9-19-2006

Super-Consistent Tests of Lp-Functional Form

Jonathan B. Hill
Department of Economics, Florida International University

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 mis-specification for any nuisance parameter value. When the sufficient condition fails an alternative moment condition is guaranteed to work. A simulation study clearly demonstrates the superiorty of a randomized test: randonly selecting the nuisance parameter leads to more power than average- and supremum-test functionals, and obtains empirical power nearly equivelant 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 mis-specification.

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 $F(\tau' x_t)$ indexed by a real-valued nuisance vector $\tau \in \Gamma \subset \mathbb{R}^k$, effectively producing uncountably many moment conditions which "reveal" model mis-specification. Expanding upon Bierens' (1990) seminal Lemma 1, Stinchcombe and White (1998) show that any real analytic

*Dept. of Economics, Florida International University, Miami, Fl; jonathan.hill@fiu.edu; http://www.fiu.edu/~hilljona
JEL classification: C12, C45, C52.
Key words: consistent test; conditional moment test; best L_p-predictor; nonlinear model.

function $F(A(x_t))$ that is non-polynomial can reveal model mis-specification, where $A : \mathbb{R}^k \to \mathbb{R}$ is affine.

Hill (2006) takes a difference tack by constructing a class of revealing weights $G_t(m) = \prod_{i=1}^{k} \Psi_i(x_t)^{m_i}$ based on integer nuisance parameters $m \in \mathbb{Z}^k$ where $\Psi : \mathbb{R}^k \to \mathbb{R}^k$ 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 $S \subset \Xi$ 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 $\mathbb{R}^k/S$ 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 $\xrightarrow{p}$ denotes convergence in probability, or finite distributions. $\Rightarrow$ denotes weak convergence on a function space. $sp(\{z_i\}_{i=1}^{n})$ denotes the span of $z_1, ..., z_n$, and $\overline{sp}(\{z_i\}_{i=1}^{n})$ denotes the closed linear span. $|\cdot|$ 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 $\{y_t, \mathbf{x}_t\} \in \mathbb{R} \times \mathbb{R}^{k-1}$ be a strictly stationary, ergodic stochastic process in $L_p(\Omega, \Im, \mu)$, $p \in (1, 2]$, with nondegenerate continuous marginal distributions, $\Im = \sigma(\cup_t \Im_t)$, $\Im_{t-1} \subseteq \Im_t = \sigma(\{\mathbf{x}_t\} : \tau \leq t+1)$. Define $x_t \equiv (1, \mathbf{x}_t')'$. The regressors $\mathbf{x}_t$ 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, and see Hill (2006) for complete details on the following set-up.

Denote by $Q_{t-1}y \equiv Q(y_t|\mathcal{F}_{t-1})$ the orthogonal L_p-metric projection of y_t onto the space spanned by $\{x_{t-i}\}_{i=0}^{\infty}$. If we write

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

then clearly $E[e_t z_{t-1}] = 0 \ \forall z_{t-1} \in sp(\{x_{t-i}\}_{i=0}^{\infty})$. If $p = 2$ then $Q_{t-1}y_t = E[y_t|\mathcal{F}_{t-1}]$. The fundamental hypotheses are

$$H_0 : P\left(Q\left(y_t|\mathcal{F}_{t-1}\right) = f(x_t, \phi_0)\right) = 1, \text{ for some } \phi_0 \in \Phi$$
$$H_1 : \sup_{\phi \in \Phi} P\left(Q\left(y_t|\mathcal{F}_{t-1}\right) = f(x_t, \phi_0)\right) < 1.$$

Under H_0 the function $f(x_t, \phi_0)$ represents the best L_p-predictor of y_t.

Write $F'(u) := (\partial/\partial u)F(u)$ and define the following class of weights:

$$\begin{aligned}\mathcal{H}_F \quad = \quad \{g : \mathbb{R}^k \to \mathbb{R} | g(x) = F(A(x)),\ A \text{ is affine},\ F : \mathbb{R} \to \mathbb{R},\\ F \text{ and } F' \text{ are analytic and non-polynomial}\\ \text{on some open interval } R_0 \subset \mathbb{R}\}.\end{aligned}$$

Assumption B Let $F \in \mathcal{H}_F$, and $(\partial/\partial u)^i F(u)|_{u=0} = c_i$ where $c_i = 0$ for only finitely many $i \in \mathbb{N}$. Let 0 lie in the interior of R_0.

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

Let $h : \mathbb{R}^k \times \Delta \to \mathbb{R}^k$ be a uniformly bounded, $\mathcal{F}_{t-1}$-measurable function, $k \geq 1$, where Δ is an arbitrary subset of $\mathbb{R}^l$ for some $l \geq 0$. Write $h_t(\delta) = h(x_t, \delta)$. 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 $E|e_t| < \infty$, and let x_t be an $\mathcal{F}_{t-1}$-measurable bounded vector in $\mathbb{R}^k$ such that $P(E[e_t|x_t] = 0) < 1$. Let Assumption B hold. For each $\delta \in \mathbb{R}^l$ the set

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

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

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

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

$$0 \in \Gamma$$

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

$$e_t = \epsilon_t^{<p-1>}$$

and define the moment

$$\eta(\gamma) := E[e_t F(\gamma' \psi(x_t))].$$

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

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

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

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

Define the vector weight function

$$(1) \quad H_t(\gamma, \gamma^{(*)}) := \left[F_t(\gamma),\ \psi_1(x_t) F_t'(\gamma^{(1)}),\ ...,\ \psi_k(x_t) F_t'(\gamma^{(k)}) \right]' \in \mathbb{R}^{k+1},$$

and define the set

$$S^{(*)} = \bigcap_{i=1}^{k} \{\gamma \in \Gamma : E[e_t H_{i,t}(\gamma, \gamma^{(*)})] = 0, \text{ and } P(\gamma' \psi(x_t) \in R_0) = 1\}.$$

Consider the moment

$$\varpi(\gamma, \gamma^{(*)}) := E\left[e_t \left(F_t(\gamma) - \sum_{i=1}^{k} \gamma_i \psi_i(x_t) F_t'(\gamma^{(i)}) \right) \right].$$

LEMMA 2 Let x_t be an $\Im_{t-1}$-measurable bounded vector in $\mathbb{R}^k$ such that $P(E[e_t|x_t] = 0) < 1$. Then $\varpi(\gamma, \gamma^{(*)}) = 0$ if and only if $\gamma = 0$.

The main result of the paper follows easily from Lemma 2.

THEOREM 3 Let x_t be an $\Im_{t-1}$-measurable bounded vector in $\mathbb{R}^k$ such that $P(E[e_t|x_t] = 0) < 1$. Then $S^* = \{0\}$ if and only if $e_t \perp \overline{sp}([\psi_i(x_t)F_t'(\gamma^{(i)})]_{i=1}^k)$, and $S^{(*)} = \{\perp\}$ otherwise.

