta \partial_h m\,(\beta, h)\, m\,(\beta, h)'\right\|\right] < \infty$, $E\left[\sup_{\beta\in B_\delta, h\in\mathcal{H}_\delta} \left\|\partial^2_{hh} m\,(\beta, h)\, m\,(\beta, h)'\right\|\right] < \infty$.*

Assumption 4 *(a) $\|\, \widehat{h} - h_0 \,\|_{\mathcal{H}} = o_p\left(T^{-1/4}\right)$;*
(b) $\widehat{v}_T\,(h) := T^{-1/2}\sum_{t=1}^{T}\{m_t\,(\beta_0, h) - E\,[m_t\,(\beta_0, h)]\}$ is stochastically equicontinuous at h_0; (c) the classes of functions $\mathcal{M}_{\partial\beta}=\{\partial_\beta m\,(\beta, h)\,, \beta \in B_\delta, h \in \mathcal{H}_\epsilon\}$, $\mathcal{M}_{(\partial\beta)^2}=\{\partial_\beta m\,(\beta, h)\,\partial_\beta m\,(\beta, h)'\,, \beta \in B_\delta, h \in \mathcal{H}_\epsilon\}$ and $\mathcal{M}_{\partial^2 hh}=\{\partial^2_{hh} m\,(\beta_0, h)\,, h \in \mathcal{H}_\epsilon\}$ have, respectively, $N_{[]}(\varepsilon, \mathcal{M}_{\partial\beta}, L_1\,(P)) < \infty$, $N_{[]}(\varepsilon, \mathcal{M}_{(\partial\beta)^2}, L_1\,(P)) < \infty$ and $N_{[]}(\varepsilon, \mathcal{M}_{\partial^2 hh}, L_1\,(P)) < \infty$, where $N_{[]}\,(\cdot)$ is the bracketing number (see for example van der Vaart and Wellner, 1996 for a definition); (d) the classes of functions $\mathcal{M}^2_{\partial\beta}=\left\{\partial_\beta m\,(\beta, h)\, m\,(\beta, h)'\,, \beta \in B_\delta, h \in \mathcal{H}_\epsilon\right\}$, $\mathcal{M}^2_{\partial^2\beta\beta}=\{\partial^2_{\beta\beta} m\,(\beta, h)\, m\,(\beta, h)'\,, \beta \in B_\delta, h \in \mathcal{H}_\epsilon\}$, $\mathcal{M}^2_{\partial h}=\left\{\partial_h m\,(\beta, h)\, m\,(\beta, h)'\,, \beta \in B_\delta, h \in \mathcal{H}_\epsilon\right\}$, $\mathcal{M}^2_{\partial\beta\partial h}=\{\partial_\beta\partial_h m\,(\beta, h)\, m\,(\beta, h)'\,, \beta \in B_\delta, h \in \mathcal{H}_\epsilon\}$ and $\mathcal{M}^2_{\partial^2 hh}=\left\{\partial^2_{hh} m\,(\beta, h)\, m\,(\beta, h)'\,, \beta \in B_\delta, h \in \mathcal{H}_\epsilon\right\}$ have, respectively, $N_{[]}(\varepsilon, \mathcal{M}^2_{\partial\beta}, L_1\,(P)) < \infty$, $N_{[]}(\varepsilon, \mathcal{M}^2_{\partial^2\beta\beta}, L_1\,(P)) < \infty$, $N_{[]}(\varepsilon, \mathcal{M}^2_{\partial h}, L_1\,(P)) < \infty$, $N_{[]}(\varepsilon, \mathcal{M}^2_{\partial\beta\partial h}, L_1\,(P)) < \infty$ and $N_{[]}(\varepsilon, \mathcal{M}^2_{\partial^2 hh}, L_1\,(P)) < \infty$.

Assumption 5 *Either (a) $\|\, E[m_t(\beta_0, \widehat{h})] \,\| = o_p\left(T^{-1/2}\right)$; or (b) $E[\partial m(z_{2t}, \theta, \tau)/\partial\tau|_{\tau=h_0}\widetilde{h}(z_t)] = 0$ $\forall \widetilde{h} \in \mathcal{H}$ and $z_{2t} \subset z_t$.*

Assumption 6 *(a)* $\widehat{h}(w) - h_0(w) = T^{-1}\sum_{t=1}^{T} \Phi_T(z_{2t}, w) \odot \phi(z_t) + r_T(w)$, *where "$\odot$" is the Hadamard product,* $\Phi_T(z_{2t}, \cdot)$ *is some weighting function,* $\sup_w \|r_T(w)\| = o_p\left(T^{-1/2}\right)$ *(possibly uniformly in* $\alpha \in A$*);*
(b) $E[\phi(z_t) | \mathcal{F}_{t,z_{2t}}] = 0$, *where* $\mathcal{F}_{t,z_{2t}}$ *is the minimum* σ*-algebra generated by* z_{2t}*;* $E\left[\phi(z_t)\phi(z_t)'\right] < \infty$*; and* $\lim_{T\to\infty}\sup_w var(T^{-(1/2+\delta)}\sum_{t=1}^{T} \Phi_T(z_{2t}, w) \odot \phi(z_t)) < \infty$, *for some* $\delta \in (0, 1/2)$.

Assumption 7 *The estimator* $\widehat{\beta}$ *is such that*

$$T^{1/2}(\widehat{\beta}-\beta_0) = A(\beta_0, g_0)^{-1} \frac{1}{T^{1/2}} \sum_{t=1}^{n} f_t(\beta_0, g_0) + o_p(1) \xrightarrow{d} N(0, A(\beta_0, g_0)^{-1} B(\beta_0, g_0) A(\beta_0, g_0)^{-1\prime}), \tag{3.1}$$

where $A(\beta_0, g_0)$ *is an* $\mathbb{R}^{b\times b}$*-valued nonsingular matrix,* $B(\beta_0, g_0)$ *is a positive definite matrix and* g_0 *is an infinite dimensional parameter.*

Assumptions 1-3 are mild regularity conditions on the dependent structure of the observations, the kernel function used to smooth the observations and the existence of certain moments. Note that 2(c) is satisfied by the Bartlett, Parzen and the quadratic kernel. Assumption 4(a) assumes uniform consistency (possibly also with respect to α) of the nonparametric estimator used for h_0. For kernel estimators Andrews (1995) provides sufficient conditions including the case of estimated random variables. Sufficient conditions for Assumption 4(b) are provided for example in Andrews (1994b) and van der Vaart and Wellner (1996); Assumptions 4(c)-(d) are satisfied for example when $\partial_\beta m(\beta, h)$, $\partial^2_{hh} m(\beta, h)$, $\partial_\beta m(\beta, h) m(\beta, h)'$, $\partial^2_{\beta\beta} m(\beta, h) m(\beta, h)'$, $\partial_h m(\beta, h) m(\beta, h)'$, $\partial_\beta\partial_h m(\beta, h) m(\beta, h)'$ and $\partial^2_{hh} m(\beta, h) m(\beta, h)'$ are smooth in β and h, B is a compact set and h belongs to a class of sufficiently smooth functions, such as a Sobolev class. Coupled with Assumption 3(a) implies that the classes of functions $\mathcal{M}_{\partial\beta}$ and $\mathcal{M}_{\partial^2 hh}$ satisfy a uniform law of large numbers. Assumptions 3(b) and 4(d) are required for the calculation of the optimal bandwidth s^*_T. Assumption 5 implies the asymptotic orthogonality between the finite dimensional and the infinite dimensional parameter. In such case, it is not necessary to account for the presence of $\widehat{h}$ in the asymptotic distribution of $\widehat{\beta}$, which greatly simplifies the calculation of the asymptotic variance. Condition 5(a) is directly assumed by Andrews (1994a) and is also considered by Hjort, McKeague, and van Keilegom (2009); condition 5(b) is assumed by Newey (1994). Note that for $h = h(z_{2t})$ sufficient conditions for condition 5(a) are Assumptions 5(b) and 4(a). Note also that Assumption 5 is satisfied by many important semiparametric models, including partially linear, single index and partially additive models. Assumption 6 provides a generic way to account for the potential estimation effect arising from the estimation of the infinite dimensional parameter. Finally Assumption 7 is satisfied by many semiparametric estimators including those based on M, GMM (Generalized Method of Moments) and GEL estimation. Let I denote the identity matrix and

