Asymptotically Nuisance-Parameter-Free Consistent Tests of L_p-Functional Form

Jonathan B. Hill*
Dept. of Economics
Florida International University

July 9, 2006

Abstract

We develop a consistent conditional moment test of L_p-best predictor functional form, $1 < p \leq 2$. Our main result is a reduction of the nuisance parameter space to the set of integers which greatly simplifies asymptotic theory, and allows for removal of the nuisance parameter in a mechanical fashion. Our results provide a fresh vantage into why Bierens' (1990) moment condition works, and uncovers a new class of weights which sharply contrasts with Stinchcombe and White's (1997) weight classification (real analytic and non-polynomial). The computation of a weighted-Average CM statistic is easy and asymptotically nuisance parameter free because it incorporates all possible nuisance parameter values. Our test serves as a consistent model check in L_p-regression environments. Finally, we provide a simple nuisance parameter free series expansion of the best L_p-predictor.

*Dept. of Economics, Florida International University, Miami, Fl; jonathan.hill@fiu.edu; http://www.fiu.edu/~hilljona

JEL classification: C12, C45, C52.

Key words: nonlinear regression models; consistent conditional moment test; nuisance parameter-free test; L_p-best predictor.

1. Introduction Parametric conditional moment (CM) tests of functional form based on a finite number of L_2-orthogonality conditions, cf. Ramsey (1969, 1970), White (1981), Newey (1985) and Tauchen (1985), are known in general not to be consistent against every alternative. Apparently the only consistent parametric CM tests are those of Bierens (1982, 1984, 1987, 1990), de Jong (1996), and the Integrated CM test of Bierens (1982) and Bierens and Ploberger (1997). See, also, White (1989), de Jong and Bierens (1994), and Corradi and Swanson (2002) for related methods. Consistency is apparently achieved by generating weight functions $F(\tau' x_t)$ indexed by a real-valued nuisance vector $\tau \in \Xi \subset \mathbb{R}^k$, effectively producing uncountably many moment conditions which "reveal" model mis-specification. Stinchcombe and White (1998) show that any real analytic function $F(A(x_t))$ that is non-polynomial can reveal model mis-specification, where $A : \mathbb{R}^k \rightarrow \mathbb{R}$ is affine. For notation conventions, see Section 2.

Although much has been said about the subset $S \subset \Xi$ on which consistent tests fail, very little has been said about noteworthy subsets of the remaining "revealing" points Ξ/S. The extant literature argues test consistency requires the nuisance parameter space Ξ to have positive Lebesgue measure and therefore contain uncountably infinitely many elements: see Bierens (1990: Lemma 1), Bierens and Ploberger (1997: Theorem 1), and Stinchcombe and White (1998: Theorems 2.2 and 2.3).

Stinchcombe and White (1998: p. 298) claim a "*remarkable feature of Bierens' approach is that a* smooth *random choice of* τ*...will deliver a consistent test.*" Such a perspective neglects to consider the trivial decomposition $\tau = [\tau_i]_{i=1}^k = [m_i \delta_i]_{i=1}^k$ for some integer $m \in \mathbb{N}^k$ and $\delta \in \mathbb{R}^k$. It is worth wondering what roles m and δ play with respect to test power.

Hansen (1996: p. 415, 419) laments the "*unpleasant dilemma*" of selecting τ arbitrarily or in a data dependent way; the extensive costliness of selecting τ from a continuous parameter space Ξ; and the necessary diminishment of power when a discreet approximation to Ξ is used (which is always the case, in practice). The source of the dilemma is the assumption that τ is selected from a set with uncountably infinitely many elements.

The ICM test of Bierens (1982) and of Bierens and Ploberger (1997) solves the choice problem by integrating the sample moment over a subset Ξ with positive Lebesgue measure. A discreet approximation to Ξ is required in practice. Moreover, the integration involves a probability measure weight that only incorporates information from the nuisance vector τ and not the actual magnitude of the sample moment evaluated at τ.

In this paper we develop CM tests of best L_p-predictor functional form that are consistent against any deviation from the null specification under a class of $\sqrt{n}$-local alternatives. We consider L_p-best prediction because the L_2-best predictor is hardly the only object of interest. Indeed, L_p-regression, L_p-GMM, M-estimation, non-Hilbertian metric projection and impulse response analysis provide important alternatives to canonical L_2-methods for both *iid* and time series data. See Koul and Zhu (1995), Arcones (1996), Koul (1996), Bantli and Hallin (2001), de Jong and Han (2002), Liebscher (2003), Cheng and De Gooijer (2005), Lai and Lee (2005), Hill (2006) and Wu (2006).

Our main contribution is a reduction of the nuisance parameter space to the set of integers. We effectively present an alternative interpretation of the power of the Bierens test: test consistency is not predicated on a smooth choice of τ, *per se*, but for *any* non-zero $\delta \in \mathbb{R}^k$ there exist infinitely many integers $m \in \mathbb{Z}^k$ such that $[\tau_i]_{i=1}^k = [m_i \delta_i]_{i=1}^k$ generates a consistent test. We provide fresh perspectives on why Bierens' (1990) exponential moment condition works, and uncover an infinitely large class of "totally revealing" weights that does not nest the class of real analytic, non-polynomial weights characterized by Stinchcombe and White (1997). Indeed, our weight functionals need not be differentiable nor, therefore, analytic.

A weighted "Average CM" test can be computed mechanically over an increasing integer subset. Asymptotic theory is greatly simplified because distribution tightness requirements are automatically satisfied. Our test provides a consistent model check for L_p-regression models of L_p-best predictors, $1 < p \leq 2$. Moreover, we use data-driven weights that place more weight on large sample moments.

Furthermore, our theory allows for a simple, asymptotically nuisance-parameter-free, L_p-norm convergent series expansion of the best L_p-predictor. This provides a simple plug-in for a consistent non-parametric test of L_p-functional form. See Lee (1988), Yatchew (1992), Hong and White (1995), Zheng (1996), Dette (1999) and Li *et al* (2003) for related non-parametric methods in L_2.

We only consider the finite dimensional case for brevity. See de Jong (1996) for an infinite dimensional extension of the Bierens (1991) test. Finally, the L_1-case involves known difficulties which deviate from the fundamental objectives of this paper. We leave this case for future consideration.

In Section 2 we construct a basic vector moment condition, and develop an integer indexed conditional moment in Section 3. Section 4 presents the Average CM test, and Section 5 concludes with a monte carlo study. Assumptions can be found in Appendix 1, all proofs are left for Appendix 2, and all tables are placed at the end of the paper.

Throughout $\rightarrow$ denotes convergence in probability or in finite dimensional distributions; $\Rightarrow$ denotes weak convergence. $|\cdot|_p$ denotes the l_p-norm for real-valued vectors or matrices, and $||\cdot||_p$ denotes the corresponding L_ρ-norm: $||x||_p = (\Sigma_{i,j} E|x_{i,j}|^p)^{1/p}$. Vector powers x^a are understood to represent $(x_1^{a_1}, ..., x_k^{a_k})'$, and $x/a = (x_i/a_i)_{i=1}^k$, $\forall \{a_i \neq 0\}_{i=1}^k \in \mathbb{R}^k$. I_k denotes a k-dimensional identity matrix and 0_k and 1_k denote k-vectors of $0's$ and $1's$. $z^{<a>}$ denotes the signed power: $sign(z) \times |z|^a$. $x \perp_p y$ *if and only if* $Ex^{<p-1>}y = 0$. $sp(\{z_i\}_{i=1}^\infty)$ denotes the span of $\{z_i\}_{i=1}^\infty$ and $\overline{sp}(\{z_i\}_{i=1}^\infty)$ the closed linear span. Let $0 \in \mathbb{N}^k$ (i.e. $\mathbb{N}^k = \mathbb{Z}_+^k$).

2. Conditional Moments

A standard preliminary result concerning "revealing" vector moments under model mis-specification is contained in Lemma 1.

Let $\{y_t, \tilde{x}_t\} \in \mathbb{R} \times \mathbb{R}^{k-1}$ be a strictly stationary, ergodic stochastic process in

$L_p(\ ,\Im,\mu)$, $p \in (1,2]$, with nondegenerate continuous marginal distributions, $\Im = \sigma(\cup_t \Im_t)$, $\Im_{t-1} \subseteq \Im_t = \sigma(\{x_\tau\} : \tau \leq t+1)$. Define $x_t \equiv (1, \tilde{x}_t')'$. The regressors x_t may contain lags of y_t as well as contemporary and lagged values of some other vector process.

Let $f_t(\phi) = f(x_t, \phi)$ denote a known response function, $f_t : \mathbb{R}^k \times \Phi \to \mathbb{R}$, measurable with respect to $\Im_{t-1}$, with Φ a compact subset of $\mathbb{R}^k$. Consult Appendix 1 for all assumptions detailed under Assumption A.

We aim for the greatest generality in order to permit consistency and asymptotic normality of the L_p-regression estimator $\hat{\phi} = \arg\min_{\phi\in\Phi}\{\sum_{t=1}^n |y_t - f_t(\phi)|^p\}$ for dependent, heterogeneous data $\{y_t, \tilde{x}_t\}$. Consult Assumption A. There exists a substantial literature on the topic of L_p-regression of linear models $f(x_t, \phi) = \phi' x_t$ (e.g. Koul and Zhu 1995, Bantli and Hallin 2001, Lai and Lee 2005, and Wu 2006), and a comparatively small literature for *nonlinear* models (e.g. Koul 1996, Cheng and De Gooijer 2005, and Liebscher 2003).

Denote by $Q_{t-1}y \equiv Q(y_t|\Im_{t-1})$ the orthogonal L_p-metric projection of y_t onto the space spanned by $\{x_{t-i}\}_{i=0}^{\infty}$. The operator Q is orthogonal: $Q_{t-1}z_t = 0\ \forall z_t \perp_p sp(\{x_{t-i}\}_{i=0}^{\infty})$; quasi-linear: $Q_{t-1}(z_t + w_t) = Q_{t-1}(z_t) + w_t\ \forall w_t \in sp(\{x_{t-i}\}_{i=0}^{\infty})$; conditional expectations: $Q_{t-1}y_t = E[y_t|\Im_{t-1}]$ sufficiently if $p = 2$; and Q_{t-1} generates a moving average decomposition with strong orthogonal innovations $\{z_t\}$ (i.e. $z_t \perp_p sp(\{x_{t-i}\}_{i=0}^{\infty}$ and $\sum_{i=0}^{h-1} z_{t-i} \perp_p sp(\{x_{t-i}\}_{i=h}^{\infty}\ \forall h \geq 0)$ *if and only if* Q iterates (i.e. $Q_{t-j}Q_{t-i}y_t = Q_{t-j}y_t$ $\forall j \geq i \geq 0$). Consult Lindenstrauss and Tzafriri (1977), Megginson (1998), and Hill

(2006).