Remark: The vector moment condition $E[e_t H_t(\gamma, \gamma^{(*)})]$ provides a two-way safety net against failing to detect model mis-specification. If a chosen $\gamma \neq 0$ implies failure of a moment condition characterized by Stinchcombe and White (1998),

$$E[e_t F_t(\gamma)] = 0,$$

then for at least one $i \in \{1, ..., k\}$ we are guaranteed a model mis-specification revealing moment

$$E\left[e_t \psi_i(x_t) F_t'(\gamma^{(i)})\right] \neq 0.$$

Conversely, because the vectors $\gamma^{(i)}$ maximize each gradient level $E[e_t \psi_i(x_t) F_t'(\gamma^{(i)})]$ and not the absolute magnitude $|E[e_t \psi_i(x_t) F_t'(\gamma^{(i)})]|$, it is possible that $E[e_t \psi_i(x_t) F_t'(\gamma^{(i)})] = 0$ for each $i = 1...k$ such that none reveal mis-specification. In such a case $E[e_t F(\gamma' x_t)] \neq 0$ is guaranteed to hold for all non-zero $\gamma \in \Gamma$.

EXAMPLE 1 Let $F(u) = \exp\{u\}$ and assume x_t is bounded. If $P[E(e_t|x_t) = 0] < 1$ then $\forall \gamma \neq 0$

$$\left[E[e_t \exp\{\gamma' x_t\}], E[e_t x_{1,t} \exp\{\gamma^{(1)\prime} x_t\}], ..., E[e_t x_{1,t} \exp\{\gamma^{(k)} x_t\}]\right]' \neq 0.$$

Stacking moment conditions popularized in the neural network and smooth transition threshold autoregression literatures generates a fail safe vector moment condition. Cf. Hornik, Stinchcombe and White (1989), Bierens (1991), and Teräsvirta (1994).

EXAMPLE 2 Let $e_t = \epsilon_t^{<p-1>}$, $1 < p \leq 2$. Suppose $\gamma^{(*)} = 0$ such that $F_t'(\gamma^{(i)}) = F_t'(0) = c_1$ by Assumption B. If $c_1 = 0$ then $0 \in S^*$. If $c_1 \neq 0$ then $0 \in S^*$ if and only if $E[\epsilon_t^{<p-1>} \psi(x_t)] = 0$. For example, if we use the weight $F(\gamma' x_t)$ to test linearity $f_t(\phi) = \phi' x_t$ then $E[\epsilon_t^{<p-1>} \psi(x_t) F_t'(\gamma^{(i)})] = E[\epsilon_t^{<p-1>} (\partial/\partial\phi) f_t(\phi)] = E[\epsilon_t^{<p-1>} x_t] = 0$ is automatically satisfied by L_p-orthogonality, hence $S^* = \{0\}$.

4. Super-Revealing CM's with Integer Nuisance Parameters

In practice computing $\gamma^{(i)} = \arg\sup_{\gamma \in \Gamma} \{(\partial/\partial\gamma_i) E[\epsilon_t F(\gamma' \psi(x_t))]\}$ and a test statistic functional over Γ may be computationally costly. Moreover, the subset Γ 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 $\psi : \mathbb{R}^k \to \mathbb{R}^k$ be any bounded, one-to-one function, and consider the weight

$$G_t(m) = \prod_{i=1}^k \psi_i(x_t)^{m_i}.$$

If $P[E(e_t|x_t) = 0] < 1$ then Theorem 3 of Hill (2006) guarantees

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

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

For example, assume x_t is bounded and use $\psi(x_t) = [\exp\{x_{1,t}\}, ..., \exp\{x_{k,t}\}]$. Then $G_t(m) = \prod_{i=1}^k \psi_i(x_t(\delta))^{m_i} = \exp\{m' x_t\}$ reveals model mis-specification for infinitely many m. If x_t is not bounded then simply substitute x_t for any bounded one-to-one function.

Now rewrite

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

$$\varpi(m, \gamma^{(*)}) = E\left[e_t\left(\exp\{m' x_t\} - \sum_{i=1}^k \gamma_i \psi_i(x_t) F_t'(\gamma^{(i)})\right)\right],$$

$$H_t(m, \gamma^{(*)}) = \left[\exp\{m' x_t\},\ \psi_1(x_t) F_t'(\gamma^{(1)}),\ ...,\ \psi_k(x_t) F_t'(\gamma^{(k)})\right]'$$

$$S_{(m)}^{(*)} = \bigcap_{i=1}^k \{m \in \mathbb{Z}^k : E[e_t H_{i,t}(m, \gamma^{(*)})] = 0\}.$$

The fact that Γ in Lemma 2 is arbitrary is advantageous here. The claim holds for all $\gamma \in \mathbb{R}^k$ and therefore for every $\gamma = m \in \mathbb{Z}^k$. That said, notice Lemma 2 exploits the fact that $E[e_t F_t \gamma)] \neq 0$ under H_1 for uncountably infinitely many γ in any arbitrary compact subset Γ, yet we are not guaranteed that $E\left[e_t G_t(m)\right] \neq 0$ for any m in any particular subset Γ. 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 $(\partial/\partial\gamma_i)\eta(\gamma^{(i)})$ is bounded we must assume $\gamma^{(i)}$ is finite.

Together with the fact that the weight $\exp\{m' x_t\}$ reveals mis-specification for infinitely many $m \in \mathbb{Z}^k \subset \mathbb{R}^k$, the following corollary to Lemma 2 is immediate.

COROLLARY 4 Assume $\gamma^{(i)} \in \Gamma$ for some compact subset $\Gamma \subset \mathbb{R}^k$. Let e_t be a random variable satisfying $E|e_t| < \infty$, and assume x_t is a bounded $\Im_{t-1}$-measurable k-vector. If $P[E(e_t|x_t) = 0] < 1$ then $\varpi(m, \gamma^{(*)}) = 0$ if and only if $m = 0$.

COROLLARY 5 Under the conditions of Lemma 4, if $P[E(e_t|x_t) = 0] < 1$ then $S_{(m)}^{(*)} = \{0\}$ if and only if $e_t \perp \overline{sp}[\psi_i(x_t) F_t'(\gamma^{(i)})]$, and $S_{(m)}^{(*)} = \{\emptyset\}$ otherwise.

5. Test Statistic

Write $\hat{\phi} = \arg\min_{\phi \in \Theta}\{\sum |y_t - f_t(\phi)|^p\}$, define $\hat{e}_t \equiv y_t - f_t(\hat{\phi})$, $\hat{e}_t \equiv \hat{e}_t^{<p-1>}$, and write $\partial f(\cdot) = (\partial/\partial\phi) f(\cdot)$. Define the sample conjugate to $\gamma^{(*)}$:

$$\hat{\gamma}^{(*)} = [\hat{\gamma}^{(i)}]_{i=1}^k = \left[\arg\inf_{\gamma \in \Gamma}\left\{1/n \sum_{t=1}^n \hat{e}_t \psi_i(x_t) F_t(\gamma)\right\}\right]_{i=1}^k.$$

In order to reduce notation write

$$\theta \equiv \{\gamma, \gamma^{(*)}\} \in \Theta \equiv \Gamma \times \Gamma^{(*)}, \quad \text{and} \quad \hat{\theta} = \{\gamma, \hat{\gamma}^{(*)}\}.$$

Define the sample vector moment

$$\hat{z}(\hat{\theta}) = 1/\sqrt{n} \sum_{t=1}^{n} \hat{\epsilon}_t^{<p-1>} H_t(\hat{\theta}) \in \mathbb{R}^{k+1},$$

where $H_t(\cdot)$ is defined in (1). We use a Pitman $\sqrt{n}$-local alternative of the form

