Consistent Model Specification Tests Against Smooth Transition Alternatives*

Jonathan B. Hill[†]
Dept. of Economics
Florida International University

August 5, 2005

Abstract

In this paper we develop tests of functional form that are consistent against a class of nonlinear "smooth transition" models of the conditional mean. Our method is an extension of the consistent model specification tests developed by Bierens (1990), de Jong (1996) and Bierens and Ploberger (1997), provides maximal power against nonlinear smooth transition ARX specifications, and is consistent against any deviation from the null hypothesis.

Of separate interest, we provide substantial detail regarding when and whether Bierens-type tests are asymptotically degenerate.

In a simulation experiment in which all parameters are randomly selected, and a linear AR null model is selected by minimizing the AIC, the proposed test has power nearly identical to a most powerful test for true STAR processes, and dominates popular tests.

1. Introduction Smooth Transition Threshold Autoregressive (STAR) models have gained significant popularity in the economics and finance literatures as a means to transcend well known explanatory and forecasting limitations of linear and binary switching models. Suggested by Chan and Tong (1986a,b) to account for sluggish regime dynamics in many time series, Teräsvirta

---

*For comments on an earliery version of the paper, I would like to thank Dick van Dijk, Phil Rothman and the participants of the 12th annual meeting of the Society for Nonlinear Dynamics in Economics; as well as Werner Ploberger and the participants of the 2004 North American Econometric Society meeting. Finally, I would like to thank two anonymous referees and an associate editor for comments that led to substantial improvements. All errors, of course, are mine alone.

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

Key Words: smooth transition, consistent test, nondegenerate test, nonlinear, neural networks.
JEL Classification: Primary C12; Secondary C32, C45.

(1994) established a composite theory of estimation, diagnostic checking and inference for smooth transition processes with exponential and logistic transition functions. See, also, Luukkonen *et al* (1988) and van Dijk *et al* (2000).

Consider a time series process $\{y_t, x_t\}$ with regressors $x_t = (1, y_{t-1}, ..., y_{t-p})'$, an innovations process $\{u_t\}$. The standard class of 2-regime STAR processes is represented as

$$y_t = \phi_0' x_t + \phi_1' x_t F(y_{t-d}, \gamma, c) + u_t, \tag{1}$$

for some transition function $F(y_{t-d}, \gamma, c) : \mathbb{R}^3 \rightarrow [0, 1]$, transition scale $\gamma > 0$, threshold variable y_{t-d}, threshold c, and delay parameter d. Tradition specifications for $F(\cdot)$ include the logistic $[1 - \exp\{-\gamma(y_{t-d} - c)\}]^{-1}$ and exponential $\exp\{-\gamma(y_{t-d} - c)^2\}$, cf. Luukkonen *et al* (1988) and Teräsvirta (1994)[1].

Tests for linearity against STAR alternatives have received almost no attention in the theory literature, although a standard practice dominates the applied literature. Luukkonen *et al* (1988), Saikkonen and Luukkonen (1988), Lin and Teräsvirta (1994), Teräsvirta (1994), Hagerud (1997), Gonzalez-Rivera (1998), Escribano and Jorda (2000), van Dijk et al (2000), Medieros and Veiga (2000), Rothman *et al* (2001), Lundberg and Teräsvirta (2002), Lundberg et al (2003), Lundberg and Teräsvirta (2005), and others proscribe a truncated Taylor expansion approximation of the transition function $F_t(d, \gamma, c)$. The technique leads to a simple polynomial auxiliary regression in the spirit of Ramsey(970), Ramsey and Schmidt (1976) and Keenan (1985), and standard F-tests of parametric zero-restrictions in order to determine whether the process is linear, exponential or logistic STAR[2].

In order for the polynomial regression to have meaning in a STAR framework, the true data generating process is simply assumed to be a STAR (see, e.g., Teräsvirta, 1994; Lundberg and Teräsvirta, 2005). Thus, the test is not a true test against smooth transition alternatives, per se. This issue is particularly relevant if we admit *any* functional alternative to explain the data provided linearity is found inadequate. The use of polynomial functionals in weight-based conditional moment tests is known not to lead to a consistent test in the sense that the asymptotic power of the test against any deviation from the null is not guaranteed to be one (Stinchcombe and White, 1998: Theorem 2.3). This shortcoming of classic weight-based conditional moment tests has been extensively documented in the inference theory and artificial neural network literatures: see, e.g., Davies (1977), Holley (1982), Bierens (1990), Kuan and White (1994), Bierens and Ploberger (1997), and Stinchcombe and White (1998).