$$K(\beta,h,g) = [I, M(\beta,h) A(\beta,g)^{-1}] V(\beta,h,g) [I, M(\beta,h) A(\beta,g)^{-1}]', \tag{3.2}$$

$$V(\beta,h,g) = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} [m_t(\beta,h)', f_t(\beta,g)']' \right),$$

$$K^e(\beta,h,g) = [I, M(\beta,h) A(\beta,g)^{-1}] V^e_\times(\beta,h,g) [I, M(\beta,h) A(\beta,g)^{-1}]', \tag{3.3}$$

$$V^e_d(\beta,h,g) = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} [m^e_t(\beta,h)', f_t(\beta,g)']' \right) \text{ and}$$

$$V^e_{nd}(\beta,h,g) = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} \left[\left(m_t(\beta,h) + h^{(1)}_T(z_t,\beta,h) \right)', f_t(\beta,g)' \right]' \right),$$

where $M(\beta,h) = E[\partial_\beta m_t(\beta,h)]$,

$$m^e_t(\beta,h) = m_t(\beta,h) + \frac{1}{(T-1)} \sum_{s=1}^{t-1} \Psi(z_s,z_t,\beta,h),$$

$$\Psi(z_s,z_t,\beta,h) = \partial_h m_t(\beta_0,h_0)' \Phi_T(z_{2s},z_{2t}) \odot \phi(z_s) + \partial_h m_t(\beta_0,h_0)' \Phi_T(z_{2t},z_{2s}) \odot \phi(z_t), \tag{3.4}$$

$$h^{(1)}_T(\cdot,\beta_0,h_0) = E[\Psi(\cdot,z_t,\beta_0,h_0)] = \int \Psi(\cdot,u,\beta_0,h_0(u)) f_{z_t}(u) du.$$

Note that $V(\beta_0,h_0,g_0)$ corresponds to the asymptotic orthogonality case implied by Assumption 5, while the two alternative formulations $V^e_d(\beta_0,h_0,g_0)$ and $V^e_{nd}(\beta_0,h_0,g_0)$ correspond to the estimation effect of Assumption 6, which can be expressed as either a degenerate or a nondegenerate U statistic (see equation (A-11) in the Appendix for more details). Let

$$\widehat{\zeta}^\times(\beta,h,g) = \frac{\operatorname{trace}[\widehat{K}^\times(\beta,h,g)^{-1} \widehat{B}^m(\beta,h)]}{\operatorname{trace}[\widehat{V}_{11}(\beta,h)^{-1} \widehat{B}^m(\beta,h)]},$$

where $\widehat{B}^m(\beta,h) = [\sum_{t=1}^{T} m_{ts}(\beta,h)/\omega_1][\sum_{t=1}^{T} m_{ts}(\beta,h)/\omega_1]'$.

Theorem 3.1 *Assume that $K(\beta_0,h_0,g_0)$ and $K^e_\times(\beta_0,h_0,g_0)$ are positive definite, and $\| \widehat{K}(\widehat{\beta},\widehat{h},\widehat{g}) - K(\beta_0,h_0,g_0) \| = o_p(1)$, $\| \widehat{K}^e_*(\beta_0,\widehat{h},\widehat{g}) - K^e(\beta_0,h_0,g_0) \| = o_p(1)$. Then under Assumptions 1-7 and the null hypothesis that (2.1) holds*

$$\widehat{\zeta}^\times(\widehat{\beta},\widehat{h},\widehat{g}) D_T,\ LM_T \xrightarrow{d} \chi^2_s.$$

We now consider the local power and consistency of the proposed test statistics. To obtain the local power we assume that

$$E[m_t(\beta_0,h_0)] = \frac{\delta}{n^{1/2}}, \tag{3.5}$$

for some fixed vector $\delta \in \mathbb{R}^s$.

Theorem 3.2 *Under the Assumptions of Theorem 3.1 and the local hypothesis (3.5)*

$$\widehat{\zeta}(\widehat{\beta}, \widehat{h}, \widehat{g})D_T, LM_T \xrightarrow{d} \chi^2_s(\kappa^{\cdot})$$

with the noncentrality parameter $\kappa^{\cdot} = \delta' K^{\times}(\beta_0, h_0, g_0)^{-1}\delta$ *and* $K^{\times}(\cdot)$ *is either (3.2) or (3.3).*

To establish the consistency of the proposed test statistics we first note that under alternative distributions the probability limits of the estimators $\widehat{\beta}, \widehat{h}, \widehat{g}$ are typically different from β_0, h_0, g_0 defined under the null hypothesis of correct specification. Thus we assume that $\| \widehat{\beta} - \overline{\beta} \| = o_p(1)$, $\| \widehat{h} - \overline{h} \|_{\mathcal{H}} = o_p(1)$ and $\|\widehat{g} - \overline{g}\|_{\mathcal{G}} = o_p(1)$ where $\overline{\beta}$, $\overline{h}$ and $\overline{g}$ are not necessarily β_0, h_0, g_0 under a given alternative distribution.

Theorem 3.3 *Under the Assumptions of Theorem 3.1 and the Assumption that* $\|E[m_t(\overline{\beta}, \overline{h})]\| > 0$,

$$\widehat{\zeta}(\widehat{\beta}, \widehat{h}, \widehat{g})D_T, LM_T \xrightarrow{p} \infty.$$

4 Example: An M Test for Omitted Variables in a Partially Linear Model with a Generated Regressor

We consider a test for the omission of a set of relevant covariates in the same partially linear model

$$y_t = x_{1t}'\theta_0 + g_0(x_{2t}) + \varepsilon_t \quad t = 1, \ldots, T, \tag{4.1}$$

where θ_0 is an $\mathbb{R}^k$-valued vector of unknown parameters and $g_0(\cdot)$ in an unknown real valued function. We assume that $E[\varepsilon_t|x_{1t}] \neq 0$ and that $x_{2t} =: x_{2t}(\alpha_0)$ is generated as a residual from the following linear regression model

$$q_t = v_t'\alpha_0 + x_{2t},$$

where α_0 is a vector of unknown parameters and v_t is a vector of auxiliary covariates such that $E(x_{2t}|v_t) = 0$. Thus (4.1) is a partially linear regression model where the x_{1t} covariates are endogenous and x_{2t} is not directly observable but it can be consistently estimated as a regression residual. Suppose that there exists a vector i_t of so-called instruments such that $E(\varepsilon_t|x_{2t}, i_t) = 0$, assumed for simplicity to have the same dimension as that of x_{1t}. Let $\widehat{x}_{2t} = q_t - v_t'\widehat{\alpha}$, where $\widehat{\alpha}$ is the least squares estimator for α_0, and let

$$\widehat{\theta} = \left(\sum_{t=1}^{T} i_t \widetilde{x}_{1t}'\right)^{-1} \sum_{t=1}^{T} i_t \widetilde{y}_t,$$

denote the semiparametric instrumental variable (SIV) estimator, where

$$\widetilde{y}_t = y_t - \widehat{E}\left(y_t|\widehat{x}_{2t}\right), \ \widetilde{x}_{1t} = x_{1t} - \widehat{E}\left(x_{1t}|\widehat{x}_{2t}\right),$$
$$\widehat{E}\left(\cdot_t|x_{2t}\right) = \frac{\sum_{s\neq t=1}^{T}(\cdot_t)K_{b_T}\left(\left(x_{2s}-x_{2t}\right)/b_T\right)}{\sum_{s\neq t=1}^{T}K_{b_T}\left(\left(x_{2s}-x_{2t}\right)/b_T\right)},$$