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

where $E[\epsilon_t^{<p-1>}|\Im_{t-1}] = 0$. We assume u_t is measurable with respect to $\Im_{t-1}$ and governed by a non-generate distribution. The null hypothesis is $u_t = 0 \; a.s.$, and a global alternative is simply

$$H_1^G : y_t = f_t(\phi_0) + u_t + \epsilon_t.$$

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

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

say, where

$$\begin{aligned}
g_t(\theta) &= H_t(\theta) - b(\theta, \phi_0) A(\phi_0)^{-1} \partial f_t(\phi_0) \in \mathbb{R}^{k+1} \\
A(\phi_0) &= (p-1) \operatorname{plim}_{n\to\infty} (1/n) \sum_{t=1}^{n} |y_t - f_t(\phi_0)|^{p-2} \partial f_t(\phi_0) \partial' f_t(\phi_0) \\
b(\theta, \phi_0) &= (p-1) \operatorname{plim}_{n\to\infty} (1/n) \sum_{t=1}^{n} |y_t - f_t(\phi_0)|^{p-2} H_t(\theta) \partial' f_t(\phi_0)
\end{aligned}$$

Write

$$\begin{aligned}
\eta(\theta) &= (p-1) \operatorname{plim}_{n\to\infty} 1/n \sum_{t=1}^{n} |\epsilon_t|^{p-2} u_t g_t(\theta) \\
\Sigma(\theta) &= \operatorname{plim}_{n\to\infty} 1/n \sum_{t=1}^{n} |\epsilon_t|^{2(p-1)} g_t(\theta) g_t(\theta)'.
\end{aligned}$$

5.1 Weak Convergence

Weak convergence on a space of continuous real functions $C[\Theta]$ requires convergence of finite distributions and tightness of the vector sequence $\{z_n(\hat{\theta})\}$ on Θ. See Pollard (1984) and Billingsley (1999). In the integer nuisance parameter case (i.e. $\gamma = m \in \mathbb{Z}^k$) tightness is trivial. See Billingsley (1999) and Hill (2006). Consider, then, $\gamma \in \Gamma \subset \mathbb{R}^k$.

If a constant term is included then the matrix $\Sigma(\theta) = \Sigma(\gamma, \gamma^{(*)})$ may be close to singular if γ is near zero. If $\gamma = 0$ then $\Sigma(0, \gamma^{(*)}) \in \mathbb{R}^{k+1 \times k+1}$ will have rank k due to $\hat{z}(\hat{\theta}) = 1/\sqrt{n}\sum_{t=1}^{n}[0,\ \psi_1(x_t)F_t'(\gamma^{(1)}),\ \ldots,\ \psi_k(x_t)F_t'(\gamma^{(k)})]'$. We ameliorate the problem by bounding γ away from 0.

For arbitrary $\xi > 0$ define the subspace

$$\Theta_\xi \equiv \Gamma_\xi \times \Gamma^{(*)}, \quad \text{where} \quad \Gamma_\xi = \{\gamma \in \Gamma : |\gamma| > \xi\}.$$

Denote by $\lambda_{\min}(\theta)$ the minimum eigenvalue of $\Sigma(\theta)$.

Assumption C $\inf_{\theta \in \Theta_\xi} \lambda_{\min}(\theta) > 0$.

LEMMA 6 Let Assumption C hold. Let $z(\theta)$ denote a Gaussian element of $C[\Theta_\xi]$ with mean function $\Sigma(\theta)^{-1/2}\eta(\theta)$ and covariance function $E[z(\theta_1)z(\theta_2)']$ $= \Sigma(\theta_1)^{-1/2}\Sigma(\theta_1,\theta_2)\Sigma(\theta_2)^{-1/2}$. Under Assumption A and H_1^L, $\Sigma(\theta)^{-1/2}z_n(\theta)$ $\to z(\theta)$ pointwise in Θ_ξ. Moreover $|\Sigma(\theta)^{-1/2}z_n(\theta)/\sqrt{n}| \to \infty$ with probability one under H_1^G for every $\theta \in \Theta_\xi$.

LEMMA 7 Under Assumptions A-C and H_1^L the sequence $\{\Sigma(\theta)^{-1/2}(z_n(\theta) - \eta(\theta))\}$ is tight on Θ_ξ.

Define

$$\hat{\Sigma}(\hat{\theta}) = 1/n\sum_{t=1}^{n}|\hat{\epsilon}_t|^{2(p-1)}\hat{g}_t(\hat{\theta})\hat{g}_t(\hat{\theta})'$$
$$\hat{g}_t(\hat{\theta}) = H_t(\hat{\theta}) - \hat{b}(\hat{\theta},\hat{\phi})\hat{A}(\hat{\phi})^{-1}\partial f_t(\hat{\phi})$$
$$\hat{A}(\hat{\phi}) = (p-1)(1/n)\sum_{t=1}^{n}|\hat{\epsilon}_t|^{p-2}\partial f_t(\hat{\phi})\partial' f_t(\hat{\phi})$$
$$\hat{b}(\hat{\theta},\hat{\phi}) = (p-1)(1/n)\sum_{t=1}^{n}|\hat{\epsilon}_t|^{p-2}H_t(\hat{\theta})\partial' f_t(\hat{\phi}).$$

The following is an immediate consequence of Assumption A, (2) and Lemmas 6 and 7.

THEOREM 8 Under Assumptions A-C and H_1^L, $\hat{\Sigma}(\hat{\theta})^{-1/2}\hat{z}(\hat{\theta}) \Rightarrow z(\theta)$ on $C[\Theta_\xi]$ where $z(\theta)$ is defined in Lemma 6.

5.3 Super-Consistent Test Statistic

Consider a standard Lagrange multiplier test statistic

$$T_n(\hat{\theta}) = T_n(\gamma, \hat{\gamma}^{(*)}) = \hat{z}(\gamma, \hat{\gamma}^{(*)})'\hat{\Sigma}(\gamma, \hat{\gamma}^{(*)})^{-1}\hat{z}(\gamma, \hat{\gamma}^{(*)}).$$

Theorem 8 and the continuous mapping theorem guarantee $T_n(\gamma, \hat{\gamma}^{(*)}) \to \chi^2(k+1)$ under H_0. Under H_1^G the test statistic $T_n(\gamma, \hat{\gamma}^{(*)})$ reveals model misspecification asymptotically with probability one for any nuisance parameter $\gamma \neq 0$: $T_n(\gamma, \hat{\gamma}^{(*)})/n \to \infty$ $a.s.$

In practice, however, the analyst may want to improve small sample power by considering continuous, $\sigma(\Im_t)$-measurable functions $h : \mathbb{R}_+ \to \mathbb{R}_+$, 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 γ is considered below.