---

[1] See also Chan and Tong (1986b) for the so-called normal STAR model based of a normal cdf transition function; Lin and Terasvirta (1994) and Lundberg *et al* (2003) for use of time as the threshold variable in the so-called time-varying smooth transition model (TV-STAR); and Lundberg and Terasvirta (2005) for the so-called STARZ model used to capture multiple thresholds (e.g. exchange rate zone targeting).

[2] Lundberg and Teräsvirta (2005) do not actually propose a test for linear autoregression against a STARZ alternative: they employ the polynomial regression test in order to analyze whether further nonlinear terms should be added to the nonlinear STARZ specification.

Moreover, the "delay" parameter d remains in the polynomial regression and is not defined under the null. Teräsvirta (1994) and many others suggest selecting that d which generates the lowest p-value. This is mathematically equivalent to maximizing the test statistic over an integer subspace which makes the choice of d data dependent, and therefore the test statistic will have a non-standard limit distribution under the null (see Hansen, 1996, for comments regarding TAR models in general). Nevertheless, in both econometrics textbooks and empirical applications regarding STAR tests the standard suggested practice is simply the employment of critical values derived from the chi-squared distribution: see, e.g. Saikkonen and Luukkonen (1988), Hagerud (1997), Gonzalez-Rivera (1998), Franses and van Dijk (2000), and Rothman *et al* (2001).

Furthermore, most smooth transition models in the literature incorporate only one threshold variable for test purposes, and in some cases only (non-stochastic) time. See, e.g., Lin and Teräsvirta (1994), van Dijk *et al* (2000), and Lundbergh *et al* (2003). Test consistency will require each stochastic variable that enters into the null specification to enter into the weight function.

In this paper, we develop a consistent conditional moment test of functional form in the tradition of Bierens (1982, 1984, 1990), de Jong (1996) and Bierens and Ploberger (1997). In particular, we develop a test statistic that directs power toward nonlinear smooth transition ARX alternatives, is consistent against any deviation from the null specification, and nests classic STAR and neural network specifications (e.g. Bierens, 1990; White, 1989; Lee *et al*, 1993; de Jong, 1996). We use a nonlinear ARX framework with a suitable notion of dependence and heterogeneity in order to promote generality. All of the theory developed here straightforwardly extends to nonlinear ARMAX models. A nonlinear ARX framework easily allows for a test of an arbitrary l-regime STARX under the null against an l + 1-regime STARX under the alternative for any $l \geq 1$.

Test consistency is expedited by considering general forms of the transition function and its argument. We consider the ramifications for consistency when standard parametric restrictions employed in the STAR literature are encountered: this includes the use of unique thresholds and scales for each threshold variable; and the use of quadratic arguments.

Furthermore, and of separate interest, we substantially improve upon results provided in Bierens (1990) and de Jong (1996) regarding when consistent conditional moment test statistics are asymptotically degenerate (singular covariance matrix: see Section 5 for details). In general we link non-degeneracy to non-consistency. Moreover, for tests of linear autoregression against standard neural network and STAR alternatives we prove the test statistic developed here, including the Bierens test as a special case, is never degenerate, except in a trivial case.

Consistent non-parametric moment based tests exist: see, e.g., Lee (1988), Yatchew (1992), Wooldridge (1992), Zheng (1996), Hong and White (1995), Li *et al* (2003), to name a few. Non-parametric methods are suitable for testing whether a particular functional specification is correct with probability one, but do not, in general, provide a better parametric specification if the null specification is determined false.

In a simulation study we provide evidence that our test dominate parametric tests by Bierens (1990), Lee *et al* (1993) and Teräsvirta (1994). Indeed, the power of the proposed tests against STAR, ANN and SETAR alternatives nearly matches that of most powerful tests. Our simulation is substantially less restrictive than previous such studies (e.g. Luukkonen *et al*, 1988; Bierens, 1990; Teräsvirta, 1994; de Jong, 1996): we do not fix slope parameters, and therefore control for the fact that conveniently chosen parameters may bias test results. In such a setting, the popular polynomial regression test is shown to be vastly dominated by our proposed test, and by conventional tests of nonlinearity.