and $K_{b_T}(\cdot) := K(\cdot)/b_T$ is a kernel function with bandwidth $b_T =: b(T)$. Under the regularity conditions given in Proposition 4.1 below, some calculations show that the SIV estimator admits the following asymptotic representation

$$T^{1/2}(\widehat{\theta}-\theta_0) = E\left[\left(x_{1t}-E\left(x_{1t}|x_{2t}\right)\right)i_t'\right]^{-1}\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(i_t\varepsilon_t + \left[\frac{i_t}{f(x_{2t})}\partial_\alpha[f(x_{2t})g_0(x_{2t},\theta_0)] - \frac{w_t[g_0(x_{2t},\theta_0)]}{f(x_{2t})}\partial_\alpha f(x_{2t})\right) r(v_t)x_{2t}\right),$$

where $f(x_{2t})$ is the marginal density of x_{2t}, $r(v_t) = E\left(v_t v_t'\right)^{-1} v_t$ and $g_0(x,\theta) = E\left(y_t - x_{1t}'\theta|x_{2t} = x\right)$.

As in Section 2, an M test for the omission of an $\mathbb{R}^s$-valued vector of relevant covariates x_{3t} can be constructed using the sample version of the same function given in (2.5), that is

$$m_t(\widehat{\theta},\widehat{h}) = \widehat{\varepsilon}_t[x_{3t} - \widehat{E}\left(x_{3t}|\widehat{x}_{2t}\right)], \tag{4.2}$$

where $\widehat{\varepsilon}_t = \widetilde{y}_t - \widetilde{x}_{1t}'\widehat{\theta}$ denote the SIV residual. In this case the three statistics M_T, D_T and LM_T are computed as

$$M_T = \frac{T}{\omega_1^2}\widehat{m}_s(\widehat{\beta},\widehat{h})'\widehat{K}^e(\widehat{\beta},\widehat{h},\widehat{g})^{-1}\widehat{m}_s(\widehat{\beta},\widehat{h}), \tag{4.3}$$
$$\widehat{\zeta}(\widehat{\beta},\widehat{h},\widehat{g})D_T \text{ with } \widehat{\zeta}(\widehat{\theta},\widehat{h},\widehat{g}) = \frac{\text{trace}[\widehat{K}^e(\widehat{\theta},\widehat{h},\widehat{g})^{-1}\widehat{B}^m(\widehat{\theta},\widehat{h})]}{\text{trace}[\widehat{V}_{11}(\widehat{\theta},\widehat{h})^{-1}\widehat{B}^m(\widehat{\theta},\widehat{h})]},$$
$$LM_T = \frac{T}{s_T^2}\widehat{\psi}'\widehat{V}_{11}(\widehat{\beta},\widehat{h})\widehat{K}^e(\widehat{\beta},\widehat{h},\widehat{g})^{-1}\widehat{V}_{11}(\widehat{\beta},\widehat{h})\widehat{\psi} \text{ and}$$
$$\widehat{K}^e(\widehat{\theta},\widehat{h},\widehat{g}) = [I,\widehat{M}(\widehat{\theta},\widehat{h})]\widehat{V}^e(\widehat{\theta},\widehat{h},\widehat{g})[I,\widehat{M}(\widehat{\theta},\widehat{h})]',$$

where $\widehat{M}(\widehat{\theta},\widehat{h})$ and $\widehat{V}^e(\widehat{\theta},\widehat{h},\widehat{g})$ are consistent estimators of

$$M(\theta_0,h_0) = E\left[(x_{3t}-E(x_{3t}|x_{2t}))(x_{1t}-E(x_{1t}|x_{2t}))'\right],\ A(\theta_0,g_0) = E\left[(x_{1t}-E(x_{1t}|x_{2t}))\, i_t'\right],$$

$$V_{11}(\theta_0,h_0) = \lim_{T\to\infty}\operatorname{var}\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T} m_t(\theta_0,h_0)\right),\ V_{22}^e(\theta_0,g_0) = A(\theta_0,g_0)^{-1}\,\Omega^e(\theta_0,g_0)\,(A(\theta_0,h_0)^{-1})',$$

$$\Omega^e(\theta_0,g_0) = \lim_{T\to\infty}\operatorname{var}\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(i_t\varepsilon_t + \left[\frac{i_t}{f(x_{2t})}\partial_\alpha[f(x_{2t})g_0(x_{2t},\theta_0)] - \frac{i_t[g_0(x_{2t},\theta_0)]}{f(x_{2t})}\partial_\alpha f(x_{2t})\right) r(v_t)x_{2t}\right)\right),$$

$$V_{12}^e(\theta_0,h_0,g_0) = \lim_{T\to\infty}\frac{1}{T}E\left[\sum_{t=1}^{T} m_t(\theta_0,h_0),[A(\theta_0,g_0)^{-1}\left(\sum_{t=1}^{T} i_t\varepsilon_t + \left[\frac{i_t}{f(x_{2t})}\partial_\alpha[f(x_{2t})g_0(x_{2t},\theta_0)] - \frac{i_t[g_0(x_{2t},\theta_0)]}{f(x_{2t})}\partial_\alpha f(x_{2t})\right) r(v_t)x_{2t}\right)'\right] \text{ and}$$

$$V^e(\theta_0,h_0,g_0) = \begin{bmatrix} V_{11}(\theta_0,h_0) & V_{12}^e(\theta_0,h_0,g_0) \\ V_{12}^e(\theta_0,h_0,g_0)' & V_{22}^e(\theta_0,h_0,g_0) \end{bmatrix}.$$

The following proposition provides sufficient conditions for the results of Theorems 3.1-3.3 to hold for the M test based on (4.2); note that because the unknown parameters enters linearly in the model, some of the conditions are weaker than those assumed in Assumptions 1-7.

Proposition 4.1 *Assume that: (a)* $\left\{z_t = [y_t, x_{1t}', x_{2t}, x_{3t}', i_t']'\right\}_{t=1}^{T}$ *is a sequence of* α*-mixing random vectors with* $\alpha(t) = o\left(t^{-2(2+\gamma)}\right)$*; (b) Assumption 2 holds; (c) the joint density* $f(z_t)$ *of* z_t *and the marginal density* $f(x_{2t})$ *of* x_{2t} *are twice continuously differentiable with bounded derivatives, the support* $\mathcal{X}_2$ *of* x_{2t} *is a compact set,* $\inf_{x_{2t}\in\mathcal{X}_2} f(x_{2t}) > 0$ *and* $f(x_{2t})$ *is continuously differentiable with respect to* α *with bounded derivative uniformly in* A*; (d)* $h_0(x_{2t})$ *is twice continuously differentiable,* $\sup_{x_{2t}\in\mathcal{X}^*}\| h_0^{(j)}(x_{2t})\| < \infty$ $(j=0,1,2)$ *uniformly in* A *where* $h_0^{(j)}(\cdot)$ *is the* j*th derivative of* $h_0(\cdot)$ *and* $h_0(x_{2t})$ *is continuously differentiable with respect to* α *with bounded derivative uniformly in* A*; (e)* $E\| i_t(y_t - E(y_t|x_{2t}) - (x_{1t}-E(x_{1t}|x_{2t}))'\theta_0)\|^{4+\gamma} < \infty$*; (f)* $\operatorname{rank}(E\left[i_t(x_{1t}-E(x_{1t}|x_{2t}))'\right]) = k$*,* $\operatorname{rank}(M(\theta_0,h_0)) = s$*, the matrix* $V^e(\theta_0,h_0,g_0)$ *in (4.3) is positive definite; (g) the function* $K(\cdot)$ *is a nonnegative second-order kernel with second order continuous bounded derivatives, and* b_T *satisfies* $T^{1/2}b_T^2 \to \infty$*,* $T^{1/2}b_T^4 \to 0$*. Moreover* $\left|K(\cdot+u) - K(u) - K^{(1)}(\cdot)u\right| \le \mathcal{K}(\cdot)u^2$ *where* $K^{(1)}(\cdot)$ *is the first derivative of the kernel function and* $\mathcal{K}(\cdot)$ *is a bounded function. Then the conclusions of Theorems 3.1-3.3 hold for the test statistics defined in (4.3), where* $h(x,\theta) := E[y_t - x_{1t}'\theta|x_{2t}=x]$ *and* $x_{2t} = q_t - v_t'\alpha_0$*.*

In the next section we operationalize these test statistics in various Monte Carlo experiments and an empirical application using bandwidths calculated by Silverman's (1986) rule of thumb. Although these are only optimal for estimation and not for testing (see, e.g., Gao and Gijbels, 2008), as shown below they seem to work quite well in practice.