Write

$$e_t = \epsilon_t^{<p-1>} = (y_t - Q_{t-1}y_t)^{<p-1>}.$$

Clearly e_t satisfies

$$E[e_t z_{t-1}] = 0 \ \forall z_{t-1} \in sp(\{x_{t-i}\}_{i=0}^{\infty}).$$

The fundamental hypotheses are

$$H_0 : P\left(Q\left(y_t|\Im_{t-1}\right) = f(x_t, \phi_0)\right) = 1, \text{ for some } \phi_0 \in \Phi$$

$$H_1 : \sup_{\phi\in\Phi} P\left(Q\left(y_t|\Im_{t-1}\right) = f(x_t, \phi_0)\right) < 1.$$

Under H_0 there exists some set ϕ_0 such that $f(x_t,\phi_0)$ is *almost surely* correctly specified as the best L_p-predictor of y_t, and e_t forms a martingale difference sequence: $E[e_t|\Im_{t-1}] = 0$. H_1 embraces any deviation from the null.

Stinchcombe and White (1998: Theorem 2.3) expand upon Bierens' (1990) Lemma 1 for the best L_2-predictor $E[y_t|\Im_{t-1}]$ by considering the class of functions

$$\mathcal{H}_F = \{g : \mathbb{R}^k \to \mathbb{R} \mid g(x) = F(A(x)),\ A \text{ affine},\ F : \mathbb{R} \to \mathbb{R}\}.$$

The authors prove that any analytic member $F \in \mathcal{H}_F$ has the desired "generically totally revealing" property *if and only if* F is non-polynomial.

Assumption B Let $F \in \mathcal{H}_F$. Assume F is analytic and non-polynomial on some open interval $R_0 \subset \mathbb{R}$. Assume $(\partial/\partial u)^i F(u)|_{u=0} = 0$ for only finitely many $i \in \mathbb{N}$. Let 0 lie in the interior of R_0.

Remark 1: We use the assumptions $(\partial/\partial u)^i F(u)|_{u=0} = 0$ for finitely many $i \in \mathbb{N}$, and $0 \in \text{interior}(R_0)$, in the main result Theorem 3.

Remark 2: That the available set of functions $F(\cdot)$ is limited under Assumption B is irrelevant for the main results of the paper.

Let $h : \mathbb{R}^k \times \Delta \rightarrow \mathbb{R}^k$ be a uniformly bounded, $\mathfrak{F}_{t-1}$-measurable function, $k \geq 1$, where Δ is an arbitrary subset of $\mathbb{R}^l$, $l \geq 0$. Write $h_t(\delta) = h(x_t, \delta)$. The following is a required, although easy, extension of Lemma 1 of Bierens (1990) and Theorem 1 of Bierens and Ploberger (1997).

LEMMA 1 *Let e_t be a random variable satisfying $E|e_t| < \infty$, and let x_t be an $\Im_{t-1}$-measurable bounded vector in $\mathbb{R}^k$ such that $P(E[e_t|x_t] = 0) < 1$. Let Assumption B hold. For each $\delta \in \mathbb{R}^l$ and each $i = 1...k$, the sets*

$$S_i = \left\{\tau \in \mathbb{R}^k : E[e_t h_{t,i}(\delta) F(\tau' x_t)] = 0\right\} \text{ and } P\left(\tau' x_t \in R_0\right) = 1\right\},$$

have Lebesgue measure zero, and are nowhere dense in $\mathbb{R}^k$.

Remark 1: The sets S_i will depend on the distribution of $\{y_t, x_t\}$, and on each point $\delta \in \mathbb{R}^l$.

Remark 2: The resulting set $S \equiv \cap S_i$, the collection of each τ such that the vector $E[e_t h_t(\delta) F(\tau' x_t)] = 0$ has Lebesgue measure zero under H_1.

Remark 3: Conditioning on x_t is equivalent to conditioning on any bounded, measurable, one-to-one function of x_t , $\Psi(x_t) : \mathbb{R}^k \rightarrow \mathbb{R}^k$, since any such functional induces the same σ-field as x_t : see Billingsley (1995: Theorem 5.1). In this case x_t

need not be bounded, cf. Bierens (1991).

3. Main Results

A preliminary result is contained in Lemma 2. The main result of the paper is contained in Theorem 3. Define

$$\Delta = \{\delta = [\delta_0 \vdots \delta_1] \in \mathbb{R}^{k\times 2} : \delta_0 \in \mathbb{R}^k,\ \delta_{1,i} \neq 0,\ i = 1...k\}.$$

3.1 Preliminary Result

Let $F(\cdot)$ satisfy Assumption B, and for any $\delta \in \Delta$ write

$$x_t(\delta) \equiv (\delta_{0,1} + \delta_{1,1}x_{1,t}, ..., \delta_{0,k} + \delta_{1,k}x_{k,t})'.$$

Consider any bounded one-to-one mapping $\Psi : \mathbb{R}^k \to \mathbb{R}^k$. For example $\Psi(x_t(\delta)) = (\exp(x_{1,t}(\delta)), ..., \exp(x_{k,t}(\delta)))'$ if x_t is bounded.

Define the set

$$T^*(\Psi(x_t(\delta)) = \{\tau \in \mathbb{R}^k : P(\tau'\Psi(x_t(\delta)) \in R_0) = 1\},$$

the set of τ such that $\tau'\Psi(x_t(\delta)$ almost surely obtains values on the interval on which $F(\cdot)$ is non-polynomial and analytic. In the exponential $F(u) = \exp\{u\}$ and logistic $F(u) = [1 + \exp\{u\}]^{-1}$ cases, $T^*(\Psi(x_t(\delta)) = \mathbb{R}^k$. Under Assumption B $0 \in T^*(\Psi(x_t(\delta))$ because $0'\Psi(x_t(\delta) \in R_0$.

Write $F^s(\cdot) \equiv (\partial/\partial u)^s F(\cdot)$ for any $s \in \mathbb{N}$. By convention $F^0(\cdot) = F(\cdot)$.

LEMMA 2 *Let e_t be a random variable satisfying $E|e_t| < \infty$, and let x_t be an F_{t-1}-measurable bounded vector in $\mathbb{R}^k$ such that $P[E(e_t|x_t) = 0] < 1$. Let*

Assumption B hold. For each point $\delta \in \Delta$ *and every* $\tau \in T^*(\Psi(x_t(\delta))$ *there exists infinitely many vectors* $m \in \mathbb{Z}^k$*, and for each* m *some scalar integer* $\tilde{s} \geq 0$*, such that*

$$(1) \quad E\left[\epsilon_t \prod_{i=1}^{k} \Psi_i(x_t(\delta))^{m_i} F^{\tilde{s}}(\tau' \Psi(x_t(\delta)))\right] \neq 0.$$

In particular, $\forall r_0 \in \mathbb{Z}^k$ *and any* $m_0 \in \mathbb{Z}^k$*,* $m_0 \geq r_0$*, (1) holds* $\forall m \geq m_0$*.*

Remark: Although there are infinitely many integer vectors m that satisfy (1), there is not necessarily a unique integer $\tilde{s}$ for each m.

3.2 Main Result

We can always set $r_0 = 0$ in Lemma 2 to ensure $m \geq 0$. Moreover, (1) holds for any $\delta \in \Delta$ and every $\tau \in T^*(\Psi(x_t(\delta))$, therefore it holds for $\delta = [0_k \vdots 1_k]$ and $\tau = 0$, cf. Assumption B.

This suggests the following class of weights:

$$\mathcal{H}_{G(m,\delta)} = \{g : \mathbb{R}^k \to \mathbb{R} \mid g(\Psi(x(\delta))^m) = \sqcap_{i=1}^{k} \Psi_i(x(\delta))^{m_i},\ \delta \in \Delta,$$
$$\Psi : \mathbb{R}^k \to \mathbb{R}^k,\ \Psi \text{ is bounded, one-to-one, } m \in \mathbb{Z}^k\}.$$

Notice if $G_t(m, \delta) = G(\Psi(x_t(\delta))^m) \in \mathcal{H}_{G(m,\delta)}$ then $G_t(0, \delta) = 1$ *a.s.*

When $\delta = [0_k \vdots 1_k]$ we write

$$\mathcal{H}_{G(m)} = \left\{g \in \mathcal{H}_{G(m,\delta)} : \delta = [0_k \vdots 1_k]\right\}$$

with elements $G_t(m) = G(\Psi(x_t)^m) \in \mathcal{H}_{G(m)}$. The following two results are immediate.

THEOREM 3 *Let* e_t *be a random variable satisfying* $E|e_t| < \infty$, *let* x_t *be an* $\Im_{t-1}$-*measurable bounded vector in* $\mathbb{R}^k$ *such that* $P[E(e_t|x_{t-1}) = 0] < 1$. *If* $G_t(m, \delta) = G(\Psi(x_t(\delta))^m) \in \mathcal{H}_{G(m,\delta)}$ *then*

$$E\left[e_t G_t(m, \delta)\right] \neq 0$$

for any $\delta \in \Delta$ *and infinitely many* $m \in \mathbb{Z}^k$ *in general, and specifically for infinitely many* $m \in \mathbb{N}^k$.

Remark: Suppose $e_t = (y_t - f_t(\phi))^{<p-1>}$. If $E\left[e_t G_t(m, \delta)\right] = 0 \; \forall m \in \mathbb{Z}^k$ and any $\delta \in \Delta$ then $P[E(e_t|x_{t-1}) = 0] = 1$ must hold, hence $Q\left(y_t|\Im_{t-1}\right) = f(x_t, \phi_0)$ *a.s.* In fact, Theorem 3 implies we need only consider $\mathbb{N}^k$: if $E\left[e_t G_t(m,\delta)\right] = 0 \; \forall m \in \mathbb{N}^k$ then $Q\left(y_t|\Im_{t-1}\right) = f(x_t, \phi_0)$ *a.s.*

Examples of weights $G_t(m)$ satisfying Theorem 3 are easily to generate.

COROLLARY 4 *Under the conditions of Theorem 3 if* $P[E(e_t|x_t) = 0] < 1$ *then*

$$E[e_t \sqcap_{i=1}^k x_{t,i}^{m_i}] \neq 0, \; E[e_t \exp\{m'x_t\}] \neq 0, \text{ and } E[e_t \exp\{m'x_t(\delta)\}] \neq 0,$$

for any $\delta \in \Delta$ *and infinitely many* $m \in \mathbb{N}^k$.

Remark 1: The moment $E[e_t \exp\{\tau' x_t\}]$ considered in Bierens (1990) is simply a special case of $E[e_t \exp\{m'x_t(\delta)\}]$ with fixed $m = 1_k$.

Remark 2: Because each weight $G_t(m, \delta)$ is a multiplicative transform of a one-to-one vector function of $x_t(\delta)$, we can always define $m = (a, \tilde{m}')'$, $a \in \mathbb{R}$ and $\tilde{m} \in \mathbb{N}^{k-1}$ whenever x_t contains a constant term.

Remark 3: The result $E\left[e_t \sqcap_{i=1}^k x_{t,i}^{m_i}\right] \neq 0$ for *infinitely many* $m \in \mathbb{N}^k$ under H_1 generalizes Bierens' (1982) proof that $E\ [e_t \sqcap_{i=1}^k x_t^{m_i}] \neq 0$ for *some* $m \in \mathbb{N}$.