In Section 2 we detail the smooth transition ARX framework. Section 3 develops smooth transition moment conditions, the main results are contained in Section 4, and Section 5 characterizes when and whether the test statistic is asymptotically degenerate. We demonstrate asymptotic validity of a simulation technique for approximating the asymptotic p-value in Section 6, and compare the proposed test to existing parametric tests in a simulation study in Section 7.

We maintain the following notation conventions. Let $|\cdot|_p$ denote the l_p-norm for real-valued vectors, and the matrix norm for real-valued matrices: $|x|_p = (\sum_{i,j} |x_{i,j}|^p)^{1/p}$. In all cases $|\cdot|$ denotes $|\cdot|_1$, the l_1-norm. Let $\|\cdot\|_p$ denote the L_p-matrix norm: $\|x\|_p = (\sum_{i,j} E|x_{i,j}|^p)^{1/p}$. For arbitrary k-vectors a and x, vector powers x^a are understood to represent $(x_1^{a_1}, ..., x_k^{a_k})'$. I_k denotes a k-dimensional identity matrix. $\rightarrow$ denotes convergence in probability or distribution; $\Longrightarrow$ denotes weak convergence with respect to the uniform metric; $[x]$ denotes the integer part of x.

2. Nonlinear STARX Framework

Let $z_t = (y_t, x_t')'$ be a stochastic process in $\mathbb{R} \times \mathbb{R}^{k-1}$, $k \geq 1$. Assume the k-vector process $\{z_t\}$ is strictly stationary, ergodic and exists in $L_2(\Omega, \Im_t, P)$ where $\Im_t$ denotes the increasing σ-algebra induced by $(y_{t-i}, x_{t-i})_{i=0}^{\infty}$. In the case of a purely autoregressive framework $k = 1$ is understood, and $\Im_t = \sigma(y_{t-i} : i \geq 0)$. For notational convenience we assume x_t does not contain a constant, and is $\Im_t$-measurable.

In order to restrict memory in the process z_t, and have an accessible uniform law of large numbers and central limit theorem applicable in heterogenous nonlinear ARMAX settings, we utilize the concept of υ-stability, cf. Bierens (1981, 1984, 1987, 1991, 1994). See Pötscher and Prucha (1991), Bierens (1994) and Davidson (1994). The following ARX framework borrows heavily from Bierens (1991, 1994) and complete details may be obtained from those sources. Consult Appendix 1 for fundamental assumptions detailed under Assumption A.

2.1 Smooth Transition ARX

Let $\tilde{z}_{it} \equiv (1, z_{t-1}', ..., z_{t-p_i}')' \in \mathbb{R}^{p_i k+1}$, $i = 1, 2$, $0 \leq p_2 \leq p_1$, where $\tilde{z}_{it,0} = 1$. Let $f_t(\phi) = f(\tilde{z}_{1t}, \phi)$ denote a known response function, $f_t : \mathbb{R}^{p_1 k+1} \times \Phi \rightarrow \mathbb{R}$, measurable with respect to $\Im_{t-1}$, with Φ a compact subset of $\mathbb{R}^m$, for some $m \geq 1$ which depends on the null specification. The data generating process of

$\{y_t\}$ has a 2-regime nonlinear smooth transition ARX form[3]

$$y_t = f_t(\phi_0) + \phi_1' \tilde{z}_{2t}(\delta) F_t(\tau) + \epsilon_t, \tag{2}$$

where $\phi_0 \in \Phi$, $\phi_1 \in \mathbb{R}^{p_2 k+1}$, $\tau \in \mathbb{R}^{p_1 k+1}$, $F_t(\tau) = F(\tau' \tilde{z}_{1t})$, and ϵ_t satisfies $E[\epsilon_t \tilde{z}_{1t}] = E[\epsilon_t (\partial/\partial\phi) f_t(\phi_0)] = 0$ under either hypothesis. Traditionally $F(\cdot)$ denotes the exponential or logistic, however we only require $F(\cdot)$ to be non-polynomial and analytic: see Section 3.