6. Monte Carlo Study

Write $\boldsymbol{x}_{j,t-1} = [x_{t-1}, ..., x_{t-p_j}]'$. We simulate 100 of the following processes

$$
\begin{aligned}
&H_0 : x_t = \phi_1'\boldsymbol{x}_{1,t-1} + \epsilon_t \\
&H_1^S : x_t = \phi_1'\boldsymbol{x}_{1,t-1} + \phi_2'\boldsymbol{x}_{2,t-1} \times I(x_{t-1} > 0) + \epsilon_t \\
&H_1^{ES} : x_t = \phi_1'\boldsymbol{x}_{1,t-1} + \phi_2'\boldsymbol{x}_{2,t-1} \times \exp\{-\gamma x_{t-1}^2\} + \epsilon_t \\
&H_1^{LS} : x_t = \phi_1'\boldsymbol{x}_{1,t-1} + \phi_2'\boldsymbol{x}_{2,t-1} \times [1 - \exp\{-\gamma x_{t-1}\}]^{-1} + \epsilon_t \\
&H_1^{AN} : x_t = \phi_1'\boldsymbol{x}_{1,t-1} + [1 - \exp\{-\gamma x_{t-1}\}]^{-1} + \epsilon_t \\
&H_1^{ALL} : \text{ randomized } H_1,
\end{aligned}
$$

where $\epsilon_t \overset{iid}{\sim} N(0,1)$. In all cases each p_j is randomly selected from the set $\{1, ..., 10\}$, and each $\phi_j \in \mathbb{R}^{p_j}$ is randomly selected from $[-.9, .9]^{p_j}$ subject to all roots lying outside the unit circle; β is randomly selected from $[-.9, .9]$; and γ is randomly selected from $[.5, 10]$. Under H_0 the process is AR(p); under H_1^S the process is SETAR(p,p); under H_1^{ES} and H_1^{LS} the process is ESTAR and LSTAR respectively; under H_1^{AN} it is an AR(p)-ANN (artificial neural network); under H_1^{ALL} the alternative is selected at random from those just described. Sample sizes are $n \in \{200, 500\}$.

We estimate a null model by fitting an AR(p) to the series $\{X_t\}_{t=1}^n$, where p is selected by minimizing the AIC. We test the residuals for omitted nonlinearity at the 5%-level using $\sup_{\gamma \in \Gamma_n^p} T_n(\gamma, \hat{\gamma}^{(*)})$ and $\sup_{m \in N_n^p} T_n(m, \hat{\gamma}^{(*)})$, and randomized tests $T_n(\gamma, \hat{\gamma}^{(*)})$ and $T_n(m, \hat{\gamma}^{(*)})$ where γ and m are randomly selected from subsets Γ_n^p and N_n^p defined below. In simulation experiments not reported here average test functionals are uniformly dominated by supremum functionals simply because the alternatives are not local, or "close" to the null. Cf. Andrews and Ploberger (1994).

For $\sup_{\gamma \in \Gamma_n^p} T_n(\gamma, \hat{\gamma}^{(*)})$ we use the sample moment conditions $\hat{z}(\gamma, \hat{\gamma}^{(*)}) = 1/\sqrt{n}\sum_{t=1}^n \hat{\epsilon}_t \hat{\sigma}_p^{-1} H_t(\gamma, \hat{\gamma}^{(*)})$ where $H_t(\gamma, \hat{\gamma}^{(*)})$ is one of the following:

$$
\left[\exp\{\gamma' x_t\},\ \left[x_{t-i}\exp\{\gamma^{(i)\prime} x_t\}\right]_{i=1}^{p}\right]'
$$

$$
\left[(1 + \exp\{\gamma' x_t\})^{-1},\ \left(x_{t-i}\exp\{\gamma^{(i)\prime} x_t\}\left(1 + \exp\{\gamma^{(i)\prime} x_t\}\right)^{-1}\right)_{i=1}^{p}\right]'.
$$

6.1 Nuisance Parameter Spaces

The set $\Gamma_n^p = \{\gamma_1, ..., \gamma_{J_n}\}$ is constructed from J_n randomly selected $\gamma_i \in [.5, 10]^p$, $J_n = [n/4]$. For $\sup_{m \in N_n^p} T_n(m, \hat{\gamma}^{(*)})$ we use only the exponential and construct the set of integers N_n^p as follows. Denote by $j_i^{(k)}$ a p-vector with the value j for the i^{th} component and the value k in all other components. Let $\bar{N}_n^p$ be a set with $[\sqrt{\ln n}]$ integer vectors randomly selected from $\{[0, ..., 0]', ..., [[\sqrt{\ln n}], ..., [\sqrt{\ln n}]]'\}$. Let N_n^p be the set of all integers in the hypercube $\{[0, ..., 0]', ..., [[(\ln n)^{1/8}], ..., [(\ln n)^{1/8}]]'\}$. Finally, let $\underline{N}_n^p$ denote the set of all integers $\{\{i_j^{(0)}\}_{i=1}^{[\sqrt{\ln n}]}\}_{j=1}^{p}$ and $\{1_1^{(1)}, 2_1^{(2)}, ..., [\sqrt{\ln n}]_1^{([\sqrt{\ln n}])}\}$. We use $N_n^p =$

$\bar{N}_n^p \cup N_n^p \cup N_n^p$, which contains simple integer vectors, randomized vectors, and $N_n^p \to \mathbb{N}^p$ as $n \to \infty$.

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 ϕ_1 and each p_i are known within the simulation, each model can be written as $y_t(\phi_1) = \phi_2' z_t(\gamma) + \epsilon_t$, where $y_t(\phi_1) = x_t - \phi_1' x_{t-1}$, $z_t(\gamma) = x_{2,t-1} \exp\{\gamma' x_{2,t-1}\}$ under H_1^E, $z_t(\gamma) = x_{2,t-1} I(x_{t-1} > 0)$ under H_1^S, and $z_t(\gamma) = [1 + \exp\{\gamma' x_{2,t-1}\}]^{-1}$ under H_1^{AN}. The errors are known to be iid standard normal, hence the UMP test statistic reduces to $W_n(\gamma) = y(\phi_1)' z(\gamma) \left[z(\gamma)' z(\gamma)\right]^{-1} z(\gamma)' y(\phi_1)$.

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 $T_n(\gamma, \hat{\gamma}^{(*)})$ 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 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 differentiable on Φ. u_t and $f_t(\phi)$ are $\mathcal{F}_{t-1}$-measurable, where $\mathcal{F}_t$ is the sequence of σ-algebras generated by $(x_\tau : \tau \le t+1)$. Moreover, $E[\epsilon_t^{<p-1>}|\mathcal{F}_{t-1}] = 0$ $a.s.$ for some $p \in (\min\{1, 1 + .25 \times \arg\sup_{\alpha>0}\{E|\epsilon_t|^\alpha < \infty\}, 2)$: hence $||\epsilon_t^{<p-1>}||_{4+\delta} < \infty$ for some $\delta > 0$.

