9-19-2006

Super-Consistent Tests of Lp-Functional Form

Jonathan B. Hill
Department of Economics, Florida International University

Follow this and additional works at: http://digitalcommons.fiu.edu/economics_wps

Recommended Citation
http://digitalcommons.fiu.edu/economics_wps/54

This work is brought to you for free and open access by the Department of Economics at FIU Digital Commons. It has been accepted for inclusion in Economics Research Working Paper Series by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.
Super-Consistent Tests of $L_p$-Functional Form

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

September 19, 2006

Abstract
This paper develops a consistent test of best $L_p$-predictor functional form for a time series process. By functionally relating two moment conditions with different nuisance parameters we are able to construct a vector moment condition in which at least one element must be non-zero under the alternative. Specifically, we provide a sufficient condition for moment conditions of the type characterized by Stinchcombe and White (1998) to reveal model misspecification for any nuisance parameter value. When the sufficient condition fails an alternative moment condition is guaranteed to work. A simulation study clearly demonstrates the superiority of a randomized test: randomly selecting the nuisance parameter leads to more power than average- and supremum-test functionals, and obtains empirical power nearly equivalent to uniformly most powerful tests in most cases.

1. Introduction
This paper develops consistent parametric tests of best $L_p$-predictor functional form for a time series process in the spirit of Bierens (1991), Bierens and Ploberger (1997) and Stinchcombe and White (1998). In our main result we utilize two interactive "revealing" moment conditions: at least one of the moment conditions must be non-zero under model misspecification.

Apparently the only consistent parametric CM tests are those of Bierens (1982, 1984, 1990), de Jong (1996), the Integrated CM test of Bierens (1982) and Bierens and Ploberger (1997). See, also, Andrews and Ploberger (1994), de Jong and Bierens (1994), and Dette (1999) for related methods. Consistency is achieved by generating weight functions $F(\tau; \nu)$, indexed by a real-valued nuisance vector $\tau \in \frac{1}{2} \mathbb{R}^k$, effectively producing uncountably many moment conditions which "reveal" model misspecification. Expanding upon Bierens' (1990) seminal Lemma 1, Stinchcombe and White (1998) show that any real analytic
function $F(A(x_t))$ that is non-polynomial can reveal model mis-specification, where $A : R^k \to R$ is affine.

Hill (2006) takes a different tack by constructing a class of revealing weights $G_t(m) = u^m(x_t)$ based on integer nuisance parameters $m \in Z^k$ where $u : R^k \to R$ is any bounded, one-to-one function. The weight $G_t(m)$ need not be differentiable with respect to $x_t$ nor, therefore, analytic, and may be polynomial.

Although much has been said about the measurability of the subset $S \subseteq \mathbb{Y}$ on which consistent tests fail, almost nothing has been said about its exact contents and how to control them. The extant literature argues $S$ has countably many elements, cf. Bierens (1990), Bierens and Ploberger (1997) and Stinchcombe and White (1998). Hill (2006) considers a set of moment conditions that contains Bierens’ (1990) exponential, and proves $R^k / S$ contains infinitely many integers.

In this paper we demonstrate that a class of moment conditions exists in which $S$ is empty, or contains only the origin. In either case we say the moment condition is "super-revealing" and an associated asymptotic power-one test is "super-consistent". We effectively provide a sufficient condition for moment conditions of the type characterized by Stinchcombe and White (1998) to reveal model mis-specification for any non-zero nuisance parameter value. When the sufficient condition fails an alternative moment condition is guaranteed to work. Stacking the two moment conditions leads to a super-consistent test statistic.

A simulation study demonstrates that our test statistic with a randomized nuisance parameter generates more empirical power than average and supremum test functionals, and obtains power nearly equivalent to uniformly most powerful tests.

In Section 2 we present a preliminary result concerning revealing moment conditions for best $L_p$-predictors. Section 3 develops a "super-revealing" class of moment conditions. In Section 4 we augment the main result to weights with integer-valued nuisance parameters in order to simplify test statistic construction. Section 5 details the construction of a test statistic, and Section 6 contains a simulation study.

Throughout this paper denotes convergence in probability, or finite distributions. denotes weak convergence on a function space. $sp(f_z g_{i=1}^n)$ denotes the span of $z_1, ..., z_n,$ and $sp(f_z g_{i=1}^n)$ denotes the closed linear span. $j \cdot j$ denotes the $l_1$-matrix norm. We write $C$ to denote a positive, finite constant whose values may change with the context.

2. Vector-Valued Conditional Moments

Let $f_{y_t, \pi_1} g_{2 R^k \in R^k}$ be a strictly stationary, ergodic stochastic process in $L_p(\cdot = \mu), p \geq 2 (1, 2),$ with nondegenerate continuous marginal distributions, $= \sigma(\cdot \pi_1)$, $= \mu = \sigma(\cdot \pi_0 (1, \pi_1)^0)$. The regressors $\pi_t$ 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, $f_t : R^k \to R$ measurable with respect to $= \pi_1$, with a compact subset of $R^k$. Consult Appendix 1 for all assumptions detailed under Assumption A, and see Hill (2006) for complete details on the following set-up.
Denote by $Q_{y_1}  y_1 y$ the orthogonal $L_2$-metric projection of $y_1$ onto the space spanned by $f_{x_{t_1},i} g_{i-1}$. If we write

$$e_t := e_t^{<1, 1>} = \langle y_t, Q_{y_1}  y_1 y \rangle^{<1, 1>}$$

then clearly $E[e_t z_{t_1} 1] = 0$ for some $z_{t_1} 2$ satisfying $f_{x_{t_1},i} g_{i-1}. \text{ If } p = 2 \text{ then } Q_{y_1}  y_1 y = E[|e_t|^2]$. The fundamental hypotheses are

$$H_0: P(Q(y_t y_1) = f(x_t, \phi_0)) = 1, \text{ for some } \phi_0 2 \inf \subset \sup \phi \exists \phi 2 \subset \sup \phi$$

Under $H_0$ the function $f(x_t, \phi_0)$ represents the best $L_2$-predictor of $y_t$. Write $F_0(u) := \langle \partial/\partial u \rangle F(u)$ and define the following class of weights:

$$H_F = f_g : R^k \setminus R^k \setminus f_g(x) = F(A(x)), A \text{ analytic and non-polynomial on some open interval } R_0 \subset R^k.$$ 

Assumption B: Let $F \in H_F$, and $(\partial/\partial u)^i F(u)]_{u=0} = c_i$ where $c_i = 0$ for only finitely many $i \leq N$. Let $0 \text{ lie in the interior of } R_0$.

Remark: If $F$ and $F^0$ are analytic on some open interval $R_0 \subset R^k$ then so are $F + c$ and $F^0 + c$, where $c$ is any real-valued constant. Trivial examples of weights satisfying Assumption B are $e^{\exp(u_i) [1 + \exp(u_i)]}$, and trigonometric functions.