5 Numerical Results

5.1 Monte Carlo Results

In this section we present Monte Carlo results for the size and power of the three statistics M_T, $\widehat{\zeta}(\widehat{\beta},\widehat{h},\widehat{g})D_T$ and LM_T used to test for omitted variables in the partially linear model with endogenous regressor and a generated regressor described in the previous section. The following design is a modified version of that used in Bravo, Chu, and Jacho-Chávez (forthcoming):

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

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

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

Let $\zeta_{lt} \sim N(0,1)$ $(l = 2,3,4,5)$ independent of v_{1t} and v_{2t}, and set $x_{12t} = v_{2t} + \zeta_{2t}$, $x_{2t} = v_{1t} + v_{2t} + \omega_{3t}$ such that $q_t = \zeta_{4t}\alpha_0 + x_{2t}$. For $\rho_1 = \rho_2 = 0.4$, $\rho_\varepsilon = \rho_u = 0.95$, and $m_0(v) = \Phi(v)$ ($\Phi(\cdot)$ is the CDF of a standard normal), we generate 1000 samples, $\{y_t, x_{11t}, x_{12t}, x_{2t}, v_{1t}, v_{2t}\}_{t=1}^T$, with $T \in \{100, 400\}$, three different scenarios $\rho_{\varepsilon u} \in \{0.1, 0.5, 0.9\}$ representing an increasing degree of endogeneity and $\theta_0 = [1,1]'$, $\pi_0 = [1,-1]'$. Finally $x_{3t} = 0.5x_{3t-1} + 0.2x_{3t-2} + \zeta_{5t}$ is the the additional covariate suspected of being erroneously omitted.

The SIV estimator $\widehat{\theta}$ is computed using as instruments $i_t := [x_{12t}, v_{1t}]'$ whereas $\widehat{h} = [\widehat{E}(y_t|\widehat{x}_{2t}), \widehat{E}(x_{12t}|\widehat{x}_{2t}), \widehat{E}(x_{3t}|\widehat{x}_{2t})]'$ is computed using the Nadaraya-Watson estimator with a Gaussian kernel and bandwidth b_T chosen by the Silverman's rule-of-thumb and $\widehat{\alpha}$ is the least squared estimator of α_0 obtained from regressing q_t on ζ_{4t}.

We calculate the $\widehat{\zeta}(\widehat{\beta},\widehat{h},\widehat{g})D_T := D_T^\times$ and $LM_T^\times$ statistics using as $\omega(\cdot)$ the Bartlett kernel with bandwidth parameter s_T^* selected using (2.7) with $\xi(q)$ chosen by Andrews's (1991) parametric specification and for three specifications of the GEL objective function $\rho(\cdot)$ given in (2.8), that is $\rho(v) = \log(1-v)$ corresponding to Empirical Likelihood (EL), $\rho(v) = -\exp(v)$ corresponding to the Exponential Tilting (ET) and $\rho(v) = -(1+v)^2/2$ corresponding to Euclidean Likelihood (EU). To assess the sensitivity of the statistics to the chosen bandwidth b_T, we also consider two other bandwidths: $b_{1T} = 0.5b_T$ and $b_{2T} = 1.5b_T$. Tables 1 and 2 report the finite sample sizes of the nine statistics for all the different scenarios, bandwidth b_T choices and the two chosen sample sizes using 5000 replications.

Table 1. Finite sample sizes for $T = 100$

	b_{1T}		b_T		b_{2T}	
$\rho_{\varepsilon u} = 0.1$	0.010	0.050	0.010	0.050	0.010	0.050
M_T	0.018	0.061	0.017	0.059	0.018	0.060
LM_T^{EL}	0.019	0.060	0.020	0.058	0.019	0.057
LM_T^{ET}	0.020	0.057	0.019	0.057	0.018	0.056
LM_T^{EU}	0.019	0.056	0.018	0.056	0.020	0.055
D_T^{EL}	0.020	0.056	0.017	0.057	0.018	0.056
D_T^{ET}	0.018	0.055	0.018	0.056	0.019	0.057
D_T^{EU}	0.021	0.054	0.019	0.055	0.019	0.056
$\rho_{\varepsilon u} = 0.5$	0.010	0.050	0.010	0.050	0.010	0.050
M_T	0.021	0.064	0.023	0.063	0.025	0.065
LM_T^{EL}	0.020	0.061	0.019	0.060	0.018	0.059
LM_T^{ET}	0.021	0.058	0.022	0.058	0.020	0.057
LM_T^{EU}	0.020	0.058	0.021	0.059	0.022	0.058
D_T^{EL}	0.018	0.056	0.017	0.055	0.016	0.054
D_T^{ET}	0.017	0.057	0.018	0.056	0.016	0.055
D_T^{EU}	0.020	0.055	0.019	0.057	0.018	0.056
$\rho_{\varepsilon u} = 0.9$	0.010	0.050	0.010	0.050	0.010	0.050
M_T	0.026	0.073	0.027	0.076	0.029	0.078
LM_T^{EL}	0.021	0.061	0.023	0.059	0.022	0.061
LM_T^{ET}	0.021	0.059	0.023	0.061	0.024	0.063
LM_T^{EU}	0.022	0.060	0.021	0.062	0.022	0.062
D_T^{EL}	0.018	0.057	0.020	0.058	0.022	0.059
D_T^{ET}	0.018	0.058	0.019	0.059	0.021	0.057
D_T^{EU}	0.022	0.058	0.021	0.059	0.020	0.061

Note: Results are based on 1000 Monte Carlo replications. *EL*, *ET*, *EU* stands for Empirical Likelihood, Exponential Tilting and Euclidean Likelihood respectively. Bandwidths b_T correspond to Silverman's rule of thumb, $b_{1T} = 0.5b_T$ and $b_{2T} = 1.5b_T$.