Remark 4: Theorem 3 provides further support for the practice of adding products and cross-products to regression models in order to improve model fit. Cf. Gallant and Souza (1991).

The facts that Bierens' chosen weight $\exp\{u\}$ is analytic and non-polynomial, and $\exp\{\tau' x_t\}$ exploits a "*smooth choice*" of $\tau \in \mathbb{R}^k$, are apparently immaterial. We are only concerned with power-products $\sqcap_{i=1}^k \Psi_i(x_t(\delta))^{m_i}$ of bounded one-to-one functions Ψ, and $\Psi(x_t(\delta))$ need not be analytic because it need not be differentiable[1]. Moreover, δ is irrelevant for consistency as long as $\delta \in \Delta$ (i.e. non-zero weight is placed on $x_{t,i}$). Thus $\exp\{\tau' x_t\} = \exp\{m' x_t(\delta)\} = \sqcap_{i=1}^k \exp\{m_i \delta_i x_{t,i}\} = \sqcap_{i=1}^k \Psi_i(x_t(\delta))^{m_i}$ delivers a consistent test for countably infinitely many integers $m \in \mathbb{N}^k$.

Stinchcombe and White's (1997) class $\mathcal{H}_F$ is not nested within $\mathcal{H}_{G(m,\delta)}$. The multiplicative logistic, for example,

$$\prod_{i=1}^k [1 + \exp\{\delta_{0,i} + \delta_{1,i} x_{t,i}\}]^{-m_i}$$

is an element of $\mathcal{H}_{G(m,\delta)}$, and cannot be represented as $F(A(x)) : \mathbb{R} \rightarrow \mathbb{R}$ for *affine* A if $k > 1$. That said, the standard logistic in general $[1 + \exp\{m' x_t(\delta)\}]^{-1} \notin \mathcal{H}_{G(m,\delta)}$ even if $k = 1$.

Stinchcombe and White (1998: Lemma 3.5) exploit results in Hornik (1991) in

[1]Stinchcombe and White (1998: Theorem 3.10) characterize revealing functionals $F(\cdot)$ that are non-real analytic. However, they still require $F(\cdot)$ to be infinitely differentiable.

order to rule out q-order polynomials because they are not comprehensive. However, it is straightforward to show that $(\tau' x_t)^q$ for any $q \in \mathbb{N}$ is simply a version of $\sqcap_{i=1}^k(\delta_{0,i} + \delta_{1,i} x_{t,i})^{m_i}$ for some $m \in \mathbb{N}^k$. Therefore $(\tau' x_t)^q \in \mathcal{H}_{G(m,\delta)}$.

3.3 Bounding the Weight

For test computation purposes a bounded weight $G_t(m, \delta)$ may be desirable given $m \in \mathbb{Z}^k$ is unrestricted. For notational simplicity fix $\delta = [0_k \vdots 1_k]$. Weights $G_t(m) \in \mathcal{H}_{G(m)}$ with trivial bounds include

$$\exp\left\{m'\tilde{x}_t / \max_{1 \le i \le q}\{|\tilde{x}_{t,i}|\}\right\} = o_p(2^{\Sigma_{j=1}^k m_j/2})$$

$$\prod_{i=1}^q \left(1 + \tilde{x}_{t,i} / \max_{1 \le i \le q}\{|\tilde{x}_{t,i}|\}\right)^{m_i} = o_p(2^{\Sigma_{j=1}^k m_j/2}).$$

Weights that are $O_p(2^{-\Sigma_{j=2}^k m_i})$ are also easy to construct. Let $m = (m_1, \tilde{m}')'$, $m_1 \in \mathbb{R}$ and $\tilde{m} \in \mathbb{N}^{k-1}$. An argument identical to Lemma 1 shows Theorem 3 holds for any $e_t h_t(\xi)$, $\xi \in \mathbb{R}^{k-1}$. Consider $h_t(\xi) = \exp\{\xi' \tilde{x}_t^2\}$ and $G_t(m) = \exp\{m' x_t\}$. Now re-parameterize: define $\gamma_i \equiv -\xi_i > 0$ and $c_i \equiv \tilde{m}_i / (\xi_i 2)$ for each $i = 1 \ldots k-1$, and fix $m_1 = \sum_{i=1}^{k-1} \tilde{m}_i^2 / (\xi_i 4)$. Then $P[E(e_t|x_t) = 0] < 1$ implies for each $\xi < 0$

$$(2) \quad E[e_t \exp\{-\xi' \tilde{x}_t^2\} \exp\{m' x_t\}] = E[e_t \exp\{-\gamma' (\tilde{x}_t - c)^2\}] \neq 0$$

for countably infinitely many $c = \tilde{m}/(\xi 2)$, $\tilde{m} \in \mathbb{N}^{k-1}$. Simply pick, say, $\gamma_i = -\xi_i = -1/2$.

COROLLARY 5 *Under the assumptions of Theorem 3, if* $P(E[e_t|x_t] = 0) < 1$, *then*

$$E[e_t \exp\{-.5 \times \Sigma_{i=1}^{k-1} (\tilde{x}_{t,i} - \tilde{m}_i)^2\}] \neq 0$$

for infinitely many $\tilde{m} \in \mathbb{N}^{k-1}$*, where* $\exp\{-.5 \times \Sigma_{i=1}^{k-1}(\tilde{x}_{t,i} - \tilde{m}_i)^2\} = O_p(2^{-\Sigma_{j=2}^{k} m_i})$.

Remark: It is easy to show $\sqcap_{i=1}^{k}[\exp\{-|x_{t,i}|\} \times sign(x_{t,i})]^{m_i}$,

$$\prod_{i=1}^{k}([1 + |x_{t,i}|] \times sign(x_{t,i}))^{-m_i} \text{ and } \sup_{\delta \in \Delta} \prod_{i=1}^{k}[1 + \exp\{\delta_{0,i} + \delta_{1,i}x_i\}]^{-m_i}$$

are also $O_p(2^{-\Sigma_{j=2}^{k} m_i})$ elements of $\mathcal{H}_{G(m)}$.

3.4 Best L_p-Predictor

The remark following Theorem 3 implies the best L_p-predictor of y_t is an element of the closed linear span of $\{G_t(m)\}_{m \in \mathbb{N}^k}$, with probability one. Note $\overline{sp}(\{G_t(m)\}_{m \in \mathbb{N}^k})$ $= \overline{sp}(1, \{G_t(m)\}_{m \in \mathbb{N}^k})$ due to $G_t(0) = 1$.

THEOREM 6 *If* $G_t(m) = G(\Psi(x_t)^m) \in \mathcal{H}_{G(m)} \subset L_p$ *and* x_t *is bounded, then* $Q(y_t|\Im_{t-1}) \in \overline{sp}(\{G_t(m)\}_{m \in \mathbb{N}^k})$ *a.s. In particular, for some sequence of real numbers* $\{\beta_m\}_{m \in \mathbb{N}^k}$, $Q(y_t|\Im_{t-1}) = \sum_{m \in \mathbb{N}^k} \beta_m G_t(m)$ *a.s., where* $\sum_{m \in \mathbb{N}^k} \beta_m G_t(m)$ *is* L_p*-norm convergent.*

Remark 1: The assumption $\mathcal{H}_{G(m)} \subset L_p$ simply ensures $\overline{sp}(\{G_t(m)\}_{m \in \mathbb{N}^k}) \subset L_p$. If $G_t(m) = O_p(2^{-\Sigma_{j=2}^{k} m_i})$ then necessarily $G_t(m) \subset L_p$ by the $\Im_{t-1}$-measurability of x_t.

It is now an empirical matter whether a truncated version of $\sum_{m \in \mathbb{N}^k} \beta_m G_t(m)$ adequately approximates $Q(y_t|\Im_{t-1})$. In practice estimation of the unique set of parameters $\{\beta_m\}$ by L_p-regression is trivial, and asymptotically all nuisance parameters in $\mathbb{N}^k$ are incorporated.

4. Test Functional For convenience restrict $m \in \mathbb{N}^k$ and $\delta = [0_k \vdots 1_k]$. In this section we analyze weak convergence of a suitable sample moment on a functional space, and design a simple test functional. Let

$$p < \min\{1 + .5 \times \arg\inf\{\alpha > 0 : E|\epsilon_t|^\alpha < \infty\}, 2\}.$$

Hence $E|\epsilon_t|^{2(p-1)+\iota} < \infty$ for some $\iota > 0$.

Write $\hat{\phi} = \arg\min_{\phi\in\Phi}\{\sum |y_t - f_t(\phi)|^p\}$, define $\hat{\epsilon}_t \equiv y_t - f_t(\hat{\phi})$ and write $\partial f(\cdot) = (\partial/\partial\phi)f(\cdot)$. Define the sample moment

$$\hat{z}(m) = 1/\sqrt{n}\sum_{t=1}^{n} \hat{e}_t G_t(m), \quad \text{where } \hat{e}_t \equiv \hat{\epsilon}_t^{<p-1.}.$$

We use a Pitman $\sqrt{n}$-local alternative of the form

$$(3) \quad H_1^L : y_t = f_t(\phi_0) + u_t/\sqrt{n} + \epsilon_t,$$

where $Q_{t-1}\epsilon_t = 0$, and u_t is measurable with respect to $\Im_{t-1}$ and governed by a non-generate distribution.

Using the mean-value-theorem and Assumption A, under H_1^L for some sequence $\{u_t^*\}$, $u_t^* \in [0, u_t]$, $u_t^* = o_p(\sqrt{n})$, we may write

$$\begin{aligned}(4) \quad \hat{z}(m) &= 1/\sqrt{n}\sum_{t=1}^{n} \epsilon_t^{<p-1>} g_t(m) \\ &\quad + (p-1)1/n\sum_{t=1}^{n} |u_t^*/\sqrt{n} + \epsilon_t|^{p-2} u_t g_t(m) + o_p(1) \\ &= z_n(m) + o_p(1)\end{aligned}$$

say, where

$$(5) \qquad g_t(m) = G_t(m) - b(m, \phi_0)' A(\phi_0)^{-1} \partial f_t(\phi_0)$$

$$A(\phi_0) = (p-1)\,\mathrm{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} |y_t - f_t(\phi_0)|^{p-2} \partial f_t(\phi_0) \partial' f_t(\phi_0)$$

$$b(m, \phi_0) = (p-1)\,\mathrm{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} |y_t - f_t(\phi_0)|^{p-2} G_t(m) \partial' f_t(\phi_0)$$

Assumption A guarantees the following result.