Let $h : R^k \subset f \subset R^k$ be a uniformly bounded, $f_1$-measurable function, $h_1 1$, where $f$ is an arbitrary subset of $R^l$ for some $l \leq N$. Write $h_i(\delta) = h_i(x_i, \delta).$ The following is a required, although easy, extension of Theorem 1 of Bierens and Ploberger (1997) and Theorem 3.9 of Stinchcombe and White (1998).

**Lemma 1** Let $e_t$ be a random variable satisfying $E|e_t | < 1$, and let $x_t$ be an $x_{t_1} 1$-measurable bounded vector in $R^k$ such that $P(E[e_t x_t] = 0) < 1$. Let Assumption B hold. For each $\delta > 0$ the set

$$S \subset \sum_{i=1}^k \delta R^k : E[e_t h_{t_i}(\delta) F(x_{t_i}^0)] = 0 \text{ and } P(x_{t_i}^0 \subset R_0) = 0$$

has Lebesgue measure zero and is nowhere dense in $R^k$.

Remark: By Assumption B and Theorem 3.9 of Stinchcombe and White (1998), Lemma 1 holds with $F(\phi)$ replaced by $F^0(\phi)$. 

3
3. **Super-Revealing Moments** Let $\mathcal{J}$ be an arbitrary compact subset of $\mathbb{R}^k$ with positive Lebesgue measure. We will require

$$0_2 \mathcal{J}$$

to expedite the proof of the main result, but sets $\mathcal{J}$ not containing zero may be considered in practice. Consider any weight function $F$ satisfying Assumption B, and let $\phi : \mathbb{R}^k \to \mathbb{R}^k$ be bounded, one-to-one $=_{\text{t-1}}$-measurable function. Write

$$e_t = e^{\phi t} 1_0$$

and define the moment

$$\eta(\gamma) := E[e_t F(\gamma \phi (x_t))]$$

Construct the set $\mathcal{J}$ of matrices $\gamma = [\gamma^{(1)}, ..., \gamma^{(k)}] \in \mathbb{R}^{k \times k}$ from

$$\gamma^{(i)} = \arg \sup_{\gamma^{(i)}} f(\partial / \partial \gamma_i) \eta(\gamma) \g.$$ 

For the sake of convention assume $(\partial / \partial \gamma_i) \eta(\gamma)_{\gamma = \gamma(\phi)} = 0$. All subsequent results carry over to the general case $(\partial / \partial \gamma_i) \eta(\gamma)_{\gamma = \gamma(\phi)} \neq 0$.

In general $\mathcal{J}$ may contain more than one element under either hypothesis, and may have zero or positive Lebesgue measure. Under the null hypothesis, for example, $(\partial / \partial \gamma_i) \eta(\gamma)_{\gamma = \gamma(\phi)} = 0$ holds with probability one for all $\gamma^{(i)} = i$, hence $\mathcal{J} = \phi \in \mathcal{J}^{(i)}$. Under $H_1$ if $\eta(\gamma)$ is non-monotonic then $\arg \sup_{\gamma^{(i)}} f E[e_t (x_t) F(\gamma)] \g$ need not be unique and may be zero. Lemma 1 implies the set of $\mathcal{J}$ on which $E[e_t \phi (x_t) F(\gamma)] = 0$ under $H_1$ has Lebesgue measure zero, but this is immaterial here.

Define the vector weight function

$$H_i(\gamma, \gamma^{(i)}) := F_i(\gamma), \phi_1(x_t) F_1(\gamma^{(1)}), ..., \phi_k(x_t) F_k(\gamma^{(k)}) \in \mathbb{R}^{k+1},$$

and define the set

$$\mathcal{S}^{(i)} = \bigcap_{i=1}^k f \mathcal{J} : E[e_t H_i, \mathcal{J} (\gamma, \gamma^{(i)})] = 0, \text{ and } P(\gamma \phi (x_t) \neq 0 \neq R_0) = 1.$$

Consider the moment

$$\varpi(\gamma, \gamma^{(i)}) := E \left[ e_t F_i(\gamma) \phi \sum_{i=1}^k \gamma_i \phi_i(x_t) F_i(\gamma^{(i)}) \right].$$

**Lemma 2** Let $x_t$ be an $=_{\text{t-1}}$-measurable bounded vector in $\mathbb{R}^k$ such that $P(\gamma \phi (x_t) = 0) < 1$. Then $\varpi(\gamma, \gamma^{(i)}) = 0$ if and only if $\gamma = 0$.

The main result of the paper follows easily from Lemma 2.
THEOREM 3 Let \( x_t \) be an \( =_t \)-measurable bounded vector in \( \mathbb{R}^k \) such that
\[
P(E[e_i|x_t] = 0) < 1.
\]
Then \( S^a = f g \) if and only if \( e_t \) s.t. \( \sum_{i=1}^{k} F_t^Q(\gamma_i^{(i)}) = 1 \), and \( S(\omega) = f g \) otherwise.

Remark: The vector moment condition \( E[e_i H_t(\gamma, \gamma_i^{(i)})] \) provides a two-way safety net against failing to detect model misspecification. If a chosen \( \gamma \in 0 \) implies failure of a moment condition characterized by Stinchcombe and White (1998),
\[
E[e_i F_t(\gamma)] = 0,
\]
then for at least one \( i \leq 1, \ldots, k \) we are guaranteed a model misspecification revealing moment
\[
h E[e_i^g(x_t) F_t^Q(\gamma_i^{(i)})] \in 0.
\]

Conversely, because the vectors \( \gamma_i^{(i)} \) maximize each gradient level \( E[e_i^g(x_t) F_t^Q(\gamma_i^{(i)})] \) and not the absolute magnitude \( \| E[e_i^g(x_t) F_t^Q(\gamma_i^{(i)})] \| \), it is possible that \( E[e_i^g(x_t) F_t^Q(\gamma_i^{(i)})] = 0 \) for each \( i = 1, \ldots, k \) such that none reveal misspecification. In such a case \( E[e_i F_t(\gamma_i^{(i)})] \in 0 \) is guaranteed to hold for all non-zero \( \gamma \neq 0 \).

EXAMPLE 1 Let \( F(u) = \exp u g \) and assume \( x_t \) is bounded. If \( P(E(e_i|x_t) = 0) < 1 \) then \( 8 \gamma \in 0 \)
\[
h E[e_i \exp x_i t \exp g(\gamma^{(i)} x_i t g)] E[e_i x_i t \exp g(\gamma^{(i)} x_i t g)] \in 0.
\]


EXAMPLE 2 Let \( e_t = e_i^{<p_i \uparrow} 1, < p \cdot 2 \). Suppose \( \gamma^{(a)} = 0 \) such that \( F_t^Q(\gamma_i^{(i)}) = F_t^Q(0) = c_1 \) by Assumption B. If \( c_1 = 0 \) then \( 0 \leq S^a \). If \( c_1 \in 0 \) then \( 0 \leq S^a \) if and only if \( E[e_i^{<p_i \uparrow} 1, i a (x_t)] = 0 \). For example, if we use the weight \( F(\gamma_i^{(a)}) \) to test linearity \( f_i(\phi) = \phi x_t \) then \( E[e_i^{<p_i \uparrow} 1, i a (x_t) F_t^Q(\gamma_i^{(i)})] = E[e_i^{<p_i \uparrow} 1, i a (\partial/\partial \phi) f_i(\phi)] = E[e_i^{<p_i \uparrow} 1, x_t (\gamma_i^{(i)})] = 0 \) is automatically satisfied by \( L_p \)-orthogonality, hence \( S^a = f g \).