Assumption A2:
$i.$ 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. For some stochastic sequence $\{u_t^*\}$ satisfying $u_t^* \in [0, u_t]$ and $u_t^* = o_p(\sqrt{n})$ the L_p-estimator $\hat\phi = \arg\min_{\phi\in\Phi}\sum_{t=1}^n |y_t - f_t(\phi)|^p$ satisfies

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

$ii.$ $\inf_{\gamma\in\Gamma}\inf_{r'r=1} r'E[e_t^2\left([\Psi_i(x_t)F_t'(\gamma)]_{i=1}^k\right) \times \left([\Psi_i(x_t)F_t'(\gamma)]_{i=1}^k\right)']r > 0$.
$iii.$ There exists some $\gamma_0^{(*)} \in \Gamma^{(*)}$ such that $|\hat\gamma^{(*)} - \gamma_0^{(*)}| = O_p(1/\sqrt{n})$ and

$$\sup_{\gamma\in\Gamma}\left|H_t(\gamma,\hat\gamma^{(*)}) - H_t(\gamma,\gamma_0^{(*)})\right| = O_p(1/\sqrt{n})$$
$$\sup_{\gamma\in\Gamma,\phi\in\Phi}\left|b(\gamma,\hat\gamma^{(*)},\phi) - b(\gamma,\gamma_0^{(*)},\phi)\right| = O_p(1/\sqrt{n})$$
$$\sup_{\gamma\in\Gamma,\phi\in\Phi}\left|\Sigma(\gamma,\hat\gamma^{(*)}) - \Sigma(\gamma,\gamma_0^{(*)})\right| = O_p(1/\sqrt{n}).$$

Assumption A3: For each $i = 1...k$ let $\hat{b}(\gamma,\boldsymbol{\gamma},\phi_0) = (p-1)(1/n)\sum_{t=1}^n |y_t - f_t(\phi_0)|^{p-2} \times H_t(\gamma,\boldsymbol{\gamma})\partial' f_t(\phi_0)$, where $\gamma \in \Gamma$ and $\boldsymbol{\gamma} \in \Gamma\times\cdots\times\Gamma \subset \mathbb{R}^{k\times k}$. Then $\sup_{\phi\in\Phi,\gamma\in\Gamma,\boldsymbol{\gamma}\in\bar\Gamma}|\hat{b}(\gamma,\boldsymbol{\gamma},\phi) - b(\gamma,\boldsymbol{\gamma},\phi)| \to 0$ where $b(\gamma,\gamma^{(i)},\phi)$ is a non-stochastic function satisfying $\sup_{\phi\in\Phi,\gamma\in\Gamma,\boldsymbol{\gamma}\in\bar\Gamma}|b(\gamma,\boldsymbol{\gamma},\phi)|_2 < C$, $\sup_{\phi\in\Phi,\gamma\in\Gamma,\boldsymbol{\gamma}\in\bar\Gamma}|(\partial/\partial\gamma)b(\gamma,\boldsymbol{\gamma},\phi)|_2 < C$ and $\sup_{\phi\in\Phi,\gamma\in\Gamma,\boldsymbol{\gamma}\in\bar\Gamma}|(\partial/\partial\boldsymbol{\gamma}_l)b(\gamma,\boldsymbol{\gamma},\phi)|_2 < C$, $l = 1...k$.

Assumption A4: Write $\theta = \{\gamma,\gamma^{(*)}\}$ and $\theta_i = \{\gamma_i,\gamma^{(*)}\}$.
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 positive definite matrix.
ii. There exists a mapping $\eta : \Gamma\times\Gamma^{(*)} \to \mathbb{R}$ satisfying $(p-1)\times(1/n)\sum_{t=1}^n |\epsilon_t|^{p-2}u_t g_t(\gamma,\gamma^{(*)}) \to (p-1)\times\lim_{n\to\infty}(1/n)\sum_{t=1}^n E[|\epsilon_t|^{p-2}u_t g_t(\gamma,\gamma^{(*)})] = \eta(\gamma,\gamma^{(*)})$.
iii. Write $\theta \in \Theta = \Gamma\times\Gamma^{(*)}$. There exists a functional $\Sigma : \Theta\times\Theta \to \mathbb{R}^{k\times k}$ satisfying $(1/n)\sum_{t=1}^n E[|\epsilon_t|^{2(p-1)}|\mathcal{F}_{t-1}]\times g_t(\theta_1)g_t(\theta_2)' \to \Sigma(\theta_1,\theta_2)$, $(1/n)\sum_{t=1}^n |\epsilon_t|^{2(p-1)} \times g_t(\theta_1)g_t(\theta_2)' \to \Sigma(\theta_1,\theta_2)$ and $(1/n)\sum_{t=1}^n E[|\epsilon_t|^{2(p-1)} \times g_t(\theta_1)g_t(\theta_2)'] \to \Sigma(\theta_1,\theta_2)$ pointwise on $\Gamma\times\Gamma^{(*)}$.
iv. For some $\delta > 0$, $\limsup_{n\to\infty}\sup_{\{\gamma,\gamma^{(*)}\}\in\Gamma} 1/n\sum_{t=1}^n E||\epsilon_t|^{p-2}u_t g_t(\gamma,\gamma^{(*)})|^{2+\delta} < \infty$.

Assumption A.5: For some $\delta > 0$, $||\sup_{\theta\in\Theta}|g_t(\theta)|||_{4+\delta} < C$ and $||\sup_{\theta\in\Theta}|(\partial/\partial\theta)g_t(\theta)|||_{4+\delta} < C$.

Appendix 2: Formal Proofs

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

$$E[e_t \psi_i(x_t)F_t'(\gamma)] \le E[e_t \psi_i(x_t)F_t'(\gamma^{(i)})] = \sup_{\gamma \in \Gamma} E[e_t \psi_i(x_t)F_t'(\gamma)].$$

Assumptions A and B imply

$$\varpi(0, \gamma^{(*)}) = E[e_t F_t(0)] = 0.$$

Now differentiate $\varpi(\gamma, \gamma^{(*)})$ with respect to each γ_j, and add and subtract $E[e_t \psi_j(x_t)F_t'(0)]$. By construction

$$(3) \qquad \left(\partial/\partial\gamma_j\right) \varpi(\gamma, \gamma^{(*)})$$
$$= E\left[e_t \psi_j(x_t)\{F_t'(\gamma) - F_t'(0)\}\right] - E[e_t \psi_j(x_t)\{F_t'(\gamma^{(j)}) - F_t'(0)\}] \le 0.$$

Thus $\varpi(\gamma, \gamma^{(*)})$ is zero at $\gamma = 0$ and is weakly decreasing in γ. From (3) we know $E[e_t \psi_j(x_t)\{F_t'(\gamma^{(j)}) - F_t'(0)\}] \ge 0 \ \forall j = 1...k$.

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

Case 1: $\quad E[e_t \psi_j(x_t)\{F_t'(\gamma^{(j)}) - F_t'(0)\}] = 0.$

Trivially

$$E[e_t \psi_j(x_t)\{F_t'(\gamma) - F_t'(0)\}]|_{\gamma=0} = 0,$$

Lemma 1 therefore implies there exists an open neighborhood $N_0 \subset \Gamma$ of zero satisfying

$$E[e_t \psi_j(x_t)\{F_t'(\gamma) - F_t'(0)\}] \ne 0 \quad \forall\gamma \in N_0/0.$$

By assumption $E[e_t \psi_j(x_t)\{F_t'(\gamma^{(j)}) - F_t'(0)\}] = 0$ hence from (4) we deduce

$$E[e_t \psi_j(x_t)\{F_t'(\gamma) - F_t'(0)\}] < 0 \quad \forall\gamma \in N_0/0.$$

Thus $\varpi(\gamma, \gamma^{(*)})$ is zero at $\gamma = 0$, strictly decreasing arbitrarily close to $\gamma = 0$, and weakly decreasing everywhere else. Thus $\varpi(\gamma, \gamma^{(*)}) \ne 0 \ \forall\gamma \ne 0$.