THEOREM 7 *Let Assumption A hold, and let $G \in \mathcal{H}_{G(m)}$ where $G_t(m) = G(\Psi(x_t)^m)$ $= O_p(2^{-\Sigma_{j=2}^{k} m_j})$. Let $z(m)$ denote a Gaussian random variable with mean $\eta(m) = (p-1)\,\mathrm{plim}_{n\to\infty} 1/n \sum_{t=1}^{n} |\epsilon_t|^{p-2} u_t g_t(m) < \infty$ and variance $\gamma(m) = \mathrm{plim}_{n\to\infty} 1/n \sum_{t=1}^{n} |\epsilon_t|^{2(p-1)} g_t(m)^2 < \infty$. Under H_1^L, $\hat{z}(m) \to N(\eta(m), \gamma(m))$ in distribution pointwise in $m \in \mathbb{N}^k$.*

4.1 Weak Convergence on $\mathbb{R}^\infty$

The random sequence $\{\hat{z}(m)\}_{m\in\mathbb{N}^k}$ does not converge on a space of continuous real functions because $\hat{z}(m)$ is a step function on $\mathbb{N}^k$. It does, however, converge on a space of countably infinite sequences.

Denote by N_n^k a monotonically increasing subset of $\mathbb{N}^k$ such that $N_n^k \to \mathbb{N}^k$ as $n \to \infty$. Let $\{\psi_{n,m}\}$ be a finite sequence of possibly stochastic real numbers, $\psi_{n,m} > 0$, $\sum_{m\in N_n^k} \psi_{n,m} \leq 1$ with probability one, and $\psi_{n,m} = O_p(2^{-2\Sigma_{j=1}^{k} m_j})$. Let $\{\psi_m\}$ be a non-stochastic infinite sequence, $\psi_m > 0$, $\sum_{m\in\mathbb{N}^k} \psi_m = 1$, $\psi_m = O(2^{-2\Sigma_{j=1}^{k} m_j})$ and $\limsup_{m\to\infty} |\psi_{n,m} - \psi_m| = o_p(1)$.

Let $\mathbb{R}^\infty \equiv (\mathbb{R}^\infty, \Re^\infty)$ be the countably infinite dimensional Euclidean space with Borel sets $\Re^\infty$. See Billingsley (1995). We require separability and the notion of a bounded inner product[2].

Write $z = \{z(m)\}_{m\in\mathbb{N}^k}$. Define the inner product space

$$\left(\mathbb{B}^\infty, \mathfrak{B}^\infty, \|\cdot\|_\psi\right) = \{z \in (\mathbb{R}^\infty, \Re^\infty)\colon\ z(m) = o_p(2^{\Sigma_{j=1}^k m_j/2}),$$
$$\|z\|_\psi = (\Sigma_{m\in\mathbb{N}^k} z(m)^2 \psi_m)^{1/2}\},$$

where $\mathfrak{B}^\infty$ denotes the associated Borel sets, and the supporting inner product is $< x, y >_\psi = \sum_{m\in\mathbb{N}^k} x(m)y(m)\psi_m$. Because $\{x, y, z\} \in \mathbb{B}^\infty$ satisfy $|x(m)y(m)|\psi_m = O(2^{-\Sigma_{j=1}^k m_j})$ and $z(m)^2\psi_m = O(2^{-\Sigma_{j=1}^k m_j})$, summations like $\sum_{m\in\mathbb{N}^k} z(m)^2\psi_m$ and $\sum_{m\in\mathbb{N}^k} x(m)y(m)\psi_m$ are well-defined.

If $G_t(m) = O_p(2^{-\Sigma_{j=1}^k m_j})$ then $\eta(m) = O(2^{-\Sigma_{j=1}^k m_j})$ and $\gamma(m) = O(2^{-\Sigma_{j=1}^k m_j})$ follows from Assumption A.4. Similarly, if $G_t(m) = o_p(2^{\Sigma_{j=1}^k m_j/2})$ then $\eta(m) = o_p(2^{\Sigma_{j=1}^k m_j/2})$ and $\gamma(m) = o_p(2^{\Sigma_{j=1}^k m_j/2})$.

LEMMA 8 $\mathbb{B}^\infty$ *has the topology of pointwise convergence.* $\mathbb{B}^\infty$ *is separable and complete, hence each sequence of measures on* $\mathbb{B}^\infty$ *is tight. The finite dimensional sets* $\{z_n(m_i)\}_{i=1}^l$, $l \geq 1$, *where* $m_i \in \mathbb{N}^k$ *and* $\sigma(\{z_n(m_i)\}_{i=1}^l) \subset \mathfrak{B}^\infty$, *are convergence determining.*

---

[2] Billingsley (1999) discusses the space $(\mathbb{R}^\infty, \Re^\infty, d_\infty)$ metricized with $d_\infty(x, y) = \sum_{m=1}^\infty |x(m) - y(m)|/(1 + |x(m) - y(m)|)$. $(\mathbb{R}^\infty, \Re^\infty, d_\infty)$ is separable and complete, but $d_\infty(x, y)$ is not induced by an inner product.

Remark: Thus, pointwise convergence is equivalent to convergence in finite dimensional distributions, which is equivalent to weak convergence. Cf. Billingsley (1999).

THEOREM 9 *Under Assumption A and H_1^L there exists a Gaussian element z of $\mathbb{B}^\infty$ with mean functions $\eta(m) = (p-1)\operatorname{plim}_{n\to\infty} 1/n \sum_{t=1}^n |\epsilon_t|^{p-2} u_t g_t(m)$ and covariance functions $\gamma(m_1, m_2) = \operatorname{plim}_{n\to\infty} 1/n \sum_{t=1}^n |\epsilon_t|^{2(p-1)} g_t(m_1) g_t(m_2)$ such that $\hat{z} \Rightarrow z$ on $\mathbb{B}^\infty$.*

4.2 Average Conditional Moment Statistic

Let $\hat{v}(m) = n^{-1}\sum_{t=1}^n |\hat{\epsilon}_t|^{2(p-1)} \hat{g}_t(m)^2$ estimate the asymptotic variance of $\hat{z}(m)$. If $G_t(m) = O_p(2^{-\Sigma_{j=1}^k m_j})$ then $\hat{z}(m)/\hat{v}(m)^{1/2}$ need not be well-defined as $m \to \infty$. For the sake of brevity we will therefore,not consider supremum $\sup_{m\in N_n^k} \hat{z}(m)^2/\hat{v}(m)$ and average $\sum_m [\hat{z}(m)^2/\hat{v}(m)]\psi_{n,m}$ statistics. See, e.g., Davies (1977, 1987), King and Shively (1993), and Andrews and Ploberger (1994, 1995).

A powerful alternative statistic is the weighted-*Average Conditional Moment* [ACM]:

$$\hat{T}_n = \sum_{m\in N_n^k} \hat{z}(m)^2 \psi_{n,m}.$$

Recall $\psi_{n,m}$ may be stochastic.

THEOREM 10 $\hat{T}_n \Rightarrow \sum_{m\in\mathbb{N}^k} z(m)^2 \psi_m \equiv T_1$ *under Assumption A and H_1^L.*

Remark 1: For a chosen sequence of weights $\{\psi_m\}$ the distribution of T_1 is nuisance parameter-free: $\sum_{m\in\mathbb{N}^k} z(m)^2\psi_m$ averages out every possible nuisance parameter m in $\mathbb{N}^k$.

Remark 2: Tightness is trivially guaranteed in the present environment, and all nuisance parameters are incorporated into T_1. Thus, an extension of Hansen's (1996) monte carlo method for asymptotic p-value approximation is straightforward.

Valid stochastic weights include

$$(6) \quad \psi_{n,m} = 2^{-2\Sigma_{i=1}^q m_i} + \frac{\hat{z}(m)^2/n}{1+\sum_{m\in N_n^k}\hat{z}(m)^2/n} \text{ or } \frac{2^{-2\Sigma_{i=1}^q m_i} + \hat{z}(m)^2/n}{1+\sum_{m\in N_n^k}\hat{z}(m)^2/n}.$$

Clearly $\{\psi_{n,m}\}$ augments the weight placed on large sample moments $\hat{z}(m)$, *ceteris paribus*. As long as $\limsup_{n\geq 1}\sum_{m\in N_n^k}\hat{z}(m)^2/n = o_p(1)$ then both $\sum_{m\in N_n^k}\psi_{n,m}\rightarrow 1$ under either hypothesis. The former condition holds if $G_t(m) = O_p(2^{-\Sigma_{j=2}^k m_i})$, or $G_t(m) = o_p(2^{\Sigma_{j=2}^k m_i/2})$ and $\max_{1\leq i\leq k}\{m_{n,i}^* : m_n^* = \max\{m\in N_n^k\}\}\leq \ln n$.

4.3 The Distribution T_1

Characterizing the limiting distribution T_1 closely follows Bierens and Ploberger (1997). The space $\mathbb{B}^\infty$ is a separable, complete inner-product space, hence a separable Hilbert space. Each separable inner product space has a countably infinite orthonormal basis, say $\{\varpi_i(m)\}_{i=1}^\infty$ with $\sum_{m\in\mathbb{N}^k}\varpi_i(m)\varpi_j(m)\psi_m = I_{i=j}$ (e.g. Giles, 2000: Theorem 3.27). Thus, for some orthonormal sequence $\{\varpi_i(m)\}_{i=1}^\infty$ each element $z\in\mathbb{B}^\infty$ admits a coordinate-wise expansion

$$(7) \quad z(m) = \sum_{i=1}^\infty \varpi_i(m)u_i \;\; a.s.$$

where $\{u_i\}_{i=1}^{\infty}$ satisfies

$$(8) \quad u_i = \langle z, \varpi_i \rangle_{\psi} = \sum\nolimits_{m \in \mathbb{N}^k} z(m) \varpi_i(m) \psi_m.$$

Moreover, because each $\Gamma(m_1, m_2)$ is a symmetric positive-semi-definite function, the linear operator $\Gamma = (\Gamma(m_1, m_2))_{m_1, m_2 \in \mathbb{N}^k}$ is a compact self-adjoint operator (e.g. Giles, 2000: Section 15). Using the spectral theorem for compact self-adjoint operators Γ on the Hilbert space $(\mathbb{B}^{\infty}, \mathfrak{B}^{\infty}, || \cdot ||_{\psi})$, there is an orthonormal basis of $(\mathbb{B}^{\infty}, \mathfrak{B}^{\infty}, || \cdot ||_{\psi})$ consisting of eigenvectors of Γ, where each eigenvalue λ_i of Γ is real and non-negative (Giles, 2000: Theorem 20.4.1). It is immediate that $\{\varpi_i(m)\}_{i=1}^{\infty}$ denotes the eigenfunctions of Γ, hence

$$\sum\nolimits_{m_2 \in \mathbb{N}^k} \Gamma(m_1, m_2) \varpi_i(m_2) \psi_{m_2} = \lambda_i \varpi_i(m_1)$$

Using Parseval's identity, (6)-(7) and the orthonormality of $\{\varpi_i(m)\}_{i=1}^{\infty}$ we obtain

$$T_1 = \sum\nolimits_{m \in \mathbb{N}^k} z(m)^2 \psi_m = \langle z, z \rangle_{\psi} = \sum\nolimits_{i=1}^{\infty} u_i^2.$$

Each $z(m)$ under H_1^L is Gaussian, therefore each Fourier coefficient $u_i = \sum_{m \in \mathbb{N}^k} z(m) \varpi_i(m) \psi_m$ is Gaussian and therefore completely characterized by means

$$\eta_i = E[u_i] = \sum\nolimits_{m \in \mathbb{N}^k} \eta(m) \varpi_i(m) \psi_m$$

and pair-wise covariances

$$\begin{aligned} &E\left(\sum\nolimits_{m \in \mathbb{N}^k} [z(m) - \eta(m)] \varpi_i(m) \psi_m\right) \\ &\qquad \times \left(\sum\nolimits_{m \in \mathbb{N}^k} [z(m) - \eta(m)] \varpi_j(m) \psi_m\right) \\ &= \sum\nolimits_{m_1 \in \mathbb{N}^k} \sum\nolimits_{m_2 \in \mathbb{N}^k} \Gamma(m_1, m_2) \varpi_i(m_1) \varpi_j(m_2) \psi_{m_1} \psi_{m_2} = \lambda_i I_{i=j}. \end{aligned}$$

Thus $\{u_i\}_{i=1}^{\infty}$ is a sequence of independent Gaussian random variables with means η_i and variances λ_i. This proves the limit distributions of the ACM statistic $\hat{T}_n$ and Bierens and Ploberger's ICM statistic are identical in form.