4. Super-Revealing CM’s with Integer Nuisance Parameters In practice computing \( \gamma_i^{(i)} = \arg \sup_{\gamma^{(i)}} f(\partial/\partial \gamma_i) E[e_i F(\gamma^{a} (x_t))] \)g and a test statistic functional over \( i \) may be computationally costly. Moreover, the subset \( i \) is itself arbitrary and may be considered to be a nuisance space for small samples under the alternative. See Hansen (1996) for comments on this problem.

Let \( a : \mathbb{R}^k ! \mathbb{R}^k \) be any bounded, one-to-one function, and consider the weight
\[
G_i(m) = u_i^{<p_i \downarrow} 1, i (x_t)^{m_i}.
\]
If $P[E(e_i|x_i) = 0] < 1$ then Theorem 3 of Hill (2006) guarantees

$$E[e_iG_i(m)] \neq 0$$

for infinitely many integers $m = [m_i]_{i=1}^k \in \mathbb{Z}^k$ in general, and specifically in... nitely many $m \in \mathbb{N}^k$.

For example, assume $x_i$ is bounded and use $\alpha (x_i) = [\exp x_i, g_i, \ldots, \exp x_k, i g_i]$.

Then $G_i(m) = \sum_{i=1}^k \alpha (x_i) = \exp m_{0i}, g_i$ reveals model mis-specification for infinitely many $m$. If $x_i$ is not bounded then simply substitute $x_i$ for any bounded one-to-one function.

Now rewrite

$$\gamma^{(i)} = \arg \sup_{\gamma \in \mathbb{R}^k} f(\partial/\partial \gamma_i)\eta(\gamma)g$$

$$\omega (m, \gamma^{(n)}) = E \exp m_{0i} x_i g_i \prod_{i=1}^k \gamma_i \cdot (x_i)F_i^Q \gamma^{(i)}$$

$$H_i(m, \gamma^{(n)}) = \exp m_{0i} x_i g_i \cdot 1(x_i)F_i^Q \gamma^{(1)}, \ldots, a_k(x_i)F_i^Q \gamma^{(k)}$$

$$S_{(m)}^{(n)} = \mathbb{E} \{ \prod_{i=1}^k f m \in \mathbb{Z}^k : E[e_i H_i, i (m, \gamma^{(n)})] = 0 \}$$

The fact that $i$ in Lemma 2 is arbitrary is advantageous here. The claim holds for all $\gamma \in \mathbb{R}^k$ and therefore for every $\gamma \in m \in \mathbb{N}^k$. That said, notice Lemma 2 exploits the fact that $E[e_i F_i] = 0$ under $H_i$ for uncountably in... nitely many $\gamma$ in any arbitrary compact subset $i$, yet we are not guaranteed that $E[e_i G_i (m)] \neq 0$ for any $m$ in any particular subset $i$. Thus, we must take the supremum $\sup_{\gamma \in \mathbb{R}^k} (\partial/\partial \gamma_i)\eta(\gamma)$ over the entire real-line $\mathbb{R}^k$. In order to ensure $(\partial/\partial \gamma_i)\eta(\gamma)$ is bounded we must assume $\gamma^{(i)}$ is... nite.

Together with the fact that the weight $\exp m_{0i} x_i g_i$ reveals mis-specification for in... nitely many $m \in \mathbb{Z}^k \mathcal{F} \mathcal{R}^k$, the following corollary to Lemma 2 is immediate.

**COROLLARY 4** Assume $\gamma^{(i)} \neq 0$ for some compact subset $i \in \mathcal{F} \mathbb{R}^k$. Let $e_i$ be a random variable satisfying $E[e_i] < 1$, and assume $x_i$ is a bounded $\gamma_{i1}$-measurable k-vector. If $P[E(e_i|x_i) = 0] < 1$ then $\omega (m, \gamma^{(n)}) = 0$ if and only if $m = 0$.

**COROLLARY 5** Under the conditions of Lemma 4, if $P[E(e_i|x_i) = 0] < 1$ then $S_{(m)}^{(n)} = f g$ if and only if $e_i? \exp m_{0i} x_i g_i$ and $S_{(m)}^{(n)} = f g$ otherwise.

5. **Test Statistic** Write $\hat{\phi} = \arg \min_{\phi \in \mathbb{R}^k} \sum_{i=1}^P y_i f_i(\phi)g_i y_{i1}$, and write $\hat{f}(\phi) = (\partial/\partial \phi) f(\phi)$. Define the sample conjugate to $\gamma^{(n)}$:

$$\gamma^{(u)} = \arg \inf_{\gamma} \hbar (1/n \sum_{i=1}^n \gamma_i x_i) F_i(\gamma)\eta(\gamma)$$

$$\omega_{(m)}^{(n)} = \arg \sup_{\gamma \in \mathbb{R}^k} f(\partial/\partial \gamma_i)\eta(\gamma)g$$

$$H_i(m, \gamma^{(n)}) = \exp m_{0i} x_i g_i \cdot 1(x_i)F_i^Q \gamma^{(1)}, \ldots, a_k(x_i)F_i^Q \gamma^{(k)}$$

$$S_{(m)}^{(n)} = \mathbb{E} \{ \prod_{i=1}^k f m \in \mathbb{Z}^k : E[e_i H_i, i (m, \gamma^{(n)})] = 0 \}$$

6.
In order to reduce notation write
\[
\theta' f_{\gamma, \gamma^{(n)}} g 2 \mathcal{L} \quad i \mathcal{L}^{(n)} \quad \text{and} \quad \hat{\theta} = f_{\gamma, \gamma^{(n)}} g.
\]

Define the sample vector moment
\[
\mathcal{L}(\hat{\theta}) = 1/\mathcal{P}_n \sum_{i=1}^{n} (y_i - f_i(\phi))_{<\mathcal{P}_i = 1>} g_i(\theta) + o_p(1)
\]
where \(H_i(\phi)\) is defined in (1). We use a Pitman \(\mathcal{P}_n\)-local alternative of the form
\[
H_{I^2}^1 : y_i = f_i(\phi_0) + u_i/\mathcal{P}_n + \epsilon_i,
\]
where \(E[\epsilon <\mathcal{P}_i = j] = 0\). We assume \(u_i\) is measurable with respect to \(=i_1\) and governed by a non-generate distribution. The null hypothesis is \(u_i = 0 \ a.s.,\) and a global alternative is simply
\[
H_{I^2}^G : y_i = f_i(\phi_0) + u_i + \epsilon_i.
\]