We define $\tilde{z}_{2t}(\delta)$ as a bounded, $\Im_{t-1}$-measurable mapping from $\mathbb{R}^{p_2 k+1} \times \Delta$ to $\mathbb{R}^{p_2 k+1}$, with Δ a compact subset of $\mathbb{R}^{(p_2 k+1)\times q}$, $q \geq 0$, and $P(\inf_{\delta\in\Delta} |\tilde{z}_{2t}(\delta)| > a_0) = 1$ and $P(\sup_{\delta\in\Delta} |\tilde{z}_{2t}(\delta)| < a_1) = 1$ for some constants $0 < a_0 < a_1 < \infty$. It is understood that $\tilde{z}_{2t}(\delta)$ is a parametric function of $\tilde{z}_{2t}$, where $\tilde{z}_{2t}(0) = \tilde{z}_{2t}$ by convention[4]. If $q = 0$, then $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t}$; if $q \geq 1$, $\tilde{z}_{2t}(\cdot)$ is a Borel measurable function of $p_2 k + 1$ vectors δ_i, $i = 1...q$, and we write $\delta = vec(\delta_i)_{i=1}^{q}$.

$f_t(\phi)$ and $F_t(\tau)$ are understood to be functions of the "complete" non-parametric regressor set $\tilde{z}_{1t}$. The weight $F(\tau' \tilde{z}_{1t})$ must utilize all regressors contained in the null specification $f(\tilde{z}_{1t}, \phi)$ as a necessary condition for test consistency.

The possibly nonlinear function $f_t(\phi)$ provides a substantial degree of flexibility, allowing for an l-regime STAR, $l \geq 1$, under the null (where $l = 1$ implies a linear autoregression), neural network terms, and much more.

2.2 Traditional STAR Specifications We utilize the term $\phi_1' \tilde{z}_{2t}(\delta)$ in order to nest LSTARX, ESTARX, and neural network alternatives. For example, an **LSTARX** model with thresholds and scales uniquely defined for each threshold variable $\tilde{z}_{1t,i}$ can be specified by setting $q = 2$, $p_1 = p_2$, and

$$\begin{aligned} \text{LSTARX:} \quad & \tilde{z}_{2t}(\delta) = \tilde{z}_{2t} \prod\nolimits_{i=1}^{p_1 k} \left[\exp\{\delta_{1i} + \delta_{2i} \tilde{z}_{1t,i}\} + 1\right]^{-1} \\ & F_t(\tau) = \exp\{\tau' \tilde{z}_{1t}\}. \end{aligned} \tag{3}$$

Notice that $F_t(\tau)$ itself is exponential[5]. Hence, a particular re-parameterization renders a "compound" logistic weight with threshold variable specific scales γ_i and thresholds c_i:

$$\tilde{z}_{2t}(\delta) F_t(\tau) = \tilde{z}_{2t} \prod\nolimits_{i=1}^{p_1 k} \left[1 + \exp\{-\gamma_i (\tilde{z}_{1t,i} - c_i)\}\right]^{-1}, \tag{4}$$

provided γ and c satisfy $\sum_{i=1}^{p_1 k} \gamma_i c_i = -\tau_0$, and for $i = 1...p_1 k$, $\gamma_i = \delta_{2i} = \tau_i$ and $\gamma_i c_i = -\delta_{1i}$.

---

[3] It would be straightforward to generalize the innovations ϵ_t to a class of linear finite dependent processes allowing for a smooth transition ARMAX representation in (2). The inclusion of moving average terms would only add unnecessary notation. See de Jong (1996).

[4] For example, $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t}$ for all δ; or $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t} \exp\{\delta' \tilde{z}_{2t}\}$.

[5] The following is in no way offered as an exhaustive treatment of possible transition function specifications. See the citiations in Section 1 for further variations of the logistic and exponential STAR models, and for the normal STAR model.

Such a parameterization directly relates δ to τ: for a chosen δ we are not free to choose τ. A consistent test with such a parametric restriction is not guaranteed to exist: see Remark 6 of Lemma 1, below.