Table 2. Finite sample sizes for $T = 400$

	b_{1T}		b_T		b_{2T}	
$\rho_{\varepsilon u} = 0.1$	0.010	0.050	0.010	0.050	0.010	0.050
M_T	0.017	0.057	0.016	0.058	0.016	0.057
LM_T^{EL}	0.018	0.060	0.018	0.056	0.019	0.057
LM_T^{ET}	0.019	0.055	0.018	0.055	0.016	0.056
LM_T^{EU}	0.019	0.054	0.017	0.055	0.019	0.055
D_T^{EL}	0.019	0.055	0.016	0.057	0.017	0.054
D_T^{ET}	0.017	0.054	0.017	0.055	0.017	0.056
D_T^{EU}	0.020	0.054	0.018	0.054	0.018	0.055
$\rho_{\varepsilon u} = 0.5$	0.010	0.050	0.010	0.050	0.010	0.050
M_T	0.020	0.062	0.022	0.063	0.024	0.062
LM_T^{EL}	0.019	0.060	0.019	0.060	0.016	0.059
LM_T^{ET}	0.020	0.057	0.022	0.058	0.020	0.057
LM_T^{EU}	0.020	0.058	0.019	0.059	0.020	0.057
D_T^{EL}	0.017	0.054	0.015	0.054	0.016	0.054
D_T^{ET}	0.017	0.056	0.018	0.056	0.014	0.055
D_T^{EU}	0.019	0.056	0.017	0.057	0.017	0.056
$\rho_{\varepsilon u} = 0.9$	0.010	0.050	0.010	0.050	0.010	0.050
M_T	0.022	0.072	0.024	0.075	0.028	0.075
LM_T^{EL}	0.020	0.056	0.021	0.057	0.022	0.061
LM_T^{ET}	0.019	0.059	0.020	0.058	0.021	0.063
LM_T^{EU}	0.020	0.056	0.018	0.057	0.020	0.060
D_T^{EL}	0.017	0.057	0.016	0.055	0.018	0.058
D_T^{ET}	0.015	0.056	0.018	0.057	0.019	0.056
D_T^{EU}	0.018	0.056	0.020	0.058	0.021	0.056

Note: Results are based on 1000 Monte Carlo replications. *EL*, *ET*, *EU* stands for Empirical Likelihood, Exponential Tilting and Euclidean Likelihood respectively. Bandwidths b_T correspond to Silverman's rule of thumb, $b_{1T} = 0.5b_T$ and $b_{2T} = 1.5b_T$.

The results of Tables 1 and 2 can be summarized as follows: first, all of the test statistics are characterized by good finite sample sizes close to the nominal level. As expected the size distortion is more evident when the degree of endogeneity is higher (that is for $\rho_{\varepsilon u} = 0.9$) and decreases when the sample sizes increases. Between the nine statistics, M_T (that is the one based on the standard M formulation) is the one with the largest size distortion, whereas the GEL objective function statistics have typically the smallest one. Between EL, ET and EU

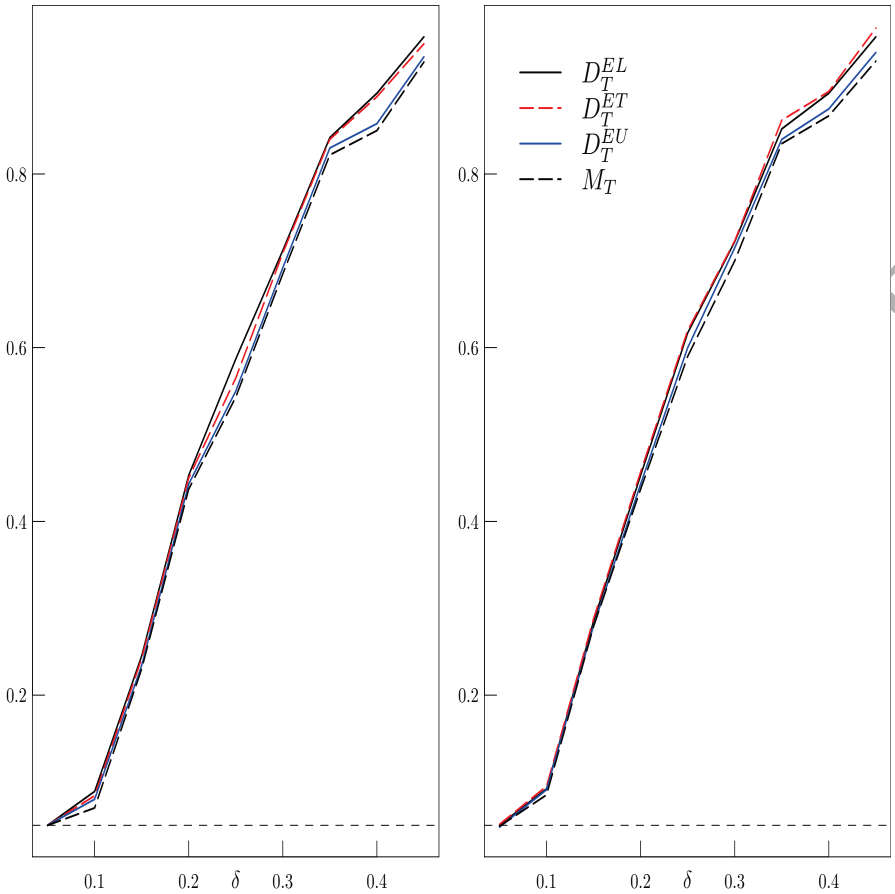


Figure 1: Finite sample size adjusted power for M_T, D_T^{EL}, D_T^{ET} and D_T^{EU} statistics. Left panel corresponds to $\rho_{\varepsilon u} = 0.1$, right panel corresponds to $\rho_{\varepsilon u} = 0.9$

likelihood, the former two seems to have a slight edge over the latter one in terms of (smaller) size distortion. Finally, the results seem to be robust to the choice of bandwidth b_T.

Figure 1 reports the finite sample (size adjusted) power of the M_T and the three GEL objective functions statistics D_T^{EL}, D_T^{ET} and D_T^{EU} for a sequence of alternative hypotheses indexed by $\delta = [0.05, 0.1, 0.2, 0.25, 0.30, 0.35, 0.40, 0.45]$ for both cases of low and high endogeneity with bandwidth set at Silverman's rule-of-thumb and sample size $T = 100$.

Figure 1 shows that all test have good finite sample power against the alternative hypothesis with that of D_T^{EL} and D_T^{ET} having an edge on that of D_T^{EU} and M_T.

Taken together the results of the simulation study suggest that all of the proposed tests are characterized by good finite sample properties that are robust to the choice of bandwidth b_T. Among the statistics considered, those based on either the empirical likelihood or exponential tilting objective functions seem to have an advantage in terms of smaller finite sample size distortion and larger finite sample power.

5.2 Empirical Application

In this section we proceed to implement the proposed test statistics to test whether lagged dependent variables, what is often called in the economic literature state dependency, have been omitted in the partially linear specification of the electricity demand function in Yatchew (2003). The data is publicly available and it comes from the Ontario Hydro Corporation. It has been previously used by Chu, Huynh, and Jacho-Chávez (2013), and it consists of 288 quarterly observations from the Canadian province of Ontario between 1971 and 1994. Yatchew's (2003, eq. 4.6.9, pp. 81) model is

$$\text{elec}_t - \text{gdp}_t = \theta_0 \text{relprice}_t + g_0(\text{temp}_t) + \varepsilon_t, \text{ for } t = 2, \ldots, 288, \tag{5.1}$$

where elec_t is the log of electricity sales in millions of Canadian dollars, gdp_t is the log of Ontario gross domestic product in millions of Canadian dollars, relprice_t is defined as the log of ratio of price electricity to the price of natural gas, and temp_t is the difference between the number of days the temperature is above 68°F and the number of days is below it. We proceed to calculate M_T, D_T and LM_T test statistics to check for the presence of state dependency or seasonal state dependency in (5.1), using, respectively, $\text{elec}_{t-1} - \text{gdp}_{t-1}$ and $\text{elec}_{t-4} - \text{gdp}_{t-4}$ as x_{3t} in (2.5). As in Section 5.1, the test statistics are calculated using a Gaussian kernel function with Silverman's rule-of-thumb bandwidths for all the conditional expectations and a Barlett kernel for $\omega(\cdot)$ with bandwidth $s_T^* = 7$. The estimator in Robinson (1988) is used to calculate $\widehat{\theta} = -0.0722$, which is statistically significant at all levels. EL is used to calculate the sample values of the test statistics M_T,D_T and LM_T, which are, respectively, 0.0007, 0.0019 and 0.0121 ($p = 1$) and 0.0112, 0.0243 and 0.118 ($p = 4$). These test statistics are statistically insignificant (at 10%) when comparing them to the 2.71 critical value from a χ_1^2 distribution under the null hypothesis. As a robustness check we also calculated all the test statistics using half and one and a half times the original rule-of-thumb bandwidths yielding the same results. For example, when using half the original rule-of-thumb bandwidths, the test statistics M_T,D_T and LM_T are respectively, 0.0000, 0.0002 and 0.0015 ($p = 1$) and 0.0061, 0.0152 and 0.0841 ($p = 4$). Similarly, when using one and a half the original rule-of-thumb bandwidths, the calculated test statistics are 0.0018, 0.0048 and 0.0276 ($p = 1$) and 0.0157, 0.0334 and 0.1576 ($p = 4$). This confirms that there is neither state nor seasonal state dependency in the demand for electricity as estimated in Yatchew (2003), and this finding seems to be robust to bandwidth choice.