From the mean-value-theorem and Assumption A, under \(H_{I^2}^1\) we may write for some sequence \(u_i^n\) satisfying \(u_i^n 2 [0, u_i]\) and \(u_i^n = o_p(\mathcal{P}_n)\)
\[
\mathcal{L}(\hat{\theta}) = 1/\mathcal{P}_n \sum_{i=1}^{n} (y_i - f_i(\phi))_{<\mathcal{P}_i = 1>} g_i(\theta) + o_p(1)
\]
\[
= 1/\mathcal{P}_n \sum_{i=1}^{n} (y_i - f_i(\phi))_{<\mathcal{P}_i = 1>} g_i(\theta)
\]
\[
+ (p i 1/n) \sum_{t=1}^{n} j_{ut_i}/\mathcal{P}_n + \epsilon_i j_{ut_i} g_i(\theta) + o_p(1)
\]
\[
= z_n(\theta) + o_p(1),
\]
say, where
\[
g_i(\theta) = H_i(\theta) i b(\theta, \phi_0)A(\phi_0) i ^{1/2} f_i(\phi_0) 2 \mathcal{P}_n\]
\[
A(\phi_0) = (p i 1) \lim_{n \to \infty} (1/n) \sum_{i=1}^{n} (y_i - f_i(\phi_0))_{<\mathcal{P}_i = 1>} 2 f_i(\phi_0) \partial^2 f_i(\phi_0)
\]
\[
b(\theta, \phi_0) = (p i 1) \lim_{n \to \infty} (1/n) \sum_{i=1}^{n} (y_i - f_i(\phi_0))_{<\mathcal{P}_i = 1>} 2 H_i(\theta) \partial^2 f_i(\phi_0)
\]
\[
\eta(\theta) = (p i 1) \lim_{n \to \infty} (1/n) \sum_{i=1}^{n} (y_i - f_i(\phi_0))_{<\mathcal{P}_i = 1>} 2 u_i g_i(\theta)
\]
\[
\Psi(\theta) = \lim_{n \to \infty} (1/n) \sum_{i=1}^{n} (y_i - f_i(\phi_0))_{<\mathcal{P}_i = 1>} 2 g_i(\theta) g_i(\theta) 0.
\]

5.1 Weak Convergence

Weak convergence on a space of continuous real functions \(C[\mathcal{L}]\) requires convergence of finite distributions and tightness of the vector sequence \(f_z_n(\hat{\theta})g\) on \(\mathcal{L}\). See Pollard (1984) and Billingsley (1999). In the integer nuisance parameter case (i.e. \(\gamma = m 2 \mathbb{Z}_+\)) tightness is trivial. See Billingsley (1999) and Hill (2006). Consider, then, \(\gamma 2 i \mathbb{Z}_+.\)

If a constant term is included then the matrix \(\Psi(\theta) = \Psi(\gamma^{(n)})\) may be close to singular if \(\gamma\) is near zero. If \(\gamma = 0\) then \(\Psi(0, \gamma^{(n)}) 2 \mathbb{R}^{k+1}\) will have
rank $k$ due to $\xi(\hat{\theta}) = 1/\sqrt{p - n} \sum_{t=1}^{n} [0, a_{1}(x_{t})F_{t}^{\gamma(1)}(z), \ldots, a_{k}(x_{t})F_{t}^{\gamma(k)}(z)]^{\text{tr}}$. We ameliorate the problem by bounding $\gamma$ away from 0.

For arbitrary $\xi > 0$ denote the subspace

$$E_{\xi} = \{ x_{t} : \gamma(0) > \xi \} \text{ where } i_{x} = f_{\gamma(0)} : j_{x} > \xi.$$ Denote by $\lambda_{\min}(\theta)$ the minimum eigenvalue of $\xi_{\theta}(\theta)$. 

**A ssumption C** $\inf_{\theta \in \xi_{\min}(\theta)} \lambda_{\min}(\theta) > 0$.

**L E M M A 6** Let Assumption C hold. Let $z(\theta)$ denote a Gaussian element of $C[\xi]_{\xi}$ with mean function $\xi(\theta)^{1/2} \eta(\theta)$ and covariance function $E[z(\theta_{1})z(\theta_{2})]$ $\xi(\theta_{1})^{1/2} \xi(\theta_{2})^{1/2}$. Under Assumption A and $H_{1}^{G}$, $\eta(\theta)^{1/2} z_{n}(\theta)$ ! $z(\theta)$ pointwise in $\xi$. Moreover $\xi(\theta)^{1/2} z_{n}(\theta) / n_{i}$ 1 with probability one under $H_{1}^{G}$ for every $\theta \in \xi$.

**L E M M A 7** Under Assumptions A-C and $H_{1}^{G}$ the sequence $f\xi(\theta)^{1/2} z_{n}(\theta) i \eta(\theta) g$ is tight on $\xi$.

**T H E O R E M 8** Under Assumptions A-C and $H_{1}^{G}$, $\xi(\theta)^{1/2} z(\theta) \xi_{\xi}(\theta)$ on $C[\xi]$ where $z(\theta)$ is defined in Lemma 6.

### 5.3 Super-Consistent Test Statistic

Consider a standard Lagrange multiplier test statistic

$$T_{n}(\theta) = T_{n}(\gamma, \delta^{(a)}) = \xi(\gamma, \delta^{(a)})(\xi(\gamma, \delta^{(a)})^{1/2} \xi(\gamma, \delta^{(a)}))^{1/2} \xi(\gamma, \delta^{(a)}).$$ Theorem 8 and the continuous mapping theorem guarantee $T_{n}(\gamma, \delta^{(a)})$ $\chi^{2}(k + 1)$ under $H_{0}$. Under $H_{1}^{G}$ the test statistic $T_{n}(\gamma, \delta^{(a)})$ reveals model misspecification asymptotically with probability one for every nuisance parameter $\gamma \in (0: T_{n}(\gamma, \delta^{(a)}) / n \in \gamma \text{ a.s.})$.

In practice, however, the analyst may want to improve small sample power by considering continuous, $f_{(1)}$-measurable functions $h : R_{+} \rightarrow R_{+}$, including supremum and average functionals. See Davies (1977), Andrews (1993), King and Shively (1993), and Andrews and Ploberger (1994). Whether such functionals actually improve small sample power over a test with a randomized nuisance parameter $\gamma$ is considered below.
6. Monte Carlo Study Write $x_{j,t_1}= [x_{t_1,1}, \ldots, x_{t_1,p_j}]$. We simulate 100 of the following processes

$$H_0: x_t = \phi_0 x_{t-1} + \epsilon_t$$

$$H^S_1: x_t = \phi_0 x_{t-1} + \phi_2 x_{t-2} \mathbb{I}(x_{t-1} > 0) + \epsilon_t$$

$$H^{ES}_1: x_t = \phi_0 x_{t-1} + \phi_2 x_{t-2} \mathbb{I} \exp \gamma x_{t-1}^2 + \epsilon_t$$

$$H^{LS}_1: x_t = \phi_0 x_{t-1} + \phi_2 x_{t-2} \mathbb{I} [1 \exp \gamma x_{t-1}^2 + \epsilon_t$$

$$H^{1N}: x_t = \phi_0 x_{t-1} + [1 \exp \gamma x_{t-1}^2 + \epsilon_t$$

$$H^{ALL}: \text{randomized } H_1,$$

where $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0,1)$. In all cases each $p_j$ is randomly selected from the set $1, \ldots, 10g$ and each $\phi_i \overset{iid}{\sim} \mathcal{N}(0,1)$ subject to all roots lying outside the unit circle; $\beta$ is randomly selected from $[0, 0.9]$ and $\gamma$ is randomly selected from $[0.5, 1.0]$ under $H_0$ the process is $AR(p)$; under $H^{S}_1$ the process is $SETAR(p, 2)$ and $H^{ES}_1$ the process is $ESTAR$ and $LSTAR$ respectively; under $H^{1N}$ it is an $AR(p)$-ANN (artificial neural network); under $H^{ALL}$ the alternative is selected at random from those just described. Sample sizes are $n = 200, 500g$.