Case 2: $\quad E[\epsilon_t \psi_j(x_t)\{F_t'(\gamma^{(j)}) - F_t'(0)\}] > 0.$

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

$$\left(\partial/\partial\gamma_j\right) \varpi(0, \gamma^{(*)}) = 0 - E\left[\epsilon_t \psi_j(x_t)\left\{F_t'(\gamma^{(j)}) - F_t'(0)\right\}\right] < 0.$$

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

Proof of Theorem 3. Assume $P[E(e_t|x_t) = 0] < 1$ and recall x_t does not contain a constant term. By Lemma 2 we know $\forall\gamma \ne 0$

$$\varpi(\gamma, \gamma^{(*)}) = E[e_t F_t(\gamma)] - \sum_{i=1}^{k} \gamma_i E[e_t \psi_i(x_t)F_t'(\gamma^{(i)})] \ne 0.$$

Trivially, therefore, at least one moment condition $E[e_tF_t(\gamma)]$, $E\left[e_t\Psi_1(x_t)F_t'(\gamma^{(1)})\right]$, ..., or $E\left[e_t\Psi_k(x_t)F_t'(\gamma^{(k)})\right]$ must be non-zero, hence $E[e_tH_t(\gamma,\gamma^{(*)})] \neq 0$ for every $\gamma \neq 0$.

Finally, under Assumptions A and B

$$E[e_tH_t(0,\gamma^{(*)})] = \left[0, E\left[e_t\Psi_1(x_t)F_t'(\gamma^{(1)})\right],\dots,E\left[e_t\Psi_k(x_t)F_t'(\gamma^{(k)})\right]\right]',$$

hence $E[e_tH_t(0,\gamma^{(*)})] = 0$ if and only if e_t is orthogonal to $\overline{sp}(\{\Psi_i(x_t) \times F_t'(\gamma^{(i)})\}_{i=1}^k)$. ■

Proof of Lemma 6. Invoking a Cramér-Wold device and Assumption A, under H_1^L it suffices to prove for any $r \in \mathbb{R}^{k+1}$, $r'r = 1$,

$$1/\sqrt{n}\sum_{t=1}^n \epsilon_t^{<p-1>} r'\Sigma(\theta)^{-1/2}g_t(\theta) \to N(0,I_k).$$

Clearly $\{\epsilon_t^{<p-1>} r'g_t(\theta), \mathcal{F}_{t-1}\}$ forms a martingale difference sequence due $E[\epsilon_t^{<p-1>}|\mathcal{F}_{t-1}] = 0$ under Assumption A. The claim under H_1^L now follows from Assumption A.4.iii, Lemma A.1, below, and Bierens' (1994:Lemma 6.1.7) generalization of McLeish's (1974) martingale difference central limit theorem.

Write $w_t(\theta) = (y_t - f_t(\phi))^{<p-1>}g_t(\theta)$. Under H_1^G and Lemma 1 we deduce $E[w_t(\theta)|\mathcal{F}_{t-1}] \neq 0\ \forall\theta \in \Theta_\xi$. Therefore, we need only show (2) implies

$$z_n(\theta)/\sqrt{n} = 1/n\sum_{t=1}^n (w_t(\theta) - E[w_t(\theta)|\mathcal{F}_{t-1}]) \to 0,$$

in probability. The limit follows from Assumption A and a martingale difference law of large numbers due to Chow (1971). Cf. Corollary 19.8 of Davidson (1994).

■

Proof of Lemma 7. For any $r \in \mathbb{R}^{k+1}$, $r'r = 1$, write $e_t = \epsilon_t^{<p-1>}$ and

$$\begin{aligned} 1/\sqrt{n}\sum_{t=1}^n e_t r'\Sigma(\theta)^{-1/2}g_t(\theta) &= 1/\sqrt{n}\sum_{t=1}^n e_t r'\psi_t(\theta) \\ &= 1/\sqrt{n}\sum_{t=1}^n e_t w_t(r,\theta), \end{aligned}$$

say, where $\psi_t(\theta) = \Sigma(\theta)^{-1/2}g_t(\theta)$ and $w_t(r,\theta) = r'\psi_t(\theta)$. Using Lemma A.1 of Bierens and Ploberger (1997) we need to show

$$\text{(4)} \qquad \begin{aligned} &\limsup_{n\to\infty} n^{-1}\sum_{t=1}^n E[e_t^2K_t^2] < \infty \\ &\sup_{r'r=1}\limsup_{n\to\infty} n^{-1}\sum_{t=1}^n E[e_t^2w_t(r,\theta_0)^2] < \infty, \end{aligned}$$

for at least one point $\theta_0 \in \Theta$, where it suffices to use

$$K_t = \sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta)w_t(r,\theta)|.$$

The second inequality in (4) follows from Assumption A and $r'r = 1$:

$$\sup_{r'r=1} E[e_t^2w_t(r,\theta)^2] \le \|e_t\|_4^2\, |\Sigma(\theta)^{-1/2}|^2 \|\sup_{\theta\in\Theta_\xi}|g_t(\theta)|\|_4^2 \le C,$$

where $|\Sigma(\theta)^{-1/2}| < \infty$ is an easy consequence of Assumptions A and C.

For the first inequality in (4) we will prove $E[e_t^2 K_t{}^2] \le M$ for some positive constant $M < \infty$. By the Cauchy-Schwartz inequality

$$\sup_{r'r=1} E[e_t^2 K_t{}^2] \le \|\epsilon_t\|_4^2 \times \|\sup_{\theta \in \Theta_\xi} |(\partial/\partial\theta)\psi_t(\theta)|\|_4^2.$$

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

$$\begin{aligned}(\partial/\partial\theta_l)\psi_{t,j}(\theta) &= \sum_{i=1}^{k+1} (\partial/\partial\theta_l)\Sigma(\theta)_{j,i}^{-1/2} \times g_{t,i}(\theta) \\ &\quad + \sum_{i=1}^{k+1} \Sigma(\theta)_{j,i}^{-1/2} \times (\partial/\partial\theta_l) g_{t,i}(\theta).\end{aligned}$$

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

$$\begin{aligned}
&\left\|\sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta)\psi_t(\theta)|\right\|_4 \\
&\quad \le \left\| \sum_{l=1}^{k(k+1)} \sum_{j=1}^{k+1} \sup_{\theta\in\Theta_\xi} \left| \sum_{i=1}^{k+1} (\partial/\partial\theta_l)\Sigma(\theta)_{j,i}^{-1/2} \times g_{t,i}(\theta) \right| \right\|_4 \\
&\qquad + \left\| \sum_{l=1}^{k(k+1)} \sum_{j=1}^{k+1} \sup_{\theta\in\Theta_\xi} \left| \sum_{i=1}^{k+1} \Sigma(\theta)_{j,i}^{-1/2} \times (\partial/\partial\theta_l) g_{t,i}(\theta) \right| \right\|_4 \\
&\quad \le k \left\| \sum_{l=1}^{k(k+1)} \sup_{\theta\in\Theta_\xi} \left| \sum_{i,j=1}^{k+1} \left|(\partial/\partial\theta_l)\Sigma(\theta)_{j,i}^{-1/2}\right| \sum_{i=1}^{k+1} |g_{t,i}(\theta)| \right| \right\|_4 \\
&\qquad + k(k+1)^2 \left\| \sup_{\theta\in\Theta_\xi} \left| \sum_{i,j=1}^{k+1} \left|\Sigma(\theta)_{j,i}^{-1/2}\right| \sum_{l=1}^{k(k+1)} \sum_{i=1}^{k+1} |(\partial/\partial\theta_l) g_{t,i}(\theta)| \right| \right\|_4 \\
&\quad \le k \sum_{l=1}^{k(k+1)} \sup_{\theta\in\Theta_\xi} \left|(\partial/\partial\theta_l)\Sigma(\theta)^{-1/2}\right| \left\| \sup_{\theta\in\Theta_\xi} |g_t(\theta)| \right\|_4 \\
&\qquad + k(k+1)^2 \sup_{\theta\in\Theta_\xi} \left|\Sigma(\theta)^{-1/2}\right| \left\| \sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta) g_t(\theta)| \right\|_4 \le C.
\end{aligned}$$