6 Conclusions

In this paper we have considered the problem of M testing in the context of smooth semiparametric models with time series observations. The statistical models we have considered are fairly general and can allow for endogeneity and generated regressors. We have derived the asymptotic properties of a number of test statistics based on a smoothed version of GEL method and illustrated them by considering a test for omitted variables in a semiparametric time series regression model with endogenous covariates and a nonparametric generated regressor. The results of a Monte Carlo study suggest that the proposed test statistics have competitive finite sample properties, and its application to test for state dependency in an estimated semiparametric electricity demand model shows its practical usefulness.

References

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

——— (1994a): "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

——— (1994b): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics, Vol. IV*, ed. by R. Engle, and D. McFadden, pp. 2247–2294. New York: North Holland.

——— (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560–596.

Bertail, P. (2006): "Empirical Likelihood in some Semiparametric Models," *Bernoulli*, 12, 299–331.

Bravo, F. (2009): "Two-step Generalised Empirical Likelihood Inference for Semiparametric Models," *Journal of Multivariate Analysis*, 100(7), 1412–1431.

Bravo, F., B. M. Chu, and D. T. Jacho-Chávez (forthcoming): "Semiparametric Estimation of Moment Condition Models with Weakly Dependent Data," forthcoming in the *Journal of Nonparametric Statistics*.

Chu, B. M., K. P. Huynh, and D. T. Jacho-Chávez (2013): "Functionals of Order Statistics and their Multivariate Concomitants with Application to Semiparametric Estimation by Nearest Neighbors," *Sankhya: Series B*, 75(2), 238–292.

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

Escanciano, J. C., D. T. Jacho-Chávez, and A. Lewbel (2014): “Uniform Convergence of Weighted Sums of Non- and Semi-parametric Residuals for Estimation and Testing,” *Journal of Econometrics*, 178, 426–443.

——— (2016): “Identification and Estimation of Semiparametric Two Step Models,” *Quantitative Economics*, 7(2), 561–589.

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

Gao, J., and I. Gijbels (2008): “Bandwidth Selection in Nonparametric Kernel Testing,” *Journal of the American Statistical Association*, 103(484), 1584–1594.

Hjort, N. L., I. W. McKeague, and I. van Keilegom (2009): “Extending the Scope of Empirical Likelihood,” *Annals of Statistics*, 37(3), 1079–1111.

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

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

Masry, E. (1996): “Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates,” *Journal of Time Series Analysis*, 17(6), 571–599.

Newey, W., and K. West (1994): “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economics Studies*, 61, 631–653.

Newey, W. K. (1985): “Generalized Method of Moments Specification Testing,” *Journal of Econometrics*, 29, 229–256.

——— (1991): “Uniform convergence in probability and stochastic equicontinuity,” *Econometrica*, 59, 1161–1167.

——— (1994): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62, 1349–1382.

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

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

Robinson, P. M. (1988): "Root n-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1 edn.

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

——— (2011): "GEL Criteria for Moment Condition Models," *Econometric Theory*, 27, 1192–1235.

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

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

——— (1994): *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge, 1 edn.

Yatchew, A. (2003): *Semiparametric Regression for the Applied Econometrician*, Themes in Modern Econometrics. Cambridge University Press, Cambridge, 1 edn.

Appendix A Proofs

Proof of Theorem 3.1: Without loss of generality we normalize the first two derivatives $\rho_j(0) = -1$ $(j = 1, 2)$ of $\rho(\cdot)$, where $\rho_j(0) := \partial^j \rho(q) / \partial q^j|_{q=0}$. Let $\Psi_T^r = \{\psi : \|\psi\| \leq R_T\}$ where $R_T = O_p(s_T/T)^{\xi}$ for $\xi < 1/2$; we first show that

$$\left\| \widehat{m}_s(\widehat{\beta}, \widehat{h}) - \omega_1 E\left[m_t(\beta_0, h_0)\right] \right\| = o_p(1), \tag{A-1}$$

$$\max_{1 \leq t \leq T} \sup_{\psi \in \Psi_T^r} \left| \psi' m_{ts}(\widehat{\beta}, \widehat{h}) \right| = o_p(1) \text{ and} \tag{A-2}$$

$$\left\| \left(\frac{1}{s_T} \sum_{s=1-T}^{T-1} \omega\left(\frac{t}{s_T}\right)^2 \right)^{-1} \frac{s_T}{T} \sum_{t=1}^{T} m_{ts}(\widehat{\beta}, \widehat{h}) m_{ts}(\widehat{\beta}, \widehat{h})' - V_{11}(\beta_0, h_0) \right\| = o_p(1). \tag{A-3}$$

By the triangle inequality

$$\begin{aligned}\left\| \widehat{m}_s(\widehat{\beta}, \widehat{h}) - \omega_1 E\left[m_t(\beta_0, h_0)\right] \right\| \leq{}& \left| \sum_{j=1-T}^{T-1} \frac{1}{s_T} \omega\left(\frac{s}{s_T}\right) \right| \sup_{\beta \in B_\delta, h \in \mathcal{H}_\delta} \left\| \widehat{m}(\beta, h) - E\left[m_t(\beta, h)\right] \right\| + \\ & \left| \sum_{s=1-T}^{T-1} \frac{1}{s_T} \omega\left(\frac{s}{s_T}\right) - \omega_1 \right| E \sup_{\beta \in B_\delta, h \in \mathcal{H}_\delta} \left\| m_t(\beta, h) \right\| + \\ & + \omega_1 \left\| E[m_t(\widehat{\beta}, \widehat{h})] - E\left[m_t(\beta, h_0)\right] \right\| = o_p(1).\end{aligned} \tag{A-4}$$

The first term on the right hand side of (A-4) converges in probability to zero by the uniform law of large number (implied by Assumptions 3(a) and 4(b)), see e.g. Newey, 1991), while the second term is $o(1)$ since $\left| \sum_{s=1-T}^{T-1} \omega(j/s_T)/s_T - \omega_1 \right| = o(1)$; finally the third term on the right hand side of (A-4) is $o(1)$ by dominated convergence hence $\sup_{\beta \in B_\delta, h \in \mathcal{H}_\delta} \left\| \widehat{m}(\beta, h) - E\left[m_t(\beta, h)\right] \right\| = o_p(1)$. To show (A-2), note that by triangle inequality and the (functional) mean value theorem one has

$$\begin{aligned}\max_{1 \leq t \leq T} \left| \psi' m_{ts}(\widehat{\beta}, \widehat{h}) \right| \leq{}& R_T \left| \sum_{s=1-T}^{T-1} \frac{1}{s_T} \omega\left(\frac{s}{s_T}\right) \right| \max_{1 \leq t \leq T} \Big[\left\| m_t(\beta_0, h_0) \right\| + \\ & \sup_{\beta \in B_\delta, h \in \mathcal{H}_\delta} \left\| \partial_h m_t(\beta, h) \right\| \left\| \widehat{h} - h_0 \right\|_{\mathcal{H}} \\ & + \sup_{\beta \in B_\delta, h \in \mathcal{H}_\delta} \left\| \partial_\beta m_t(\beta, h) \right\| \left\| \widehat{\beta} - \beta_0 \right\| \Big] = o_p(1),\end{aligned}$$