We estimate a null model by fitting an $AR(p)$ to the series $f X_i g_{i,1}^{n}$, where $p$ is selected by minimizing the AIC. We test the residuals for omitted nonlinearity at the 5%-level using $\sup T_{n}(\gamma, \phi^{(n)})$ and $\sup T_{n}(m, \phi^{(n)})$, and randomized tests $T_{n}(\gamma, \phi^{(n)})$ and $T_{n}(m, \phi^{(n)})$ where $\gamma$ and $m$ are randomly selected from subsets $i_n^m$ and $N^p_n$, respectively; under $H^{1N}$ it is an $AR(p)$-ANN (artificial neural network); under $H^{ALL}$ the alternative is selected at random from those just described. Sample sizes are $n = 200, 500g$.

For $\sup T_{n}(\gamma, \phi^{(n)})$, we use the sample moment conditions $\mathbb{E}(\gamma, \phi^{(n)}) = 1/p_n P_{i=1}^{n} \mathbb{E}(\gamma, \phi^{(n)}) T_{n}(\gamma, \phi^{(n)})$, where $H_i(\gamma, \phi^{(n)})$ is one of the following:

$$H_0: \text{exp} \gamma x_{t} g_{i}^{1}, \text{exp} \gamma x_{t} g_{i}^{0}$$

$$H_1: 1 + \text{exp} \gamma x_{t} g_{i}^{1}, \text{exp} \gamma x_{t} g_{i}^{0}$$

6.1 Nuisance Parameter Spaces

The set $i_n^p = f \gamma_1, \ldots, \gamma_{j_n} g$ is constructed from $J_n$ randomly selected $\gamma_i \overset{iid}{\sim} \mathcal{N}(0,1)$. For $\sup T_{n}(m, \phi^{(n)})$ we use only the exponential and construct the set of integers $N^p_n$ as follows. Denote by $j^{(k)}_p$ a $p$-vector with the value $j$ for the $k^{th}$ component and the value $k$ in all other components. Let $N^p_n$ be a set with $\lfloor \log n \rfloor$ integer vectors randomly selected from $[0, 0.101, \ldots, 0.999, 1.0]$ and $\lfloor \log n \rfloor$ vectors $\lfloor \log n \rfloor$ to 0.1 $1.0/10$, $\lfloor \log n \rfloor/10$, $\lfloor \log n \rfloor/100/81$, $\lfloor \log n \rfloor/100/81$, $\lfloor \log n \rfloor/100/81$. Finally, let $N^p_n$ denote the set of all integers $f^{(0)} \phi_{j,1}^{(0)} g_{j,1}^{(0)} 0$ and $f^{(1)} \phi_{j,2}^{(1)}$.
\( N^n_p \) [\( N^n_p \) [\( N^n_p \) [\( N^n_p \)], which contains simple integer vectors, randomized vectors, and \( N^n_p \) \( N^n \) as \( n \to 1 \).

6.2 Uniformly Most Powerful [UMP] Tests

In order to gauge the power of the proposed tests we compute UMP tests against each alternative. Because \( \phi_1 \) and each \( \gamma_i \) are known within the simulation, each model can be written as

\[
y_t(\phi_1) = \phi_0 z_t(\gamma) + \epsilon_t,
\]

where \( y_t(\phi_1) = x_t^i \phi_0^i x_t^i 1, z_t(\gamma) = x_t^i 1 \exp(x_t^i \gamma) \) under \( H^S_1 \), \( z_t(\gamma) = x_t^i 1 \) under \( H^U_1 \), and \( z_t(\gamma) = [1 + \exp(x_t^i \gamma)]^{-1} \) under \( H^A_1 \). The errors are known to be iid standard normal, hence the UMP test statistic reduces to

\[
W_n(\gamma) = y(\phi_1) \phi_0 \gamma [z(\gamma) \phi_0 \gamma]^{-1} z(\gamma) \phi_0 y(\phi_1).
\]

6.2 Simulation Results

See Table 1 for sup-tests and Table 2 for randomized tests. Two important observations are immediately apparent. First, with respect to non-UMP tests constructing a test functional in order to improve small sample power is utterly ineffective. Indeed, the randomized test exhibits a non-negligible power improvement over the sup-test in nearly all cases, although we expect the power improvement to shrink with the sample size (see below). Second, in some cases the sup-test obtains power relatively near UMP tests. The randomized super-consistent test, however, generates empirical power nearly identical to UMP tests in most cases. Only the randomized logistic test \( T_n(\gamma, \gamma^{(n)}) \) is noticeably dominated by the associated UMP test.

Finally, we perform a simulation study in which only ESTAR and LSTAR processes are generated for \( n \in \{50, 75, \ldots, 1975, 2000\} \). Figure 1 plots the resulting rejection frequencies of the randomized super-consistent and UMP tests (tests are performed at the 5% level). The relative performance of the weight-specific super-consistent tests corresponds to the above two cases \( n \in \{200, 500\} \). When the alternative is LSTAR the super-consistent exponential test power (nearly) converges to the UMP test power at roughly \( n = 500 \). In this case logistic test power is much slower to converge, roughly matching UMP test power at \( n = 3000 \) (not shown).
Appendix 1: Assumptions

Assumption A 1: The parameter space is a compact subset of $\mathbb{R}^k$. $\phi_0 = \arg\inf_{\phi \in \mathcal{D}} E_j [y_t i f_i(\phi)]$ is twice continuously differentiable on $\mathcal{D}$, for some $c > 1$ and $f_i(\phi)$ are $\mathcal{D}$-measurable, where $\mathcal{D}$ is the sequence of $\sigma$-algebras generated by $(x_t : \tau \cdot t + 1)$. Moreover, $E_j [\epsilon_i^{pi} < 1 > | j = i | = 0$ a.s. for some $p > 2$. For some $\phi_0 = \arg\inf_{\phi \in \mathcal{D}} E_j [y_t i f_i(\phi)]$ and $\phi_0 = \arg\inf_{\phi \in \mathcal{D}} E_j [y_t i f_i(\phi)]$, where $A_1(\phi) = \arg\min_{\phi \in \mathcal{D}} E_j [y_t i f_i(\phi)]$, where $A_1(\phi)$ is a non-stochastic matrix such that $A_1(\phi)$ is positive definite. For some stochastic sequence $p_1 g_1$ satisfying $u_t^2 = 0$, $u_t^1 = 0$ and $u_t^1 = 0$ the $L_2$ estimator $\hat{\phi} = \arg\min_{\phi \in \mathcal{D}} E_j [y_t i f_i(\phi)]$ satisfies

\[ p_n^{-1} \hat{\phi}_n = A_2(\phi_0) \bar{X}_n \leq \frac{\partial f_i(\phi)}{\partial \phi} \frac{p_n}{n} + \frac{\bar{X}_n}{n} \epsilon_i + \frac{p_n}{n} u_t \frac{\partial f_i(\phi)}{\partial \phi} + o_p(1). \]