∎

Appendix 3: Supporting Lemmata

LEMMA A.1 Under Assumption A for each $\theta \in \Theta_\xi$ and every $r \in \mathbb{R}^{k+1}$, $r'r = 1$

$$(5)\quad \operatorname{plim}_{n\to\infty} 1/n \sum_{t=1}^{n} \left(\epsilon_t^{<p-1>} r'\Sigma(\theta)^{-1/2} g_t(\theta)\right)^2 = \lim_{n\to\infty} 1/n \sum_{t=1}^{n} E\left[\left(\epsilon_t^{<p-1>} r'\Sigma(\theta)^{-1/2} g_t(\theta)\right)^2\right] = 1,$$

and for some $\kappa > 0$

$$(6)\quad \operatorname*{plim}_{n\to\infty} \sum_{t=1}^{n} E\left|\epsilon_t^{<p-1>} r'\Sigma(\theta)^{-1/2} g_t(\theta)/\sqrt{n}\right|^{2+\kappa} = 0.$$

LEMMA A.2 Assumption A implies:

i. $\sup_{\theta\in\Theta_\xi} |V(\theta)^{-1/2}|^2 \le C(k+1)\left[\inf_{\theta\in\Theta_\xi} \lambda_{\min}(V(\theta))\right]^{-1} \le C$;

ii. $\sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta_l)V(\theta)^{-1/2}| \le C$, $l = 1 \ldots k+1$.

Proof of Lemma A.1. From the normalization $\pi'\pi = 1$ and Assumption A it is easy to show $E[(\epsilon_t^{<p-1>}\pi'\Sigma(\theta)^{-1/2}g_t(\theta))^2] = 1$ for all $t \in \mathbb{N}$. Limit (5) then follows from Assumption A.4:

$$\sup_{\theta\in\Theta}\left|1/n \sum_{t=1}^{n} |\epsilon_t|^{2(p-1)} g_t(\theta)g_t(\theta)' - \Sigma(\theta)\right| = o_p(1),$$

Limit (6) follows from the following bound. By l_1-norm properties, the envelope inequality, and Assumption A, for some small $\kappa > 0$ and some finite $M > 0$

$$\begin{aligned} &E|w_t(\theta)|^{2+\kappa} \\ &\quad \le |r|^{2+\kappa}|\Sigma(\theta)^{-1/2}|^{2+\kappa}E|g_t(\theta)|^{2+\kappa} \\ &\quad \le |r|^{2+\kappa}\sup_{\theta\in\Theta_\xi}|\Sigma(\theta)^{-1/2}|^{2+\kappa}\|\sup_{\theta\in\Theta_\xi}|g_t(\theta)|\|_{2+\kappa}^{2+\kappa} \le M, \end{aligned}$$

where $|r|^{2+\kappa} < \infty$ is trivial. Thus, $\sum_{t=1}^{r} E|w_t(\theta)|^{2+\kappa}/n^{1+\kappa/2} = o(1/n^{\kappa/2})$. ■

Proof of Lemma A.2.

i. Liaponov's inequality implies for some finite $B > 0$:

$$\begin{aligned} |\Sigma(\theta)^{-1/2}|^2 &\le B|\Sigma(\theta)^{-1/2}|_2^2 = B \times Tr\left(\Sigma(\theta)^{-1/2\prime}\Sigma(\theta)^{-1/2}\right) \\ &= B \times Tr\left[\Sigma(\theta)^{-1}\right] = B \times \sum_{i=0}^{pk} \lambda_i(\Sigma(\theta)^{-1}) \\ &= B \times \sum_{i=0}^{pk} 1/\lambda_i(\Sigma(\theta)) \\ &\le B \times (k+1)\,\lambda_{\min}(\Sigma(\theta))^{-1} \end{aligned}$$

hence

$$\sup_{\theta\in\Theta_\xi} |\Sigma(\theta)^{-1/2}|^2 \leq B \times (k+1)\left[\inf_{\theta\in\Theta_\xi} \lambda_{\min}(\Sigma(\theta))\right]^{-1} \leq C,$$

which is guaranteed for some finite C by Assumption C: $\inf_{\theta\in\Theta_\xi} \lambda_{\min}(\Sigma(\theta)) > 0$.

ii. By standard properties of matrix differentiation

$$(\partial/\partial\theta_l)\Sigma(\theta)^{-1/2} = -(1/2)[\Sigma(\theta)^{-1/2} \times (\partial/\partial\theta_l)\Sigma(\theta) \times \Sigma(\theta)^{-1}].$$

Hence, for some finite $B > 0$, by Liaponov's inequality and (i),

$$\begin{aligned}
&\sup_{\theta\in\Theta_\xi} \left|(\partial/\partial\theta_l)\Sigma(\theta)^{-1/2}\right| \\
&\quad\leq \sup_{\theta\in\Theta_\xi} \left|\Sigma(\theta)^{-1/2} \times (\partial/\partial\theta_l)\Sigma(\theta) \times \Sigma(\theta)^{-1}\right| \\
&\quad\leq \sup_{\theta\in\Theta_\xi} |\Sigma(\theta)^{-1/2}|^3 |(\partial/\partial\theta_l)\Sigma(\theta)| \\
&\quad\leq B(k+1)^{3/2}\left[\inf_{\theta\in\Theta_\xi} \lambda_{\min}(\Sigma(\theta))\right]^{-3/2} \sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta_l)\Sigma(\theta)|
\end{aligned}$$

where $\inf_{\theta\in\Theta_\xi} \lambda_{\min}(\Sigma(\theta)) > 0$ by Assumption C. The proof is complete when we show the l_1-normed $|(\partial/\partial\theta_l)\Sigma(\theta)|$ is uniformly bounded by some finite $M > 0$.