THEOREM 11 *Under Assumption A and H_1^L there exists a sequence $\{\xi_i\}_{i=1}^{\infty}$ of iid standard normal random variables such that*

$$T_1 = \sum\nolimits_{m \in \mathbb{N}^k} z(m)^2 \psi_m = \sum\nolimits_{i=1}^{\infty} [\xi_i \lambda_i^{1/2} + \eta_i]^2.$$

Remark: Because $\sum_{i=1}^{\infty} [\xi_i \lambda_i^{1/2} + \eta_i]^2$ identically represents the limiting distribution form of the ICM test, all of the implied properties of the ICM test carry over to the ACM test, including properties under the null, global alternative, "large" local alternatives, and asymptotic admissibility for normally distributed errors.

5. Monte Carlo Study In this final section we perform a limited monte carlo study. We draw 100 samples of *iid* standard normal random variables ϵ_t, $n \in \{400, 800\}$ and we simulate q-order autoregression (AR), self-exciting-threshold-autogression [SETAR], and bilinear [BILIN] random variables. Write $\tilde{x}_t = [x_{t-1}, ..., x_{t-q}]'$. We compute

AR: $x_t = \phi_1' \tilde{x}_t + \epsilon_t$

SETAR: $x_t = \phi_1' \tilde{x}_t \times I(x_{t-1} > 0) + \epsilon_t$

BILIN: $x_t = \phi_1' \tilde{x}_t \times \epsilon_{t-1} + \epsilon_t$.

The order q is randomly selected from $\{1, ..., 5\}$, ϕ_1 and ϕ_2 are randomly selected from $[-.9, .9]^q$ contingent on all roots being outside the unit circle. Because ϵ_t is symmetric *iid* and $\tilde{x}_t$ is $\Im_{t-1}$-measurable, trivially the best L_p-predictor for any $p \in (1, 2]$ satisfies $Q(x_t|\Im_{t-1}) = \phi_1' \tilde{x}_t$ in the AR case, etc.

For each sample we estimate an AR(q) null model by L_p-regression for each $p \in \{1.1, 1.25, 1.75, 2\}$, and select the order q by minimizing the AIC over $\{0, ..., 10\}$.

For the ACM statistic $\hat{T}_n = \sum_{m \in N_n^k} \hat{z}(m)^2 \psi_{n,m}$ we use the stochastic weight

$$\psi_{n,m} = \left(2^{-2\Sigma_{i=1}^q m_i} + \hat{z}(m)^2/n\right) \times \left[1 + \sum\nolimits_{m \in N_n^k} \hat{z}(m)^2/n\right]^{-1}.$$

For the moment weights we use

$$G_t^{(E1)}(m) = \exp\left\{m' \tilde{x}_t / \max\nolimits_{1 \le i \le q}\{|\tilde{x}_{t,i}|\}\right\}$$

$$G_t^{(E2)}(m) = \prod\nolimits_{i=1}^q [\exp\{-|\tilde{x}_{t,i}|\} \times sign(\tilde{x}_{t,i})]^{m_i}$$

$$G_t^{(L2)}(m) = \prod\nolimits_{i=1}^q [1 + \exp\{-\tilde{x}_{t,i}\}]^{-m_i}$$

$$G_t^{(P1)}(m) = \prod\nolimits_{i=1}^q (1 + \tilde{x}_{t,i} / \max\nolimits_{1 \le i \le q}\{|\tilde{x}_{t,i}|\})^{m_i}$$

$$G_t^{(P2)}(m) = \prod\nolimits_{i=1}^q \left(\frac{1}{(1 + |\tilde{x}_{t,i}|) \times sign(\tilde{x}_{t,i})}\right)^{m_i}$$

Notice $\{G_t^{(E1)}(m), G_t^{(P1)}(m)\}$ are $o_p(2^{\Sigma_{i=1}^q m_i/2})$ and $\{G_t^{(E2)}(m), G_t^{(L2)}(m), G_t^{(P2)}(m)\}$ are $O_p(2^{-\Sigma_{i=1}^q m_i})$.

For any moment condition weight discussed in Section 3.3 $\sum_{m \in N_n^k} \psi_{n,m} \to 1$ holds sufficiently if $\max_{1 \le i \le k}\{m_{n,i}^* : m_n^* = \max\{m \in N_n^k\}\} \le \ln n$. Denote by $j_i^{(k)}$ a q-vector with the value j for the i^{th} component and the value k in all other components. For example, $2_3^{(0)} = [0,0,2,0,...,0]'$ and $2_1^{(2)} = [2,2,...,2]'$. Let $\tilde{N}_n^q$ be a set with $[\sqrt{\ln n}]$ integer vectors randomly selected from $\{[0,...,0]', ..., [[\sqrt{\ln n}],...,[\sqrt{\ln n}]']\}$. Let $\check{N}_n^q$ be the set of all integers in the hypercube $\{[0,...,0]', ..., [[(\ln n)^{1/8}],...,[(\ln n)^{1/8}]]'\}$. Finally, let $\breve{N}_n^q$ denote the set of all simple integers $\{\{i_j^{(0)}\}_{i=1}^{[\sqrt{\ln n}]}\}_{j=1}^{q}$ and $\{1_1^{(1)}, 2_1^{(2)}, ..., [\sqrt{\ln n}]_1^{([\sqrt{\ln n}])}\}$. The nuisance integer m is taken from the set

$$N_n^q = \tilde{N}_n^q \cup \check{N}_n^q \cup \breve{N}_n^q.$$

Thus N_n^q contains n_m-vectors ranging from $[1,0,...,0]$ to $[[\sqrt{\ln n}],...,[\sqrt{\ln n}]]'$. Clearly $\breve{N}_n^q \to \mathbb{N}^q$ hence $N_n^q \to \mathbb{N}^q$.

Test results are located in Table 1. Each ACM test generates a reasonable empirical size at the nominal 5% level given the rejection rates have 99% bounds $.05 \pm .0195$. The best weights with respect to empirical power and all $p \in \{1.1, 1.25, 1.75, 2\}$ are $G_t^{(E1)}(m) = \exp\{m'\tilde{x}_t / \max_{1\leq i \leq q}\{|\tilde{x}_{t,i}|\}\}$ and $G_t^{(L2)}(m) = \sqcap_{i=1}^{q}[1+\exp\{-\tilde{x}_{t,i}\}]^{-m_i}$. For $p = 2$ the weight $G_t^{(P2)}(m) = \sqcap_{i=1}^{q}[(1+|\tilde{x}_{t,i}|) \times sign(\tilde{x}_{t,i})]^{m_i}$ works extremely well. Recalling that all model parameters are chosen randomly, and not purposefully to enhance power, empirical power resulting from the weights $\{G_t^{(E1)}(m), G_t^{(L12)}(m), G_t^{(P2)}(m)\}$ reaches 80%-87% for $n = 800$.

Appendix 1: Assumptions

Assumption A1: *The parameter space* Φ *is a compact subset of* $\mathbb{R}^k$. $\phi_0 = \arg\inf_{\phi\in\Phi} E|y_t - f_t(\phi)|^p \in \text{interior}\{\Phi\}$, $p \in (1,2]$. $f_t(\phi)$ *is twice continuously dif-*

ferentiable on Φ. u_t *and* $f_t(\phi)$ *are* $\Im_{t-1}$*-measurable, where* $\Im_t$ *is the sequence of* σ*-algebras generated by* $(x_\tau : \tau \leq t+1)$. *Moreover,* $E[\epsilon_t^{<p-1>}|\Im_{t-1}] = 0$ *a.s. for some* $p \in \min\{(1, 1 + .25 \times \arg\sup_{\alpha>0}\{E|\epsilon_t|^\alpha < \infty\}, 2\}$.

ASSUMPTION A2: *Let* $A_n(\phi) = (p-1)(1/n)\sum_{t=1}^n |y_t - f_t(\phi)|^{p-2}\partial f_t(\phi)\partial' f_t(\phi)$, *where* $A_n(\phi) \to A(\phi)$ *uniformly on* Ξ, *where* $A(\phi)$ *is a non-stochastic matrix such that* $A(\phi_0)$ *is positive definite. Moreover, the* L_p*-estimator* $\hat{\phi} = \arg\min_{\phi\in\Phi}\sum_{t=1}^n |y_t - f_t(\phi)|^p$ *satisfies for some stochastic sequence* $\{u_t^*\}$, $u_t^* \in [0, u_t]$, $u_t^* = o_p(\sqrt{n})$,

$$\sqrt{n}\left(\hat{\phi} - \phi_0\right) = A(\phi_0)^{-1}\left(\sum_{t=1}^n \frac{\epsilon_t^{<p-1>}}{\sqrt{n}}\frac{\partial f_t(\phi_0)}{\partial \phi} + \frac{1}{n}\sum_{t=1}^n \left|\epsilon_t + \frac{u_t^*}{\sqrt{n}}\right|^{p-2} u_t \frac{\partial f_t(\phi_0)}{\partial \phi}\right) + o_p(1).$$

ASSUMPTION A3: *Let* $\hat{b}(m, \phi_0) = (p-1)(1/n)\sum_{t=1}^n |y_t - f_t(\phi_0)|^{p-2} \times G_t(m)\partial' f_t(\phi_0)$, *where* $G_t(m) = O_p(2^{-\Sigma_{j=2}^k m_j})$. *Then* $\hat{b}(m, \phi) \to b(m, \phi)$ *uniformly on* $\mathbb{N}^k \times \Xi$ *where* $b(m, \phi)$ *is a non-stochastic function satisfying* $\sup_{\phi\in\Phi, m\in\mathbb{N}^k} |b(m, \phi)|_2 < \infty$.