Assumption A 2: For each $i = 1 \ldots k$, let $H_i(\gamma, \tau, \phi_0) = \arg\min_{\gamma, \tau} E [\epsilon_i^{pi} (\tau) f_i(\phi)]$, where $\tau 2 = 0$ and $\tau 1 = \epsilon_1 \epsilon_2 \epsilon_\tau \epsilon_{0,2} R^{2k}$. Then $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} H_i(\gamma, \tau, \phi_0) < C$ and $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} b_i(\gamma, \tau, \phi_0) = 0$ where $b_i(\gamma, \tau, \phi_0)$ is a non-stochastic function satisfying $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} b_i(\gamma, \tau, \phi_0) = C$ and $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} b_i(\gamma, \tau, \phi_0) = C$.

Assumption A 3: For each $i = 1 \ldots k$, let $\delta_i(\gamma, \tau, \phi_0) = \arg\min_{\gamma, \tau} E [\epsilon_i^{pi} (\tau) f_i(\phi)]$, where $\tau 2 = 0$ and $\tau 1 = \epsilon_1 \epsilon_2 \epsilon_{0,2} R^{2k}$. Then $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) < C$ and $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) = 0$ where $\delta_i(\gamma, \tau, \phi_0)$ is a non-stochastic function satisfying $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) = C$ and $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) = C$.

Assumption A 4: Write $\theta = f_1(\gamma, \gamma^{(n)})$ and $\theta_2 = f_1(\gamma, \gamma^{(n)})$. Then $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) < C$ and $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) = C$.

Assumption A 5: For some $\delta > 0$, $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) < C$ and $\sup_{\phi_2 \in \mathcal{D}, \tau_2 \in \tau} \max_{i} \delta_i(\gamma, \tau, \phi_0) < C$.
Appendix 2: Formal Proofs

Proof of Lemma 2. The construction of $\gamma^{(i)}$ implies for all $\gamma \geq 2$:

$$E[e_i^a_j(x_i)F_i^0(\gamma)] \cdot E[e_i^a_j(x_i)F_i^0(\gamma^{(i)})] = \sup_{\gamma \geq 2} E[e_i^a_j(x_i)F_i^0(\gamma)].$$

Assumptions A and B imply

$$\omega(0, \gamma^{(n)}) = E[e_iF_i(0)] = 0.$$

Now differentiate $\omega(\gamma, \gamma^{(n)})$ with respect to each $\gamma_j$, and add and subtract $E[e_i^a_j(x_i)F_i^0(0)]$. By construction

$$\omega(\gamma, \gamma^{(n)}) = E[e_i^a_j(x_i)f F_i^0(\gamma)] - E[e_i^a_j(x_i)f F_i^0(\gamma^{(j)})] = 0.$$

Thus $\omega(\gamma, \gamma^{(n)})$ is zero at $\gamma = 0$ and is weakly decreasing in $\gamma$. From (3) we know $E[e_i^a_j(x_i)f F_i^0(\gamma^{(j)})] = 0$, $8j = 1...k$.

In order to sharpen the weak inequality in (3) consider two possible cases.

Case 1: $E[e_i^a_j(x_i)f F_i^0(\gamma^{(j)})] = 0$.

Trivially

$$E[e_i^a_j(x_i)f F_i^0(\gamma)] = 0.$$  

Lemma 1 therefore implies there exists an open neighborhood $N_0 \ni \gamma$ of zero satisfying

$$E[e_i^a_j(x_i)f F_i^0(\gamma)] = 0 \quad 8j \geq 2 N_0/0.$$  

By assumption $E[e_i^a_j(x_i)f F_i^0(\gamma^{(j)})] = 0$ hence from (4) we deduce

$$E[e_i^a_j(x_i)f F_i^0(\gamma)] < 0 \quad 8j \geq 2 N_0/0.$$  

Thus $\omega(\gamma, \gamma^{(n)})$ is zero at $\gamma = 0$, strictly decreasing arbitrarily close to $\gamma = 0$, and weakly decreasing everywhere else. Thus $\omega(\gamma, \gamma^{(n)}) \geq 0$.

Case 2: $E[e_i^a_j(x_i)f F_i^0(\gamma^{(j)})] = 0$.

Using (3) we easily deduce $8j = 1...k$

$$\omega(0, \gamma^{(n)}) = \sup_{\gamma \geq 2} E[e_i^a_j(x_i)f F_i^0(\gamma^{(j)})] = 0.$$  

Again, $\omega(\gamma, \gamma^{(n)})$ is zero at $\gamma = 0$, strictly decreasing at $\gamma = 0$ and weakly decreasing everywhere else.

Proof of Theorem 3. Assume $P(E[e_i] = 0) < 1$ and recall $x_i$ does not contain a constant term. By Lemma 2 we know $8j \geq 0$

$$\omega(\gamma, \gamma^{(n)}) = E[e_i^aF_i(\gamma)] = \sup_{\gamma \geq 2} E[e_i^a_j(x_i)f F_i^0(\gamma^{(j)})] = 0.$$  

12
Trivially, therefore, at least one moment condition $E[e_t F_t(\gamma)]$ must be non-zero, hence $E[e_t H_t(\gamma, \gamma^{(0)})] \equiv 0$ for every $\gamma \equiv 0$.

Finally, under Assumptions A and B

$$E[e_t H_t(0, \gamma^{(n)})] = 0, E[e_t^a 1(x_i) F_t^{(1)}(\gamma^{(1)})] \ldots, E[e_t^a \varphi(x_i) F_t^{(1)}(\gamma^{(k)})]$$

hence $E[e_t H_t(0, \gamma^{(n)})] = 0$ if and only if $e_t$ is orthogonal to $\bar{a}_p \varphi_i(x_i)$ for $F_t^{(1)}(\gamma^{(i)}) g_{t-1}^i$.