In general, however, $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t}$ and $F_t(\tau) = [1 + \exp\{\tau'\tilde{z}_{1t}\}]^{-1}$ suffice for test consistency (cf. Lemma 1, below). Similarly (3), without parametric restrictions on δ and τ nests (4), and also suffices for test consistency.

ESTAR models typically have the transition form $\exp\{-\gamma'(\tilde{z}_{2t} - c)^2\}$, $\gamma > 0$, in order to capture "inner" versus "outer" regimes[6]. Set $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t}\exp\{\sum_{i=1}^{p_1 k}\delta_i \tilde{z}_{1t,i}^2\}$ (i.e. $\delta_0 = 0$) and $F_t(\tau) = \exp\{\sum_{i=0}^{p_1 k}\tau_i \tilde{z}_{1t,i}\}$, hence

$$\text{ESTARX:}\ \ \tilde{z}_{2t}(\delta)F_t(\tau) = \tilde{z}_{2t}\exp\{\sum\nolimits_{i=1}^{p_1 k} -\gamma_i(\tilde{z}_{1t,i} - c_i)^2\}, \qquad (5)$$

provided $\{\gamma_i\}_{i=1}^{p_1 k}$ and $\{c_i\}_{i=1}^{p_1 k}$ satisfy $\gamma_i = -\delta_i > 0$, $\gamma_i c_i = \tau_i/2$ for $i = 1...p_1 k$, and $\tau_0 = -\sum_{i=1}^{p_1 k}\gamma_i c_i^2$. Notice $\{\gamma_i\}_{i=1}^{p_1 k}$ and $\{c_i\}_{i=1}^{p_1 k}$ may be solved as functions of $\{\delta_i\}_{i=1}^{p_1 k}$ and $\{\tau_i\}_{i=1}^{p_1 k}$ without implicitly relating δ_i and τ_i: For any $\delta > 0$ and τ, set $\gamma_i = -\delta_i > 0$, then $c_i = -\tau_i/(\delta_i 2)$, and solve $\tau_0 = \sum_{i=1}^{p_1 k}\tau_i^2/(\delta_i 4)$. The fact that τ_0 is a function of $\{\tau_i\}_{i=1}^{p_1 k}$ and $\{\delta_i\}_{i=1}^{p_1 k}$ is irrelevant in the present case: see Remark 4 of Lemma 1, below.

Finally, a **neural network** setting has $q = p_2 = 0$ such that $\tilde{z}_{2t}(\delta) = 1$, and $y_t = f_t(\phi_0) + \phi_1 F_t(\tau) + \epsilon_t$. Cf. White (1989), Bierens (1990) and de Jong (1996).

In general, and in all that follows, we do not place any restrictions on $\tilde{z}_{2t}(\delta)$ and $F_t(\tau)$, a la functional form and δ and τ, other than those detailed in Assumptions A and B, below.

2.3 Hypotheses The null hypothesis is simply $\phi_1 = 0$. In a general framework, the hypotheses of interest are as follows:

$$\begin{aligned} H_0 &: P\left(E\left[y_t - f_t(\phi_0)|\mathfrak{F}_{t-1}\right] = 0\right) = 1, \text{ for some } \phi_0 \in \Phi \qquad (6) \\ H_1 &: \sup\nolimits_{\phi \in \Phi} P\left(E\left[y_t - f_t(\phi)|\mathfrak{F}_{t-1}\right] = 0\right) < 1. \end{aligned}$$

Under H_0 the function $f_t(\phi_0)$ is *almost surely* correctly specified for $E\left[y_t|\mathfrak{F}_{t-1}\right]$ such that ϵ_t forms a martingale difference sequence, $E[\epsilon_t|\mathfrak{F}_{t-1}] = 0$. The general alternative H_1 is simply that the null model is mis-specified, embracing any deviation from the null. See, e.g., Bierens (1990) and Bierens and Ploberger (1997).

3. Smooth Transition Conditional Moments For compactness, define $\theta \equiv (\delta', \tau')'$. If $\tilde{z}_{2t}(\delta)$ does not depend on δ (e.g. $q = 0$), then it is understood that $\theta = \tau$.