ASSUMPTION A4:

i. $(1/n)\sum_{t=1}^n E[|\epsilon_t|^{2(p-1)}(\partial/\partial\phi)f_t(\phi)(\partial/\partial\phi')f_t(\phi)] \to A_2$, *a finite non-stochastic matrix.*

ii. *There exists a mapping* $\eta : \mathbb{Z}^k \to \mathbb{R}$ *such that* $(p-1) \times (1/n)\sum_{t=1}^n |\epsilon_t|^{p-2} u_t g_t(m) \to (p-1) \times \lim_{n\to\infty}(1/n)\sum_{t=1}^n E[|\epsilon_t|^{p-2} u_t g_t(m)] = \eta(m)$. *If* $G_t(m)$ is $O_p(q(m))$ *or* $o_p(q(m))$ *for some* $q : \mathbb{Z}^k \to \mathbb{R}$ *then* $\eta(m) = O(q(m))$ or $o(q(m))$.

iii. *There exists a functional* $\gamma(m_1, m_2)$ *on* $\mathbb{N}^{k\times k}$ *such that* $(1/n)\sum_{t=1}^n E[|\epsilon_t|^{2(p-1)}|\Im_{t-1}] \times g_t(m_1)g_t(m_2) \to \gamma(m_1, m_2)$, $(1/n)\sum_{t=1}^n |\epsilon_t|^{2(p-1)} \times g_t(m_1)g_t(m_2) \to \gamma(m_1, m_2)$ *and* $(1/n)\sum_{t=1}^n E[|\epsilon_t|^{2(p-1)} \times g_t(m_1)g_t(m_2)] \to \gamma(m_1, m_2)$ *pointwise on* $\mathbb{N}^{k\times k}$. *If* $|G_t(m)| = O_p(q(m))$ *or* $o_p(q(m))$ *for some* $q : \mathbb{Z}^k \to \mathbb{R}$ *then* $\gamma(m_1, m_2) = O(q(m))$ *or* $o(q(m))$.

iv. *For some* $\delta > 0$, $\limsup_{n\to\infty} \sup_{m\in\mathbb{N}^k} 1/n \sum_{t=1}^{n} E||\epsilon_t|^{p-2} u_t g_t(m)|^{2+\delta} < \infty$.

Appendix 2: Proofs of Main Results

Proof of Lemma 1. Using Assumption B, Theorem 2.3 of Stinchcombe and White (1998) implies the closure of each set S_i has empty interior, and therefore S_i is nowhere dense in $\mathbb{R}^{p_1 k+1}$. In particular, each S_i has Lebesgue measure zero. ■

Proof of Lemma 2. The claim follows from Lemmas A.1-A.4, below. Under H_1, Lemma A.2 proves for any $\delta \in \Delta$, some set $S \subset T^*(\Psi(x_t(\delta))$ with Lebesgue measure zero, and every $\tau \in T^*(\Psi(x_t(\delta)))/S$,

$$(9) \quad E\left[e_t \prod_{i=1}^{k} \Psi_i(x_t(\delta))^{m_i} F^{\tilde{s}}(\tau'\Psi(x_t(\delta)))\right] \neq 0$$

for some $m \in \mathbb{Z}^k$ and some scalar integer $\tilde{s} \geq 0$. Lemma A.3 proves (8) holds for each $\tau \in S$. Trivially $T^*(\Psi(x_t(\delta))) = T^*(\Psi(x_t(\delta)))/S \cup S$, hence (8) holds $\forall\tau \in T^*(\Psi(x_t(\delta)))$.

Finally, for each $\tau \in T^*(\Psi(x_t(\delta)))$ Lemma A.4 implies the moment condition holds for infinitely many $m \in \mathbb{Z}^k$ and some integer $\tilde{s} \geq 0$ for each m. ■

Proof of Corollary 5. Except for $O_p(2^{-\Sigma_{j=1}^{k-1}\tilde{m}_j^2})$-bound the result is an immediate consequence of Lemma 1 and Corollary 4. For the bound we may write

$$\begin{aligned}
&\exp\{-.5\Sigma_{i=1}^{k-1}(\tilde{x}_{t,i} - \tilde{m}_i)^2\} \\
&\quad = \exp\{-.5\Sigma_{i=1}^{k-1}\tilde{x}_{t,i}^2\}\left(\exp\left\{.5 - \Sigma_{i=1}^{k-1}\tilde{x}_{t,i}\,(\tilde{m}_i/m_1)\right\}\right)^{-m_1}
\end{aligned}$$

where $m_1 = \Sigma_{j=1}^{k-1}\tilde{m}_j^2$. Thus $\exp\{-.5\Sigma_{i=1}^{k-1}(\tilde{x}_{t,i} - \tilde{m}_i)^2\} = O_p(2^{-\Sigma_{j=1}^{k-1}\tilde{m}_j^2}) = O_p(2^{-\Sigma_{j=1}^{k-1}\tilde{m}_j})$ given x_t is bounded and $m \in \mathbb{N}^k$. ■

Proof of Theorem 6. We prove the claim in two steps. Step 1 proves the result contingent on a preliminary claim proved in Step 2. We only consider the scalar case $k = 1$ for notational simplicity. The general case $k \geq 1$ follows by an identical argument.

Step 1: For each $i = 1, 2, ...$ project $G_t(i)$ onto $\overline{sp}(\{G_t(l)\}_{l=0}^{i-1})$ by L_p-orthogonal metric projection, write $e_i(x_t) \equiv G_t(i) - Q(G_t(i)|\overline{sp}(\{G_t(l)\}_{l=0}^{i-1}))$ and form an L_p-orthonormal functional $\{\psi_i(x_t)\}$:

$$\begin{aligned}
\psi_i(x_t) &= e_i(x_t)/(E|e_i(x_t)|^p)^{1/p}, \text{ if } E|e_i(x_t)|^p > 0 \\
&= 0, \text{ if } E|e_i(x_t)|^p = 0.
\end{aligned}$$

Clearly $e_j(x_t) \in \overline{sp}(\{G_t(l)\}_{l=0}^{i-1})$ $\forall j < i$ hence

$$\begin{aligned}
E\left[\psi_i(x_t)^{<p-1>}\psi_j(x_t)\right] &= \frac{E\left[e_i(x_t)^{<p-1>}e_j(x_t)\right]}{(E|e_i(x_t)|^p)^{(p-1)/p}(E|e_j(x_t)|^p)^{1/p}} = 0, \ \forall i > j \\
&= 1, \ \forall i = j.
\end{aligned}$$

The Banach space $\overline{sp}(\{\psi_i(x_t)\}_{i=0}^{\infty})$ forms a Schauder basis (see Step 2, below), which in turn guarantees for each element $z_t \in \overline{sp}(\{\psi_i(x_t)\}_{i=0}^{\infty})$ the existence of a sequence of real numbers $\{\gamma_i\}_{i=0}^{\infty}$ such that $z_t = \sum_{i=0}^{\infty}\gamma_i\psi_i(x_t)$ where $\sum_{i=0}^{\infty}\gamma_i\psi_i(x_t)$ is L_p-norm convergent. See Megginson (1998: Proposition 4.1.24).

Project y_t onto $\overline{sp}(\{\psi_i(x_t)\}_{i=0}^{\infty})$: for some sequence of real constants $\{\gamma_i\}_{i=0}^{\infty}$

$$E\left(y_t - \sum\nolimits_{i=0}^{\infty}\gamma_i\psi_i(x_t)\right)^{<p-1>} z_t = 0, \ \forall z_t \in \overline{sp}(\{\psi_i(x_t)\}_{i=0}^{\infty}),$$

hence

$$E\left(y_t - \sum\nolimits_{i=0}^{\infty}\gamma_i\psi_i(x_t)\right)^{<p-1>} G_t(j) = 0, \ j = 0, 1, ...$$

By Remark 1 of Theorem 3 we deduce $Q(y_t|\Im_{t-1}) = \sum_{i=0}^{\infty}\gamma_i\psi_i(x_t)$.

Finally by construction $e_i(x_t) \in \overline{sp}(\{G_t(l)\}_{l=0}^{i})$ for each i, and each element of a finite closed linear span $\overline{sp}(\{G_t(l)\}_{l=0}^{i})$ has a finite series expansion $e_i(x_t) = \sum_{j=0}^{i}\pi_{i,j}G_t(j)$ for some sequence of real numbers $\{\pi_{i,j}\}_{j=0}^{i}$. From the construction of $\{\psi_i(x_t)\}$ we conclude $Q(y_t|\Im_{t-1}) = \sum_{i=0}^{\infty}\beta_i G_t(i)\ a.s.$, where $\beta_i = \gamma_i(E|e_i(x_t)|^p)^{-1/p}\sum_{j=0}^{i}\pi_{i,j}$.

Step 2: We now prove $\overline{sp}(\{\psi_i(x_t)\}_{i=0}^{\infty})$ forms a Schauder basis. By the L_p-orthonormal construction of $\{\psi_i(x_t)\}_{i=0}^{\infty}$, for any $m < m'$

$$\sum_{i=0}^{m}\pi_i\psi_i(x_t)\perp_p\sum_{i=m+1}^{m'}\pi_i\psi_i(x_t).$$

Moreover, L_p-orthogonality $U\perp_p V$ between arbitrary subspaces $\{U,V\}\subseteq L_p$ implies James orthogonality $||u+\lambda v||_p \geq ||u||_p$ for all $u\in U$ and $v\in V$, and $\forall\lambda\in\mathbb{R}$. Hence, for all $\lambda\in\mathbb{R}$

$$\left\|\sum_{i=0}^{m}\pi_i\psi_i(x_t)+\lambda\sum_{i=m+1}^{m'}\pi_i\psi_i(x_t)\right\|_p \geq \left\|\sum_{i=0}^{m}\pi_i\psi_i(x_t)\right\|_p.$$

Setting $\lambda = 1$ we deduce

$$\left\|\sum_{i=0}^{m'}\pi_i\psi_i(x_t)\right\|_p \geq \left\|\sum_{i=0}^{m}\pi_i\psi_i(x_t)\right\|_p.$$

The latter inequality implies $\overline{sp}(\{\psi_i(x_t)\}_{i=0}^{\infty})$ forms a Schauder basis: see Megginson (1998). ■

Proof of Theorem 7. Let $z = \{z(m)\}_{m\in\mathbb{N}^k}$ be a Gaussian element of $\mathbb{R}^{\infty}$ with mean function $\eta(m) = \text{plim}_{n\rightarrow\infty} n^{-1}\sum_{t=1}^{n}|\epsilon_t|^{p-2}u_t g_t(m)$ and covariance function $\gamma(m_1,m_2)$ $= \text{plim}_{n\rightarrow\infty} n^{-1}\sum_{t=1}^{n}|\epsilon_t|^{2(p-1)}g_t(m_1)g_t(m_2)$. The weak functional limit $\{z_n(m)\}_{m\in\mathbb{N}^k}$

$\Rightarrow \{z(m)\}_{m\in\mathbb{N}^k}$ follows from Assumption A, Theorem 6.1.7 of Bierens (1994), and the fact that all distributions on $\mathbb{B}^\infty$ are tight. The weak limit $\{\hat{z}(m)\}_{m\in\mathbb{N}^k} \Rightarrow \{z(m)\}_{m\in\mathbb{N}^k}$ is then a consequence of (4). ■