**Proof of Lemma 6.** Invoking a Cramér-Wold type device and Assumption A, under $H_t^{2}$ it suffices to prove for any $r \geq 1$, $r_0 \equiv 1$,

$$1/n \sum_{i=1}^{n} e_{t}^{<r_i> 1/2} g_i(\theta) = N(0, I)$$

Clearly $f_{<r_i> 1/2} g_{t-1}(\theta) = _{0}^{1/2} \psi_i(\theta)$ forms a martingale difference sequence due $E[e_t^{<r_i> 1/2} g_i(\theta)] = 0$ under Assumption A. The claim under $H_t^{2}$ now follows from Assumption A.iii, Lemma A.1, below, and Bierens’ (1994: Lemma 6.1.7) generalization of Mclish’s (1974) martingale difference central limit theorem.

Write $w_t(\theta) = (y_t, f_\theta(\psi))^{<r_i> 1/2} g_i(\theta)$. Under $H_t^{2}$ and Lemma 1 we deduce $E[w_t(\theta)]=_{0}^{1/2} \psi_i(\theta)$ for $r \geq 2 \epsilon$ $\epsilon$. Therefore, we need only show (2) implies

$$z_{n}(\theta) / \sqrt{n} = 1/n \sum_{i=1}^{n} (w_t(\theta) - E[w_t(\theta)]) = 0$$


**Proof of Lemma 7.** For any $r \geq 1$, write $e_t = e_{t}^{<r_i> 1/2}$ and

$$1/n \sum_{i=1}^{n} e_{t}^{<r_i> 1/2} g_i(\theta) = 1/n \sum_{i=1}^{n} e_{t}^{r_0} \psi_i(\theta)$$

say, where $\psi_i(\theta) = g_i(\theta) / 2$ and $w_t(r, \theta) = r_0 \psi_i(\theta)$. Using Lemma A.1 of Bierens and Ploberger (1997) we need to show

$$\limsup_{n \to \infty} 1/n \sum_{i=1}^{n} E[e_t^2 K_i^2] < 1$$

$$\sup_{r_0=1} \limsup_{n \to \infty} 1/n \sum_{i=1}^{n} E[e_t^2 w_t(r, \theta_0)^2] < 1$$

for at least one point $\theta_0$, we then use

$$K_i = \sup_{\theta \in \epsilon} j(\theta / \theta_0) w_t(r, \theta).$$

The second inequality in (4) follows from Assumption A and $r_0 = 1$.

$$\sup_{r_0=1} E[e_t^2 w_t(r, \theta)^2] \cdot k e_i k_2 \psi_i(\theta) / 2 \sup_{r_0 \epsilon} j_2 j^2 \sup_{r_0 \in \epsilon} j \psi_i(\theta) j^2 \epsilon^2 < C$$
where \( j \| \theta \|^ {1/2} < 1 \) is an easy consequence of Assumptions A and C.

For the first inequality in (4) we will prove \( E[\epsilon_i^2 K_i^2] \). By the Cauchy-Schwartz inequality

\[
\sup_{\psi_j} E[\epsilon_i^2 K_i^2] \cdot k_i K_i^2 \leq j \sup_{\partial / \partial \theta} \psi_j(\theta) \| \psi_j(\theta) \|^2.
\]

The \((l,j)^{th}\)-component \((\partial / \partial \theta_l)\psi_{t,j}(\theta)\) of the \(k(k+1)\) \((k+1)\)-matrix \((\partial / \partial \theta)\psi_j(\theta)\) is exactly

\[
(\partial / \partial \theta_l) \psi_{t,j}(\theta) = X_{i=1}^{k+1} (\partial / \partial \theta_l) \| \theta \|^{1/2} g_{t,i}(\theta)
+ X_{i=1}^{k+1} \| \theta \|^{1/2} (\partial / \partial \theta_l) g_{t,i}(\theta).
\]

Using Minkowski's inequality repeatedly and Lemma A.2,

\[
\sup_{\partial / \partial \theta} \| \psi_j(\theta) \|^2
\leq X_{l=1}^{k(k+1)} X_{j=1}^{k+1} \sup_{\partial / \partial \theta} \| \theta \|^{1/2} g_{t,i}(\theta)
+ X_{i=1}^{k+1} \| \theta \|^{1/2} (\partial / \partial \theta_l) g_{t,i}(\theta)
+ \| g_{t,i}(\theta) \|^{1/2} k(k+1)^2 \sup_{\partial / \partial \theta} \| \theta \|^{1/2} (\partial / \partial \theta_l) g_{t,i}(\theta)
+ \| g_{t,i}(\theta) \|^{1/2} k(k+1)^2 \sup_{\partial / \partial \theta} \| \theta \|^{1/2} (\partial / \partial \theta_l) g_{t,i}(\theta).
\]

\( C \).
**Appendix 3: Supporting Lemmata**

**Lemma A.1** Under Assumption A for each \( \theta \in \xi \) and every \( r \in \mathbb{R}^{k+1}, r^0 \)

\[
\lim_{n \to \infty} 1/n \sum_{t=1}^{n} \epsilon^r_{(t)} r^0 = 1,
\]

and for some \( \kappa > 0 \)

\[
\lim_{n \to \infty} \left( \frac{1}{n} \sum_{t=1}^{n} E \epsilon^r_{(t)} r^0 \right)^{2+\kappa} = 0.
\]

**Proof of Lemma A.1.** From the normalization \( \pi \rho_\infty = 1 \) and Assumption A it is easy to show \( E\left[ (\epsilon^r_{(t)} r^0)^2 \right] = 1 \) for all \( t \in \mathbb{N} \). Limit (5) then follows from Assumption A.4:

\[
\sup_{\theta \in \xi} \frac{1}{n} \sum_{t=1}^{n} j^2(\theta) g_\rho(\theta)^0 \equiv o_p(1),
\]

Limit (6) follows from the following bound. By \( \ell_1 \)-norm properties, the envelope inequality, and Assumption A, for some small \( \kappa > 0 \) and some \( \sigma \)-finite \( M > 0 \)

\[
E j^2(\theta)^{2+\kappa} = \sup_{\theta \in \xi} \frac{1}{n} \sum_{t=1}^{n} j^2(\theta) g_\rho(\theta)^0 \equiv o_p(1),
\]

where \( j^2 < 1 \) is trivial. Thus, \( p \)

\[
\lim_{n \to \infty} \left( \frac{1}{n} \sum_{t=1}^{n} E j^2(\theta)^{2+\kappa} = o(1/n^{2+\kappa}) \right).
\]

**Proof of Lemma A.2.**

i. Liapunov's inequality applies for some \( \sigma \)-finite \( B > 0 \):

\[
\left[ B(\theta)^{1/2} \right]^2 = B \left( \frac{1}{2} \sum_{i=0}^{\infty} \lambda_i (\theta) \right)^{1/2} = B \left( \frac{1}{2} \sum_{i=0}^{\infty} \lambda_i (\theta) \right)^{1/2}
\]

where \( j^2 < 1 \) is trivial. Thus, \( p \)

\[
\lim_{n \to \infty} \left( \frac{1}{n} \sum_{t=1}^{n} E j^2(\theta)^{2+\kappa} = o(1/n^{2+\kappa}) \right).
\]
hence
\[
\sup_{0 < \ell \leq C} j \cdot (\theta)^i 1/2 \cdot B \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \cdot 1/2 < C,
\]
which is guaranteed for some finite $C$ by Assumption C: $\inf_{0 < \ell \leq C} \lambda_{\min}(\theta) > 0$.

ii. By standard properties of matrix differentiation
\[
(\partial/\partial \theta_i) (\theta)^i 1/2 = \frac{1}{2} (\theta)^i 1/2 E \cdot (\partial/\partial \theta_i) (\theta) E \cdot (\theta)^i 1/2.
\]
Hence, for some finite $B > 0$, by Liapunov's inequality and (i),
\[
\sup_{0 < \ell \leq C} \cdot \left( (\partial/\partial \theta_i) (\theta)^i 1/2 \cdot B \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \right)^{1/2} \sup_{0 < \ell \leq C} \cdot j \cdot (\partial/\partial \theta_i) (\theta) j \]
where $\inf_{0 < \ell \leq C} \lambda_{\min}(\theta) > 0$ by Assumption C. The proof is complete when we show the $L_1$-normed $j j (\partial/\partial \theta_i) (\theta) j j$ is uniformly bounded by some finite $M > 0$.