Denote by $\hat{s}_n(\phi_0, \phi_1, \theta)$ the nonlinear least squares score associated with (2). Evaluating the score under the null $\phi_1 = 0$ and using the nonlinear least squares

---

[6] See Luukkonen *et al* (1988), Saikkonen and Luukkonen, 1988) and Teräsvirta (1994).

estimator $\hat{\phi}$ for ϕ_0, it is easy to show

$$\hat{s}_n(\hat{\phi}, 0, \theta) = n^{-1} \sum\nolimits_{t=1}^{n} \hat{\epsilon}_t \tilde{z}_{2t}(\delta) F_t(\tau), \quad p_2 k + 1 \times 1, \tag{7}$$

where $\hat{\epsilon}_t = y_t - f_t(\hat{\phi})$.

In Lemma 1, below, we prove the corresponding vector moment condition $E[\epsilon_t \tilde{z}_{2t}(\delta) F_t(\tau)]$ has the same "totally revealing" properties as the orthogonality conditions considered in Stinchcombe and White (1998). Such a property, along with weak convergence of a properly scaled $\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ on a space of continuous functions under the null, will deliver a consistent LM test statistic in Section 4 (Theorems 2 and 3). In Section 5 we then analyze when and whether the associated covariance matrix of $\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ is singular, rendering the test statistic asymptotically degenerate.

Stinchcombe and White (1998) prove that any non-polynomial real analytic function F[7] has the desired "totally revealing" property (i.e. promoting Lemma 1, below). This includes the exponential $F(u) = \exp\{u\}$ and logistic $F(u) = (1 + \exp\{u\})^{-1}$ [8].

Assumption B The function $F : \mathbb{R} \rightarrow \mathbb{R}$ is non-polynomial and analytic on some open interval R_0 of $\mathbb{R}$.

Lemma 1 *Let ϵ_t be a random variable satisfying $E|\epsilon_t| < \infty$, and let $\tilde{z}_{1t}$ be an $\mathfrak{F}_{t-1}$-measurable bounded vector in $\mathbb{R}^{p_1 k+1}$ such that $P(E[\epsilon_t|\tilde{z}_{1t}] = 0) < 1$. Let Assumption B hold for $F(\cdot)$. For each $\delta \in \Delta$ and each $i = 1...p_2 k + 1$, the sets*

$$S_i = \left\{\tau \in \mathbb{R}^{p_1 k+1} : E[\epsilon_t \tilde{z}_{2t,i}(\delta) F(\tau' \tilde{z}_{1t})] = 0\right\} \text{ and } P\left(\tau' \tilde{z}_{1t} \in R_0\right) = 1\right\}, \tag{8}$$

have Lebesgue measure zero, and are nowhere dense in $\mathbb{R}^{p_1 k+1}$.

Remark 1: In general the sets S_i will depend on the distribution of $\{y_t, x_t\}$, and on each point $\delta \in \Delta$ (unless $q = 0$).

Remark 2: The set $S \equiv \cap S_i$, the collection of each τ such that the vector $E[\epsilon_t \tilde{z}_{2t}(\delta) F(\tau' \tilde{z}_{1t})] = 0$ has Lebesgue measure zero because subsets of measure zero sets have measure zero.

Remark 3: Conditioning on $\tilde{z}_{1t}$ is equivalent to conditioning on any bounded, measurable, one-to-one function of $\tilde{z}_{1t}$, say $\Psi(\tilde{z}_{1t}) : \mathbb{R}^{p_1 k+1} \rightarrow \mathbb{R}^{p_1 k+1}$, since any such functional induces the same σ-field as $\tilde{z}_{1t}$: see Billingsley (1995: Theorem 5.1).

Remark 4: Consider the ESTAR weight (5) re-parameterization $\gamma_i = -\delta_i > 0$ and $c_i = -\tau_i/(\delta_i 2)$ for $i = 1...pk$, and $\tau_0 = \sum_{i=1}^{p_1 k} \tau_i^2/(\delta_i 4)$. The

---

[7] A real function $F : \mathbb{R} \rightarrow \mathbb{R}$ is analytic on a domain $R_0 \subseteq \mathbb{R}$ if it possesses derivatives of all orders at each point in R_0, and is equal to its Taylor series in a neighborhood of every point of the domain R_0. See Evgrafov (1978: Theorem 3.3). This includes the exponential, logistic, trigonometric functions (e.g. sin + cos) and the polynomials.