since the Borel-Cantelli lemma and the moment conditions in Assumptions 3 imply that

$$\begin{aligned}& \max_{1 \leq t \leq T} \left\| m_t(\beta_0, h_0) \right\|, \ \max_{1 \leq t \leq T} \sup_{\beta \in B_\delta, h \in \mathcal{H}_\delta} \left\| \partial_h m_t(\beta, h) \right\|, \\ & \sup_{\beta \in B_\delta, h \in \mathcal{H}_\delta} \left\| \partial_\beta m_t(\beta, h) \right\|\end{aligned}$$

are all $o_p\left(T^{1/2}\right)$. Finally from the triangle inequality

$$\left\|\frac{s_T}{T}\sum_{t=1}^{T} m_{ts}(\widehat{\beta},\widehat{h})m_{ts}(\widehat{\beta},\widehat{h})' - \omega_2 V_{11}\left(\beta_0,h_0\right)\right\| \leq \left\|\frac{s_T}{T}\sum_{t=1}^{T} m_{ts}\left(\beta_0,h_0\right)m_{ts}\left(\beta_0,h_0\right)' - \omega_2 V_{11}\left(\beta_0,h_0\right)\right\| +$$
$$2\left\|\frac{s_T}{T}\sum_{t=1}^{T} m_{ts}\left(\beta_0,h_0\right)[\sum_{t=1}^{T} m_{ts}(\widehat{\beta},\widehat{h}) - m_{ts}\left(\beta_0,h_0\right)]'\right\| + \left|\frac{s_T}{T}\sum_{t=1}^{T}\left\|m_{ts}(\widehat{\theta},\widehat{h}) - m_{ts}\left(\theta_0,h_0\right)\right\|^2\right|;$$

a similar calculation to that used by Smith (2011) shows that

$$\left\|\frac{s_T}{T}\sum_{t=1}^{T} m_{ts}\left(\beta_0,h_0\right)m_{ts}\left(\beta_0,h_0\right)' - \omega_2 V_{11}\left(\beta_0,h_0\right)\right\| = o_p\left(1\right),$$

while

$$\left|\frac{s_T}{T}\sum_{t=1}^{T}\left\|m_{ts}(\widehat{\theta},\widehat{h}) - m_{ts}\left(\theta_0,h_0\right)\right\|^2\right| \leq \sup \|\ \widehat{h} - h_0\ \|_{\mathcal{H}}^2 \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta}\left\|\partial_h m_t\left(\theta,h\right)\right\|^2 = o_p\left(1\right) \tag{A-5}$$

by the uniform convergence of kernel estimators, see for example Masry (1996) and the uniform law of large numbers (implied by Assumptions 3(a) and 4(b)); finally by the Cauchy-Schwarz inequality and (A-5)

$$\left\|\frac{s_T}{T}\sum_{t=1}^{T} m_{ts}\left(\beta_0,h_0\right)[\sum_{t=1}^{T} m_{ts}(\widehat{\beta},\widehat{h}) - m_{ts}\left(\beta_0,h_0\right)]'\right\| \leq$$
$$\left(\frac{s_T}{T}\sum_{t=1}^{T}\left\|m_{ts}\left(\beta_0,h_0\right)\right\|^2\right)^{1/2}\left(\frac{s_T}{T}\sum_{t=1}^{T}\left\|m_{ts}(\widehat{\theta},\widehat{h}) - m_{ts}\left(\theta_0,h_0\right)\right\|^2\right)^{1/2} = o_p\left(1\right).$$

The continuous mapping theorem implies that for $j = 1, 2$

$$\sup_{\psi\in\Psi_T^r}\max_{1\leq t\leq T}\left|\rho_j(\psi' m_{ts}(\widehat{\beta},\widehat{h}))) - \rho_j\left(0\right)\right| = o_p\left(1\right),$$

thus by a second-order Taylor expansion about $\psi = 0$, we have that

$$D_T = \frac{T}{\omega_1^2}\widehat{m}_s(\widehat{\beta},\widehat{h})'\widehat{V}_{11}(\widehat{\beta},\widehat{h})^{-1}\widehat{m}_s(\widehat{\beta},\widehat{h}) + o_p\left(1\right), \tag{A-6}$$

where we have used the fact that

$$T^{1/2}\frac{\widehat{\psi}}{s_T} = V_{11}\left(\beta_0,h_0\right)^{-1}\frac{T^{1/2}}{\omega_1}\widehat{m}_s(\widehat{\beta},\widehat{h}) + o_p\left(1\right)$$

(see e.g. Newey and Smith, 2004). A mean value expansion, Assumption 7, similar arguments to those used in (A-4) and the uniform law of large numbers (implied by Assumptions 3 and

4(c)) show that

$$\frac{T^{1/2}}{\omega_1}\widehat{m}_s(\widehat{\beta},\widehat{h}) = \frac{T^{1/2}}{\omega_1}\widehat{m}_s(\beta_0,\widehat{h}) + \frac{\partial \widehat{m}_s(\overline{\beta},\widehat{h})}{\omega_1 \partial \beta^T} A\left(\beta_0, g_0\right)^{-1} \frac{1}{T^{1/2}} \sum_{t=1}^{n} f_t\left(\beta_0, g_0\right) + o_p\left(1\right)$$

$$= T^{1/2}\widehat{m}(\beta_0,\widehat{h}) + M\left(\beta_0, h_0\right) A\left(\beta_0, g_0\right)^{-1} \frac{1}{T^{1/2}} \sum_{t=1}^{n} f_t\left(\beta_0, g_0\right) + o_p\left(1\right).$$

Simple algebra shows that

$$T^{1/2}\widehat{m}(\beta_0,\widehat{h}) = [\widehat{v}_T(\widehat{h}) - \widehat{v}_T\left(h_0\right)] + T^{1/2}\widehat{m}(\beta_0,h_0) + T^{1/2}E[m_t(\beta_0,\widehat{h})],$$

where $\widehat{v}_T\left(h\right) = T^{-1/2}\sum_{t=1}^{T}\left\{m_t\left(\beta_0,h\right) - E\left[m_t\left(\beta_0,h\right)\right]\right\}$, so that by Assumptions 4(b), 5(a) and a standard central limit theorem for α-mixing random vectors (Doukhan, 1994), we have

$$T^{1/2}\widehat{m}(\beta_0,\widehat{h}) \xrightarrow{d} \left(0, K\left(\beta_0, h_0, g_0\right)\right),$$

hence the conclusion follows by the continuous mapping theorem and standard results on quadratic forms of normal random vectors (see e.g. White, 1994). Under Assumption 6, a further Taylor expansion shows that

$$m_t(\widehat{\beta},\widehat{h}) = m_t(\beta_0,h_0) + \sup_{\beta\in B_\delta} \partial_\beta m_t(\beta,h)\left(\widehat{\beta}-\beta_0\right) + \partial_h m_t\left(\beta_0,h_0\right)(\widehat{h}-h_0) +$$
$$\frac{1}{2}\int_0^1 \partial^2_{hh} m^h_t(\beta_0, h_0 + \xi(\widehat{h}-h_0))d\xi,$$

where $\partial^2_{hh} m^h_t\left(\cdot\right) = \sum_{j=1}^{l_h}(\widehat{h}-h_0)_j \partial^2_{hh_j} m_t\left(\cdot\right)(\widehat{h}-h_0)$. Using the linear representation of $\widehat{h}-h_0$ given in Assumption 6(a) ,we have that

$$\frac{T^{1/2}}{\omega_1}\widehat{m}_s(\beta_0,\widehat{h}) = T^{1/2}\widehat{m}(\beta_0,h_0) + \frac{1}{T^{3/2}\omega_1}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)\partial_h m_{t-s}\left(\beta_0,h_0\right)\times \tag{A-7}$$
$$\sum_{\tau=1,\tau\neq t}^{T}\Phi_T\left(z_{2t},z_{2t-\tau}\right)\odot\phi\left(z_t\right) + \frac{1}{T^{3/2}\omega_1}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)r_T\left(z_{2t-s}\right) +$$
$$\frac{1}{T^{3/2}\omega_1}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)\int_0^1 \partial^2_{hh} m_{t-s}(\beta_0, h_0+\xi(\widehat{h}-h_0))d\xi.$$