Proof of Lemma 8. A proof that $(\mathbb{B}^\infty, \mathfrak{B}^\infty, ||\cdot||_\psi)$ is separable and complete simply mimics arguments in Billingsley (1999: p.10), or Theorem 5.15 of Davidson (1994). Tightness now follows from Theorem 1.3 of Billingsley (1999). The fact that finite dimensional sets $\{z_n(m_i)\}_{i=1}^l$ form a convergence determining class is analogous to Example 2.4 and Theorem 2.4 of Billingsley (1999). ■

Proof of Theorem 10. Recall $\sup_{m\in\mathbb{N}^k} |\hat{z}(m) - z_n(m)| = o_p(1)$, cf. (4). Using Theorem 9 it suffices to prove $|\sum_{m\in N_n^k} z_n(m)^2\psi_{n,m} - \sum_{m\in\mathbb{N}^k} z(m)^2\psi_m| = o_p(1)$. Let $k = 1$ for notational convenience (i.e. $m \in \mathbb{N}$), and recall $\psi_m = O(2^{-2\Sigma_{j=1}^k m_j})$, $\limsup_{m\to\infty} |\psi_m - \psi_{n,m}| = o_p(1)$, and $E\left[z(m)^2\right] = o(2^{\Sigma_{j=1}^k m_j/2}$ by construction and Assumption A.4. Let $\psi_{n,m} = 0\ \forall m > n$ and recall $\sum_{m=1}^{N_n} \psi_{n,m} \leq 1$ with probability one. For some $K > 0$

$$
\begin{aligned}
&\left\|\sum_{m=1}^{N_n} z_n(m)^2\psi_{n,m} - \sum_{m=1}^{\infty} z(m)^2\psi_m\right\|_1 \\
&\leq \sum_{m=1}^{N_n} \left\|z_n(m)^2 - z(m)^2\right\|_1 \psi_{n,m} + \sum_{m=N_n+1}^{\infty} E\left[z(m)^2\right]\psi_m \\
&\quad + \sum_{m=N_n+1}^{\infty} E\left[z(m)^2\right]|\psi_{n,m} - \psi_m| + \sum_{m=1}^{\infty} E\left[z(m)^2\right]|\psi_{n,m} - \psi_m| \\
&\leq \sum_{m=1}^{N_n} \left\|z_n(m)^2 - z(m)^2\right\|_1 \psi_{n,m} \\
&\quad + K\sum_{m=N_n+1}^{\infty} 2^{-2m} + 2K\sum_{m=1}^{\infty} 2^{-m}|\psi_{n,m} - \psi_m| \\
&= o_p(1).
\end{aligned}
$$

The last line follows from the construction of $\{\psi_{n,m}\}$ and $\{\psi_m\}$, weak convergence $z_n(m) \Rightarrow z(m)$, the continuous mapping theorem, and the Helly-Bray Theorem: $Y_n \equiv |z_n(m)^2 - z(m)^2| \Rightarrow 0$ implies $E[Y_n] \to 0$. ■

Appendix 3: Supporting Lemmeta

LEMMA A.1 *Let e_t be a random variable satisfying $E|e_t| < \infty$, and let x_t be an $\Im_{t-1}$-measurable bounded k-vector such that $P[E(e_t|x_t) = 0] < 1$, and let Assumption B hold. Then for each $\delta \in \Delta$ and any $r \in \mathbb{Z}^k$ the set*

$$S = \{\tau \in \mathbb{R}^k : E\left[e_t \prod_{i=1}^{k} \Psi_i(x_t(\delta))^{r_i} F\left(\tau'\Psi(x_t(\delta))\right)\right] = 0$$
$$\text{and } P\left(\tau'\Psi(x_t(\delta)) \in R_0\right) = 1\}.$$

has Lebesgue measure and is nowhere dense in $\mathbb{R}^k$.

LEMMA A.2 *Let the assumptions of Lemma A.1 hold. If $P[E(e_t|x_t) = 0] < 1$, then for each $\delta \in \Delta$, some set $S \subset T^*(\Psi(x_t(\delta))$ with Lebesgue measure zero, and every $\tau \in T^*(\Psi(x_t(\delta))/S$ there exists an integer vector $m \in \mathbb{Z}^k$ and scalar integer $\tilde{s} \geq 0$ such that*

$$(10) \quad E\left[e_t \prod_{i=1}^{k} \Psi_i(x_t(\delta))^{m_i} F^{\tilde{s}}(\tau'\Psi(x_t(\delta)))\right] \neq 0.$$

In particular, $\tilde{s} = \sum_{i=1}^{k} s_i$ where $s = m - r$ for some $r \in \mathbb{Z}^k$, $m \geq r$.

LEMMA A.3 *The conclusion of Lemma A.2 holds for each $\tau_s \in S$.*

LEMMA A.4 *Let* $P[E(e_t|x_t) = 0] < 1$. *For any* $m_0 \in \mathbb{Z}^k$ *and scalar integer* $\tilde{s}_0 \geq 0$ *such that Lemmas A.2 and A.3 hold, the results hold for some* $m_1 > m_0$ *and* $\tilde{s}_1 \geq 0$.

Proof of Lemma A.1. The claim follows immediately from Lemma 1, and the fact that mapping $\Psi(x_t(\delta)$ is for each $\delta \in \Delta$ a one-to-one function of x_t. ■

Proof of Lemma A.2. Denote by $N_\xi(\tau)$ an open ξ-ball of τ, $\{\tau_0 \in \mathbb{R}^k : ||\tau - \tau_0|| < \xi\}$ for some $\xi > 0$. By construction $N_\xi(\tau)$ has positive Lebesgue measure.

Let H_1 hold. Applying Lemma A.1 for any $\delta_0 \in \Delta$ and any $r \in \mathbb{Z}^k$, the set S in (9) has Lebesgue measure zero, where $S \subset T^*(\Psi(x_t(\delta_0)) = \{\tau \in \mathbb{R}^k : P(\tau'\Psi(x_t(\delta)) \in R_0) = 1\}$ by construction.

Because $F(\cdot)$ is analytic on the interval R_0, for each $\delta_0 \in \Delta$, any $\tau \in T^*(\Psi(x_t(\delta_0))/S$, and every τ_0 in some open neighborhood $N_\xi(\tau)$ we may expand $F(\tau'\Psi(x_t(\delta_0)))$ around each scalar component $\tau_{0,i}$, $i = 1...k$,

(11)

$$\begin{aligned} &F\left(\tau'\Psi(x_t(\delta_0))\right) \\ &\quad = \sum\nolimits_{m_1=0}^{\infty} F^{m_1}\left(\tau_{0,1}\Psi_1(x_t(\delta_0)) + \sum\nolimits_{i=2}^{k} \tau_i \Psi_i(x_t(\delta_0))\right) \\ &\quad\quad \times \Psi_1(x_t(\delta_0))^{m_1}[(\tau_1 - \tau_{0,1})^{m_1}]/m_1! \end{aligned}$$

$$= \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} F^{m_1+m_2} \left(\sum_{i=1}^{2} \tau_{0,i} \Psi_i(x_t(\delta_0)) + \sum_{i=3}^{k} \tau_i \Psi_i(x_t(\delta_0)) \right)$$

$$\times \Psi_1(x_t(\delta_0))^{m_1} \Psi_2(x_t(\delta_0))^{m_2} [(\tau_1 - \tau_{0,1})^{m_1}]/m_1! [(\tau_2 - \tau_{0,2})^{m_2}]/m_2!$$

$$= ...$$

$$= \sum_{m \in \mathbb{N}^k}^{\infty} \prod_{i=1}^{k} \Psi_i(x_t(\delta_0))^{m_i} F^{\Sigma_{i=1}^k m_i} (\tau_0' \Psi(x_t(\delta_0))) B(\tau, \tau_0, m),$$

where $B(\tau, \tau_0, m) = \prod_{i=1}^k [(\tau_i - \tau_{0,i})^{m_i}]/m_i!$. Combining (9) and (11), for each $\delta_0 \in \Delta$, each $\tau \in T^*(\Psi(x_t(\delta_0)))/S$, every $\tau_0 \in N_\xi(\tau)$ and any $r \in \mathbb{Z}^k$

$$\begin{aligned} 0 &\neq E\left[e_t \prod_{i=1}^{k} \Psi_i(x_t(\delta_0))^{r_i} F\left(\tau' \Psi(x_t(\delta_0))\right) \right] \\ &= \sum_{m \in \mathbb{N}^k}^{\infty} E\left[e_t \prod_{i=1}^{k} \Psi_i(x_t(\delta_0))^{r_i + m_i} F^{\Sigma_{i=1}^k m_i} (\tau_0' \Psi(x_t(\delta_0))) \right] \times B(\tau, \tau_0, m). \end{aligned}$$

Using a simple re-parameterization, we conclude there exists at least one set of integer vectors $m \geq r$ and $s = m - r \geq 0$ such that for every $\tau_0 \in N_\xi(\tau)$ and every $\tau \in T^*(\Psi(x_t(\delta_0)))/S$

$$(12) \quad E\left[e_t \prod_{i=1}^{k} \Psi_i(x_t(\delta))^{m_i} F^{\tilde{s}} (\tau_0' \Psi(x_t(\delta_0))) \right] \neq 0,$$

where $\tilde{s} = \sum_{i=1}^k s_i \geq 0$. Note $m \geq r \in \mathbb{Z}^k$ hence $m \in \mathbb{Z}^k$. ■

Proof of Lemma A.3. Lemma A.2 holds for every τ_0 in an open neighborhood $N_\xi(\tau)$ of every $\tau \in T^*(\Psi(x_t(\delta_0)))/S$. Thus, Lemma A.2 holds for every

$$\tau_0 \in \bigcup_{\tau \in T^*(\Psi(x_t(\delta_0)))/S} N_\xi(\tau).$$

It suffices to prove $S \subset \cup_{\tau \in T^*(\Psi(x_t(\delta_0)))/S} N_\xi(\tau)$. By Lemma A.1 the set S has Lebesgue measure zero, therefore its closure has empty interior, which implies S is equivalent to its boundary. Moreover $S \subseteq T^*(\Psi(x_t(\delta_0))$ by construction. For each $\tau_s \in S$ it follows that

$$\inf_{\tau \in T^*(\Psi(x_t(\delta_0)))/S} ||\tau - \tau_s|| = 0.$$

Thus there exists some $\tau \in T^*(\Psi(x_t(\delta_0)))/S$ arbitrarily close to each $\tau_s \in S$. Therefore each $\tau_s \in S$ is an element of some open neighborhood $N_\xi(\tau)$ with positive Lebesgue measure such that (12) holds. But this implies $S \subset \cup_{\tau \in T^*(\Psi(x_t(\delta_0)))/S} N_\xi(\tau)$.