[8] Interestingly, even some non-analytic functions have the desired properties, including the normal cdf and pdf, cf. Theorem 3.10 of Stinchcombe and White (1998). Along with Lemma 1, this substantiates the normal STAR model of Chan and Tong (1986b).

restriction $\tau_0 = \sum_{i=1}^{p_1k} \tau_i^2/(\delta_i 4)$ may always be enforced: under H_1, for any δ we have $E[\epsilon_t \tilde{z}_{2t,i}(\delta)\exp\{\tau' \tilde{z}_{1t}\}] \neq 0$ except for $\tau \in S$, hence $E[\epsilon_t \tilde{z}_{2t,i}(\delta)\exp\{(\tau_0 + a) + \sum_{i=1}^{p_1k} \tau_i \tilde{z}_{1t,i}\}] \neq 0$ for such $\{\tau_i\}_{i=0}^{p_1k}$ and any $a \in \mathbb{R}$, and therefore for $\tau_0 = \sum_{i=1}^{p_1k} \tau_i^2/(\delta_i 4)$. Moreover, because Lemma 1 holds for any δ, and because the re-parameterization does not restrict our choice of τ, under the alternative, $E[\epsilon_t \tilde{z}_{2t} \exp\{\sum_{i=1}^{p_1k} -\gamma_i(\tilde{z}_{1t,i} - c_i)^2\}] \neq 0$ except for countably many c (i.e. except for countably many τ).

Thus, the power of the test does not rely on the scale γ (provided $\gamma \neq 0$). Rather, the power of a conditional moment test directed toward a traditional ESTAR alternative intimately depends on the choice of the threshold c. Indeed, if $\{c_i\}_{i=1}^{p_1k} = 0$ then $\{\tau_i\}_{i=1}^{p_1k} = 0$ must be true and $E[\epsilon_t \tilde{z}_{2t,i}(\delta) F(\tau' \tilde{z}_{1t})]$ reduces to $F(\tau_0) E[\epsilon_t \tilde{z}_{2t} \exp\{\sum_{i=1}^{p_1k} \delta_i \tilde{z}_{1t,i}^2\}]$. The argument $\sum_{i=1}^{p_1k} \delta_i \tilde{z}_{1t,i}^2$ is not a one-to-one function of $\tilde{z}_{1t}$ hence Lemma 1 is not guaranteed to hold. Therefore, we cannot arbitrary fix $c = 0$ as is often suggested in the smooth transition conditional volatility literature: see, e.g., Gonsalez-Rivera (1998) and van Dijk and Franses (1999).

Remark 5: The preceding discussion demonstrates that the argument of the exponential weight *need not* incorporate a one-to-one function of $\tilde{z}_{1t}$: compare $\exp\{-\sum_{i=1}^{p_1k} \gamma_i(\tilde{z}_{1t,i} - c_i)^2\}$ for $c \neq 0$ to $\exp\{-\gamma' \tilde{z}_{1t}^2\}$.

Remark 6: Lemma 1 states the sets S_i will have Lebesgue measure zero *for each* δ: for a chosen δ a consistent test may be generated by noting $E[\epsilon_t \tilde{z}_{2t}(\delta) F(\tau' \tilde{z}_{1t})] \neq 0$ for all τ in a compact subset of $\mathbb{R}^{p_1k+1}$ with positive Lebesgue measure. In the LSTARX specification (4), δ and τ are directly related via $\delta_{2i} = \tau_i$, hence *for each* δ there is *only one* τ that can be considered, and of course this τ may be in $S = \cap S_i$. However, specification (3), without parametric restrictions on δ and τ, nests (4) and satisfies Lemma 1.