The covariance matrix derivative $(\partial/\partial \theta_i) (\theta)$ is computed as
\[
(\partial/\partial \theta_i) (\theta) = \frac{1}{2} E \cdot (\partial/\partial \theta_i) _{t i} (\theta) _{t i} (\theta) \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \cdot \sup_{0 < \ell \leq C} \cdot j j (\partial/\partial \theta_i) (\theta) j j .
\]
By the envelope and repeated Cauchy-Schwartz inequalities,
\[
\sup_{0 < \ell \leq C} \cdot \left( (\partial/\partial \theta_i) (\theta)^i 1/2 \cdot B \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \right)^{1/2} \sup_{0 < \ell \leq C} \cdot j j (\partial/\partial \theta_i) (\theta) j j \]

\[
= 2 \cdot \inf_{0 < \ell \leq C} \cdot \left( (\partial/\partial \theta_i) (\theta)^i 1/2 \cdot B \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \right)^{1/2} \sup_{0 < \ell \leq C} \cdot j j (\partial/\partial \theta_i) (\theta) j j .
\]

\[
= 2 \cdot \inf_{0 < \ell \leq C} \cdot \left( (\partial/\partial \theta_i) (\theta)^i 1/2 \cdot B \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \right)^{1/2} \sup_{0 < \ell \leq C} \cdot j j (\partial/\partial \theta_i) (\theta) j j .
\]

\[
= 2 \cdot \inf_{0 < \ell \leq C} \cdot \left( (\partial/\partial \theta_i) (\theta)^i 1/2 \cdot B \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \right)^{1/2} \sup_{0 < \ell \leq C} \cdot j j (\partial/\partial \theta_i) (\theta) j j .
\]

\[
= 2 \cdot \inf_{0 < \ell \leq C} \cdot \left( (\partial/\partial \theta_i) (\theta)^i 1/2 \cdot B \cdot \inf_{0 < \ell \leq C} \lambda_{\min}(\theta) \right)^{1/2} \sup_{0 < \ell \leq C} \cdot j j (\partial/\partial \theta_i) (\theta) j j .
\]
References


Table 1: Supremum Tests

<table>
<thead>
<tr>
<th></th>
<th>$H_0$</th>
<th>$H_1^L$</th>
<th>$H_1^E$</th>
<th>$H_1^2$</th>
<th>$H_1^{AN}$</th>
<th>$H_1^{ALL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-sup$_{\gamma_2} T_n(\gamma)$</td>
<td>0.05</td>
<td>.49 .56</td>
<td>.25 .39</td>
<td>.45 .55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-sup$_{\gamma_2} T_n(\gamma)$</td>
<td>0.06</td>
<td>.73 .67</td>
<td>.50 .41</td>
<td>.55 .55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sup$_{n^2 \hat{\gamma}} T_n(m)$</td>
<td>0.06</td>
<td>.67 .66</td>
<td>.48 .37</td>
<td>.54 .54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-sup$_{\gamma_2} W_n(\gamma)$</td>
<td>0.05</td>
<td>.91 .84</td>
<td>.68 .67</td>
<td>.72 .72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-sup$_{\gamma_2} W_n(\gamma)$</td>
<td>0.06</td>
<td>.89 .74</td>
<td>.70 .66</td>
<td>.73 .73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sup$_{n^2 \hat{\gamma}} W_n(m)$</td>
<td>0.06</td>
<td>.82 .77</td>
<td>.86 .70</td>
<td>.75 .75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes: a. Super-consistent tests (L = logistic; E = exponential).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Uniformly most-powerful tests.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Randomized Tests

<table>
<thead>
<tr>
<th></th>
<th>$H_0$</th>
<th>$H_1^L$</th>
<th>$H_1^E$</th>
<th>$H_1^2$</th>
<th>$H_1^{AN}$</th>
<th>$H_1^{ALL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-$T_n(\gamma)$</td>
<td>0.09</td>
<td>.75 .69</td>
<td>.56 .55</td>
<td>.64 .69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-$T_n(\gamma)$</td>
<td>0.04</td>
<td>.80 .75</td>
<td>.72 .51</td>
<td>.66 .66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_n(m)$</td>
<td>0.05</td>
<td>.82 .74</td>
<td>.51 .65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-$W_n(\gamma)$</td>
<td>0.03</td>
<td>.83 .72</td>
<td>.72 .57</td>
<td>.71 .71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-$W_n(\gamma)$</td>
<td>0.02</td>
<td>.82 .81</td>
<td>.79 .53</td>
<td>.72 .72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_n(m)$</td>
<td>0.03</td>
<td>.83 .80</td>
<td>.82 .54</td>
<td>.67 .67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-$T_n(\gamma)$</td>
<td>0.08</td>
<td>.74 .69</td>
<td>.67 .73</td>
<td>.72 .72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-$T_n(\gamma)$</td>
<td>0.07</td>
<td>.85 .82</td>
<td>.80 .69</td>
<td>.82 .82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_n(m)$</td>
<td>0.05</td>
<td>.95 .79</td>
<td>.81 .79</td>
<td>.88 .88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-$W_n(\gamma)$</td>
<td>0.03</td>
<td>.94 .92</td>
<td>.84 .74</td>
<td>.81 .81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-$W_n(\gamma)$</td>
<td>0.03</td>
<td>.86 .93</td>
<td>.91 .70</td>
<td>.85 .85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_n(m)$</td>
<td>0.07</td>
<td>.96 .84</td>
<td>.88 .80</td>
<td>.90 .90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes: a. Real-valued $\gamma$ are randomly selected from $[.5,10]$.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Integer-valued $m$ are randomly selected from $N$.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Empirical Power Against STAR Alternatives

Exponential Super-Consistent CM Test and UMP Test Power Against LSTAR Alternative

Logistic Super-Consistent CM Test and UMP Test Power Against LSTAR Alternative

Exponential Super-Consistent CM Test and UMP Test Power Against ESTAR Alternative

Logistic Super-Consistent CM Test and UMP Test Power Against ESTAR Alternative