Assumption 6(a) implies that

$$\left\|\frac{1}{T^{3/2}\omega_1}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)r_T\left(z_{2t-s}\right)\right\| \leq \frac{1}{s_T\omega_1}\sum_{s=1-T}^{T-1}\left|\omega\left(\frac{s}{s_T}\right)\right|\frac{\left\|r_T\left(z_{2t-s}\right)\right\|_{\mathcal{H}}}{T^{1/2}} = o_p(1), \tag{A-8}$$

whereas by the uniform law of large numbers $\left\|\sup_{h\in\mathcal{H}_\delta}\partial^2_{hh}\widehat{m}\left(\beta_0,h\right)-E[\sup_{h\in\mathcal{H}_\delta}\partial^2_{hh}m_t\left(\theta_0,h\right)]\right\|=o_p\left(1\right)$, which implies that

$$\left\|\frac{1}{T^{3/2}\omega_1}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)\int_0^1\partial^2_{hh}m_{t-s}(\beta_0,h_0+\xi(\widehat{h}-h_0))d\xi\right\|\leq \tag{A-9}$$
$$\frac{1}{T^{1/2}s_T\omega_1}\sum_{s=1-T}^{T-1}\left|\omega\left(\frac{s}{s_T}\right)\right|\frac{1}{T}\sum_{t=1}^{T}\left|\int_0^1(1-\xi)\partial^2_{hh}m_t\left(\beta_0,\xi\left(h-h_0\right)\right)d\xi\right|=O_p\left(T^{-1/2}\right).$$

Note that

$$\frac{1}{T^{3/2}\omega_1}\sum_{t=1}^{T}\frac{1}{s_T}\sum_{s=1-T}^{t-1}\omega\left(\frac{s}{s_T}\right)\partial_h m_{t-s}\left(\beta_0,h_0\right)\sum_{\tau=1,\tau\neq t}^{T}\Phi_T\left(z_{2t},z_{2t-\tau}\right)\odot\phi\left(z_t\right)=$$
$$\frac{1}{s_T\omega_1}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)\frac{1}{T^{3/2}}\sum_{t=\max(1,1-s)}^{\min(T,T-s)}\sum_{\tau=1,\tau\neq t}^{T}\partial_h m_t\left(\beta_0,h_0\right)\Phi_T\left(z_{2t},z_{2t-\tau}\right)\odot\phi\left(z_t\right)=$$
$$\frac{1}{s_T}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)U_{T,s},$$

and that the difference between $U_{T,s}$ and $U_T=\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\partial_h m_t\left(\beta_0,h_0\right)\Phi_T\left(z_{2t},z_{2t-\tau}\right)\odot\phi\left(z_t\right)/T^{3/2}$ consists of s terms. The Markov inequality and Assumption 6(b) yield

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

so that again by Markov inequality and Assumption 2(a)

$$P\left(\left|\frac{1}{s_T}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)\left\{U_{T,s}-U_T\right\}\right|>\epsilon\right)\leq\frac{1}{\epsilon}\frac{1}{s_T}\sum_{s=1-T}^{T-1}\omega\left(\frac{s}{s_T}\right)E\left|U_{T,s}-U_T\right| \tag{A-10}$$
$$\leq CT^{\delta}\frac{1}{s_T}\sum_{s=1-T}^{T-1}\frac{|s|}{T}\left|\omega\left(\frac{s}{s_T}\right)\right|=O(T^{\delta-\eta-1/2})=o(1).$$

Combining (A-8), (A-9) and (A-10) we have that (A-7) can be written as

$$\frac{T^{1/2}}{\omega_1}\widehat{m}_s(\beta_0,\widehat{h})=T^{1/2}\widehat{m}(\beta_0,h_0)+T^{1/2}U_T^*\left(\beta_0,h_0\right)+o_p\left(1\right),$$

where $U_T^*(\beta_0, h_0)$ can be represented as a second order U-statistic with a varying symmetric kernel, that is

$$U_T^*(\beta_0, h_0) = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1, s\neq t}^{T} \left(\widetilde{\Phi}_T(z_{2s}, z_{2t}) + \widetilde{\Phi}_T(z_{2t}, z_{2s}) \right) \tag{A-11}$$

where $\widetilde{\Phi}_T(z_{2s}, z_{2t}) = \partial_h m_t(\beta_0, h_0)\, \Phi_T(z_{2s}, z_{2t}) \odot \phi(z_s)$. The asymptotic normality of $T^{1/2}\widehat{m}_s(\beta_0, \widehat{h})/\omega_1$ follows by the continuous mapping theorem, combining the asymptotic normality of $T^{1/2}\widehat{m}(\beta_0, h_0)$ with the asymptotic normality of $T^{1/2}U_T^*(\beta_0, h_0)$, which follows by a central limit theorem for second order degenerate or nondegenerate U-statistics of α mixing random vectors (see e.g. Bravo, Chu, and Jacho-Chávez, forthcoming), hence

$$T^{1/2}\widehat{m}_s(\widehat{\beta}, \widehat{h})/\omega_1 \xrightarrow{d} N\left(0, K^e(\beta_0, h_0, g_0)\right).$$

The conclusion follows by the continuous mapping theorem and standard results on the distribution of quadratic forms of normal random vectors (see e.g. White, 1994). ■

Proof of Theorem 3.2: The same arguments used in the proof of Theorem 3.1 imply that under the local hypothesis (3.5)

$$T^{1/2}\widehat{m}_s(\widehat{\beta}, \widehat{h})/\omega_1 \xrightarrow{d} N\left(\delta, K^{\times}(\beta_0, h_0, g_0)\right),$$

and first conclusion follows by the quadratic approximation A-6, the continuous mapping theorem and standard result on quadratic forms of nonzero mean normal random vectors (see e.g. White, 1994). ■

Proof of Theorem 3.3: By the same arguments as those used to show A-1

$$\left\| \widehat{m}_s(\widehat{\beta}, \widehat{h}) - \omega_1 E\left[m_t\left(\overline{\beta}, \overline{h}\right)\right] \right\| = o_p(1),$$

and the conclusion follows by continuous mapping theorem as $\widehat{\rho}^{\times}(\widehat{\beta}, \widehat{h}, \widehat{g})D_T/T$ and $LM_T/T = O_p(1)$. ■

Proof of Proposition 4.1: We verify Assumptions 1- 4 and 6-7. Assumptions 1 and 2 are assumed in (a) and (b); the linearity in both θ and h imply that Assumption 3 is stronger than necessary and can be replaced by the moment Assumption (d). Assumption (c) is sufficient for using the uniform consistency results of kernel estimators of Andrews (1995) to imply Assumptions 4(a) and 4(c) for an appropriate choice of the bandwidth. The stochastic equicontinuity Assumption 4(b) follows directly by the results of Andrews (1994b). Assumptions 6 (a) and (b) hold with $\Phi_T(z_t, w) = \Phi_T(x_{2t}, x) = f_{x_{2t}}(x)\, K_{b_T}(x_{2t} - x)$, $\phi(z_t) = y_t - x_{1t}'\theta_0$ using a standard kernel calculation, whereas Assumption 6(c) is not necessary. Finally Assumption 7 follows by (c), (e), (f), standard algebra of least square estimation and the uniform consistency of kernel estimators. ■