The covariance matrix derivative $(\partial/\partial\theta_l)\Sigma(\theta)$ is computed as

$$\begin{aligned}
(\partial/\partial\theta_l)\Sigma(\theta) &= (\partial/\partial\theta_l)E\left[\epsilon_t^2 g_t(\theta) g_t(\theta)'\right] \\
&= \left(E\left[\epsilon_t^2 (\partial/\partial\theta_l) g_{t,i}(\theta) g_{t,j}(\theta)\right]\right)_{i,j} \\
&\qquad + \left(E\left[\epsilon_t^2 g_{t,i}(\theta)(\partial/\partial\theta_l) g_{t,j}(\theta)\right]\right)_{i,j}.
\end{aligned}$$

By the envelope and repeated Cauchy-Schwartz inequalities,

$$\begin{aligned}
&\sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta_l)\Sigma(\theta)| \\
&\quad\leq 2\sum_{i,j=1}^{k+1} \sup_{\theta\in\Theta_\xi} \left|E\left[\epsilon_t^2 (\partial/\partial\theta_l) g_{t,i}(\theta) g_{t,j}(\theta)\right]\right| \\
&\quad\leq 2E\left[\epsilon_t^2 \sum_{i=1}^{k+1} \sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta_l) g_{t,i}(\theta)| \sum_{i=1}^{k+1} \sup_{\theta\in\Theta_\xi} |g_{t,i}(\theta)|\right] \\
&\quad\leq 2\|\epsilon_t\|_4^2 \left\|\sum_{i=1}^{k+1} \sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta_l) g_{t,i}(\theta)|\right\|_4 \left\|\sum_{i=1}^{k+1} \sup_{\theta\in\Theta_\xi} |g_{t,i}(\theta)|\right\|_4 \\
&\quad= 2\|\epsilon_t\|_4^2 \left\|\sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta_l) g_t(\theta)|\right\|_4 \left\|\sup_{\theta\in\Theta_\xi} |g_t(\theta)|\right\|_4 \leq C.
\end{aligned}$$

■

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] Bierens, H. J. (1982). "Consistent Model Speci…cation Tests," Journal of Econometric, 20, 105-134

[3] Bierens, H. J. (1984). "Model Speci…cation Testing of Time Series Regressions," Journal of Econometrics, 26, 323-353.

[4] Bierens, H. J. (1990). "A Consistent Conditional Moment Test of Functional Form," Econometrica, 58, 1443-1458.

[5] Bierens, H. J. (1994). "Topics in Advanced Econometrics: Estimation, Testing, and Speci…cation of Cross-Section and Time Series Models," Cambridge Univ. Press: Cambridge.

[6] Bierens, H. J. and W. Ploberger (1997). "Asymptotic Theory of Integrated Conditional Moment Tests," Econometrica, 65, 1129-1151.

[7] Billingsley, P. (1999). "Convergence of Probability Measures," John Wiley & Sons: New York.

[8] Chow, Y. S. (1971). On the L_p Convergence for $n^{-1/p}S_n$, $0 < p < 2$, Annals of Mathematical Statistics 36, 393-394.

[9] Davies, R.B. (1977). "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," Biometrika, 64, 247-254.

[10] de Jong, R. (1996). "The Bierens Test under Data Dependence,"Journal of Econometrics, 72, 1-32.

[11] Dette, H. (1999). "A Consistent Test for the Functional Form of a Regression Based on a Di¤erence of Variance Estimators," Annals of Statistics, 27, 1012-40.

[12] Gallant, A.R, and H. White (1988). "A Uni…ed of Estimation and Inference for Nonlinear Dynamic Models," Basil Blackwell: New York.

[13] Hansen, B. (1996). "Inference When a Nuisance Parameter Is Not Identi…ed Under the Null Hypothesis," Econometrica, 64, 413-430.

[14] Hill, J. B. (2006). "Asymptotically Nuisance-Parameter-Free Consistent Tests of L_p-Functional Form", Working Paper 06-08, Dept. of Economics, Florida International University, available at http://www.…u.edu/orgs/economics/wp2006/06-08.pdf.

[15] Holley, A. (1982). "A Remark on Hausman's Speci…cation Test," Econometrica, 50, 749-760.

[16] Hornik, K., M. Stinchcombe and H. White (1989). "Multilayer Feedforward Networks are Universal Approximators," Neural Networks, 2, 359-366.

[17] Hong, Y. and H. White (1995). Consistent Speci…cation Testing via Nonparametric Series Regression," Econometrica, 63, 1133-1159.

[18] King, M.L., and T.S. Shively (1993). "Locally Optimal Testing when a Nuisance Parameter is Present Only Under the Alternative," Review and Economics and Statistics 75, 1-7.

[19] McLeish, D. (1974). Dependent Central Limit Theorems and Invariance Principles, Annals of Probability 2, 620-628.

[20] Newey, W.K. (1985)."Maximum Likelihood Speci…cation Testing and Conditional Moment Tests," Econometrica, 53, 1047-1070.

[21] Stinchcombe, M.B. and H. White (1998). "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," Econometric Theory, 14, 295-325.

Table 1: Supremum Tests

	H_0	H_1^L	H_1^E	H_1^S	H_1^{AN}	H_1^{ALL}
n = 200						
L-$\sup_{\gamma \in \Gamma_n^p} T_n(\gamma)^a$	.05	.49	.56	.25	.39	.45
E-$\sup_{\gamma \in \Gamma_n^p} T_n(\gamma)$	.06	.73	.66	.50	.41	.55
$\sup_{m \in N_n^p} T_n(m)$	.06	.67	.63	.48	.37	.54
L-$\sup_{\gamma \in \Gamma_n^p} W_n(\gamma)^b$	.05	.91	.84	.68	.67	.72
E-$\sup_{\gamma \in \Gamma_n^p} W_n(\gamma)$	.06	.89	.68	.74	.66	.73
$\sup_{m \in N_n^p} W_n(m)$	.06	.82	.77	.86	.50	.75
n = 500						
L-$\sup_{\gamma \in \Gamma_n^p} T_n(\gamma)$	.04	.67	.53	.55	.64	.69
E-$\sup_{\gamma \in \Gamma_n^p} T_n(\gamma)$	.04	.81	.67	.70	.46	.65
$\sup_{m \in N_n^p} T_n(m)$	.05	.86	.79	.76	.70	.77
L-$\sup_{\gamma \in \Gamma_n^p} W_n(\gamma)$	.04	.95	.92	.91	.73	.87
E-$\sup_{\gamma \in \Gamma_n^p} W_n(\gamma)$	.05	.89	.85	.89	.68	.85
$\sup_{m \in N_n^p} W_n(m)$	.05	.93	.85	.92	.76	.86

Notes: a. Super-consistent tests (L = logistic; E = exponential).
b. Uniformly most-powerful tests.

Table 2: Randomized Tests

	H_0	H_1^L	H_1^E	H_1^S	H_1^{AN}	H_1^{ALL}
n = 200						
L-$T_n(\gamma)^a$	.09	.75	.69	.56	.55	.64
E-$T_n(\gamma)$	.04	.80	.75	.72	.51	.66
$T_n(m)^b$	.05	.82	.77	.74	.51	.65
L-$W_n(\gamma)$	.03	.83	.78	.72	.57	.71
E-$W_n(\gamma)$	.02	.82	.81	.79	.53	.72
$W_n(m)$	.03	.83	.80	.82	.54	.67
n = 500						
L-$T_n(\gamma)$	.08	.74	.72	.69	.73	.72
E-$T_n(\gamma)$	.07	.85	.82	.80	.69	.82
$T_n(m)$	.05	.95	.79	.81	.79	.88
L-$W_n(\gamma)$	.03	.94	.92	.84	.74	.81
E-$W_n(\gamma)$	.03	.86	.93	.91	.70	.85
$W_n(m)$	.07	.96	.84	.88	.80	.90

Notes: a. Real-valued γ are randomly selected from $[.5,10]^p$.
b. Integer-valued m are randomly selected from N_n^p.

Figure 1: Empirical Power Against STAR Alternatives