■

Proof of Lemma A.4. For any pair (δ_0, τ_0), $\delta_0 \in \Delta$ and $\tau_0 \in T^*(\Psi(x_t(\delta_0)))$, let $m_0 \in \mathbb{Z}^k$ and $\tilde{s}_0 \geq 0$ satisfy Lemmas A.2 and A.3. Now apply Lemmas A.2 and A.3 again: for the same $\delta_0 \in \Delta$, and each $\tau_1 \in T^*(\Psi(x_t(\delta_0)))$, $\tau_1 \neq \tau_0$, there exists an integer vector m_1 and scalar integer $\tilde{s}_1 \geq 0$ such that

$$(13) \quad E\left[e_t \prod_{i=1}^k \Psi_i(x_t(\delta_0))^{m_{1,i}} F^{\tilde{s}_1}(\tau_1' \Psi(x_t(\delta_0)))\right] \neq 0.$$

Note $\tilde{s}_1 = \sum_{i=1}^k s_1$, where $s_1 = m_1 - r_1$ for some $r_1 \in \mathbb{Z}^k$ and $m_1 \geq r_1$. But r_1 is arbitrary, hence we can always set $r_1 = m_0 + 1_k$ such that $m_1 \geq m_0 + 1_k$ and $r_1 \geq m_0 \geq r_0$.

Now expand (13) around each $\tau_{0,i}$, $i = 1...k$. Using the same argument from the line of proof of Lemma A.2,

$$\begin{aligned} 0 &\neq E\left[e_t \prod_{i=1}^k \Psi_i(x_t(\delta_0))^{m_{1,i}} F^{\tilde{s}_1}(\tau_1' \Psi(x_t(\delta_0)))\right] \\ &= \sum_{m \in \mathbb{N}^k}^{\infty} E\left[e_t \prod_{i=1}^k \Psi_i(x_t(\delta_0))^{m_{1i}+m_i}\} F^{\Sigma_{i=1}^k m_i + \tilde{s}_1}(\tau_0' \Psi(x_t(\delta_0)))\right] \\ &\qquad \times B(\tau, \tau_0, m). \end{aligned}$$

Therefore at least one moment

$$(14) \quad E\left[e_t \prod_{i=1}^{k} \Psi_i(x_t(\delta_0))^{m_{2,i}} F^{\tilde{s}_2}(\tau_0' \Psi(x_t(\delta_0)))\right] \neq 0$$

holds for some $m_2 \geq m_0 + 1_k > m_0$ and $\tilde{s}_2 \geq 0$. In particular, there exists some $m \in \mathbb{N}^k$ such that (14) holds for $m_2 = m + m_1 \geq m_1 \geq r_1 \geq m_0$ and $\tilde{s}_2 = \sum_{i=1}^{k} s_2 \geq 0$, where $s_2 = m + s_1 = m + m_1 - r_1 = m_2 - r_1 \geq 0$, and $r_1 \geq r_0$. ■

References

[1] Andrews, D.W.K. and W. Ploberger (1994). "Optimal Tests when a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 82, 1383-1414.
[2] Andrews, D.W.K. and W. Ploberger (1995). "Admissibility of the Likelihood Ratio Test when a Nuisance Parameter is Present Only under the Alternative," *The Annals of Statistics*, 23, 1609-1629.
[3] Arcones, M. (1996). The Bahadur-Kiefer Representation of Lp Regression Estimates, Econometric Theory 12, 257-283.
[4] Bantli, F.E., and M. Hallin (2001). "Asymptotic Behavior of M-Estimators in AR(p) Models under Non-standard Conditions," *Canadian Journal of Statistics* 29, 155-168.
[5] Bierens, H. J. (1982). "Consistent Model Specification Tests," *Journal of Econometric,* 20, 105-134.
[6] Bierens, H. J. (1984). "Model Specification Testing of Time Series Regressions," *Journal of Econometrics*, 26, 323-353.
[7] Bierens, H. J. (1987). "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands," *Journal of Econometrics*, 35, 161-190.
[8] Bierens, H. J. (1990). "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.
[9] Bierens, H. J. (1991)." Least Squares Estimation of Linear and Nonlinear ARMAX Models under Data Dependence," *Annales d'Economie et de Statistique* 20/21, 143-169.
[10] Bierens, H. J. (1994). "Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models," Cambridge Univ. Press: Cambridge.
[11] Bierens, H. J. and W. Ploberger (1997). "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129-1151.
[12] Billingsley, P., 1995. Probability and Measure (John Wiley & Sons: New York).
[13] Billingsley, P. (1999). "*Convergence of Probability Measures,*" John Wiley & Sons, New York.
[14] Cheng and De Gooijer (2005). Bahadur Representation for the Nonparametric M-Estimator under α-Mixing Dependence, Tingbergen Institute Working Paper #05-067/4.
[15] Corradi, V. and N. Swanson (2002). "A Consistent Test for Nonlinear Out of Sample Predictive Accuracy," *Journal of Econometrics*, 110, 353-381.
[16] Davies, R.B. (1977). "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247-254.
[17] Davies, R.B. (1987). "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33-43.
[18] de Jong, R. (1996). "The Bierens Test under Data Dependence,"*Journal of Econometrics*, 72, 1-32.
[19] de Jong, R., and C. Han (2002). "Properties of Lp-GMM estimators," *Econometric Theory* 18, 491-504.
[20] Dette, H. (1999). "A Consistent Test for the Functional Form of a Regression

Based on a Difference of Variance Estimators," *Annals of Statistics*, 27, 1012-40.
[21] Gallant, A.R, and G. Souza (1991). "On the Asymptotic Normality of Fourier Flexible Form Estimates," *Journal of Econometrics*, 50, 329-353.
[22] Gallant, A.R, and H. White (1988). "*A Unified of Estimation and Inference for Nonlinear Dynamic Models*". Basil Blackwell: New York.
[23] Giles, J.R. (2000). "*Introduction to the Analysis of Normed Linear Spaces,*" Cambridge Univ. Press: U.K.
[24] Han, C., and R. de Jong (2004). Closest Moment Estimation under General Conditions, Annales d'Economie et de Statistique 74, 1-13.
[25] Hansen, B. (1996). "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413-430.
[26] Hill, J. B. (2005). "Consistent Model Specification Tests Against Smooth Transition Alternatives," Dept. of Economics, Florida International University; available at http://econwpa. wustl.edu:80/eps/em/papers /0402/0402004.pdf.
[27] Hill, J. B. (2006). "Strong Orthogonal Decompositions and Non-Linear Impulse Response Functions for Infinite Variance Processes", *Canadian Journal of Statistics*, forthcoming.
[28] Holley, A. (1982). "A Remark on Hausman's Specification Test," *Econometrica*, 50, 749-760.
[29] Hong, Y. and H. White (1995). Consistent Specification Testing via Nonparametric Series Regression," *Econometrica*, 63, 1133-1159.
[30] Hornik, K. (1991). Approximation Capabilities of Multilayer Feedforward Networks, *Neural Networks* 4, 251-257.
[31] Koul, H.L. (1996). Asymptotics of Some Estimators and Sequential Residuals in Nonlinear Time Series, Annals of Statistics 24, 380-404.
[32] Koul, H.L., and Z. Zhu (1996). "Bahadur-Kiefer Representation for GMM-Estimators in Autoregression Modes," *Stochastic Processes and their Applications* 57, 167-189.
[33] Lai, P.Y. and S.M.S. Lee (2005). An Overview of Asymptotic Properties of Lp-Regression under General Classes of Error Distributions, *Journal of the American Statistical Association* 100, 446-458.
[34] Lee, B.J. (1988). "A Nonparametric Model Specification Test Using a Kernel Regression Method," Ph.D. Dissertation, University of Wisconsin, Madison.
[35] Li, Q., C. Hsiao, and J. Zinn (2003). "Consistent Specification Tests for Semi-parametric/Nonparametric Models Based on Series Estimation Methods,"*Journal of Econometrics*, 112, 295-325.
[36] Liebscher, E. (2003). Strong Convergence of Estimators in Nonlinear Autoregressive Models, Journal of Multivariate Analysis 84, 247-261.
[37] Lindenstrauss, J. and L. Tzafriri (1977). "*Classical Banach Spaces,*" Springer: New York.
[38] Megginson, R. (1998). "*An Introduction to Banach Space Theory,*" Springer-Verlag: New York.
[39] Newey, W.K. (1985)."Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047-1070.
[40] Stinchcombe, M.B. and H. White (1998). "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14,

295-325.
[41] White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.
[42] White, H. (1989). "An Additional Hidden Unit Test of Neglected Nonlinearity in Multilayer Feedforward Networks," in *Proceedings of the International Joint Conference on Neural Networks*, Washington D.C., Vol. 2 (IEEE Press: New York).
[43] Wu, W.B. (2006). "M-Estimation of Linear Models with Dependent Errors," *Annals of Statistics*: forthcoming.
[44] Yatchew, A.J. (1992). "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435-451.
[45] Zheng, J. (1996). "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263-289.

Table 1 - ACM

	p	1.10		1.25		1.75		2.00	
Model	n	400	800	400	800	400	800	400	800
AR	E1-ACM[a]	.02	.02	.06	.04	.04	.05	.03	.05
	E2-ACM[b]	.03	.06	.03	.02	.01	.00	.03	.04
	L2-ACM[c]	.04	.02	.05	.06	.01	.00	.03	.05
	P1-ACM[d]	.06	.07	.06	.05	.02	.01	.04	.05
	P2-ACM[e]	.08	.03	.03	.03	.02	.03	.02	.04
SETAR	E1-ACM	.66	.69	.58	.67	.68	.83	.65	.82
	E2-ACM	.52	.69	.59	.71	.38	.64	.54	.87
	L2-ACM	.67	.78	.63	.79	.56	.79	.67	.87
	P1-ACM	.57	.41	.46	.43	.45	.54	.39	.38
	P2-ACM	.26	.56	.31	.54	.26	.50	.33	.79
BILIN	E1-ACM	.54	.82	.66	.73	.61	.71	.64	.80
	E2-ACM	.32	.57	.36	.56	.41	.53	.68	.52
	L2-ACM	.56	.79	.52	.72	.49	.73	.68	.75
	P1-ACM	.45	.60	.58	.43	.49	.51	.42	.63
	P2-ACM	.28	.42	.25	.37	.30	.39	.40	.52

Notes: a. The weight is $\exp\{m'\tilde{x}_t / \max_{1\leq i\leq q}\{|\tilde{x}_{t,i}|\}\}$.
b. The weight is $\sqcap_{i=1}^{q}[\exp\{-|\tilde{x}_{t,i}|\} \times sign(\tilde{x}_{t,i})]^{m_i}$.
c. The weight is $\sqcap_{i=1}^{q}[1+\exp\{-\tilde{x}_{t,i}\}]^{-m_i}$.
d. The weight is $\sqcap_{i=1}^{q}(1+\tilde{x}_{t,i}/\max_{1\leq i\leq q}\{|\tilde{x}_{t,i}|\})^{m_i}$.
e. The weight is $\sqcap_{i=1}^{q}[(1+|\tilde{x}_{t,i}|) \times sign(\tilde{x}_{t,i})]^{-m_i}$.