4. Consistent Test of Linearity Against Smooth Transition Alternatives

Let $\theta = (\delta', \tau')' \in \Theta$, where Θ is a compact subset of $\mathbb{R}^{p_1k+1+q(p_2k+1)}$ with positive Lebesgue measure. If $q = 0$, then $\theta = \tau \in \Theta \subseteq \mathbb{R}^{p_1k+1}$ is understood. The asymptotics of continuous test statistic functionals will require weak convergence of a scaled sample score $\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ under the null to a Gaussian element (understood to be mean-zero) in $\mathbb{C}[\Theta]$. $\mathbb{C}[\Theta]$ denotes the vector metric space of continuous real functions on Θ, endowed with uniform metric $\sup_{\theta\in\Theta} |w_1(\theta) - w_2(\theta)|$. Cf. Pollard (1984) and Billingsley (1999).

Write $\partial f_t(\phi) \equiv (\partial/\partial\phi) f_t(\phi)$. For any point $\phi \in \Phi$ define

$$s_n(\phi, 0, \theta) = n^{-1} \sum_{t=1}^{n} \epsilon_t g_t(\theta) \text{ - } p_2k + 1 \times 1, \tag{9}$$

where $\epsilon_t = y_t - f_t(\phi)$, and

$$\begin{aligned}
g_t(\theta) &= F_t(\tau)\tilde{z}_{2t}(\delta) - b(\theta)A^{-1}\partial f_t(\phi_0) && p_2k + 1 \times 1 \\
b(\theta) &= E\left[F_t(\tau)\tilde{z}_{2t}(\delta)\partial' f_t(\phi_0)\right] && p_2k + 1 \times m \\
A &= E\left[\partial f_t(\phi_0)\partial' f_t(\phi_0)\right] && m \times m \\
V(\theta_1, \theta_2) &= E\left[\epsilon_t^2 g_t(\theta_1) g_t(\theta_2)'\right] && p_2k + 1 \times p_2k + 1.
\end{aligned} \tag{10}$$

For the covariance function $V(\theta_1, \theta_2)$ we write $V(\theta) = V(\theta, \theta)$. Similarly, define the nonlinear least squares sample conjugates $\hat{\epsilon}_t = y_t - f_t(\hat{\phi})$ and

$$\begin{aligned}
\hat{g}_t(\theta) &= F_t(\tau)\tilde{z}_{2t}(\delta) - \hat{b}(\theta)\hat{A}^{-1}\partial f_t(\hat{\phi}) \\
\hat{b}(\theta) &= n^{-1}\sum\nolimits_{t=1}^{n} F_t(\tau)\tilde{z}_{2t}(\delta)\partial' f_t(\hat{\phi}) \\
\hat{A} &= n^{-1}\sum\nolimits_{t=1}^{n} \partial f_t(\hat{\phi})\partial' f_t(\hat{\phi}) \\
\hat{V}(\theta) &= n^{-1}\sum\nolimits_{t=1}^{n} \hat{\epsilon}_t^2 \hat{g}_t(\theta)\hat{g}_t(\theta)'.
\end{aligned} \tag{11}$$

4.1 Weak Convergence of $\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ In order for $V(\theta)^{-1/2}$ to exist[9], for now we invoke the following assumption. For arbitrary $\xi \geq 0$ define the compact subspace

$$\Theta_\xi = \{\theta = (\delta', \tau')' \in \Theta\text{: } |\tau| \geq \xi\}. \tag{12}$$

By convention we write $\Theta_0 = \Theta$. Thus for all $\xi > 0$, $0 \notin \Theta_\xi$. For each point $\theta \in \Theta_\xi$ denote by $\lambda_0(V(\theta))$ the minimum eigenvalue of $V(\theta)$, $\lambda_0(\cdot) \leq \lambda_1(\cdot)$, ...

Assumption C $\inf_{\theta \in \Theta_\xi} \lambda_0(V(\theta)) > 0$.

Remark 1: The condition $\inf_{\theta \in \Theta_\xi} \lambda_0(V(\theta)) > 0$ ensures the covariance matrix $V(\theta)$ is "uniformly positive definite" in Θ_ξ. We leave for Section 5 an analysis of when Assumption C is necessarily satisfied for any $\xi > 0$.

Remark 2: Bounding $|\tau| \geq \xi > 0$ is required in order to demonstrate both pointwise convergence and tightness of $\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ (consult Lemmas A.3, A.5, A.7 and A.8 in Appendix 3), and is reminiscent of of bounding requirements in Andrews (1993).

Weak convergence of $\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ in $\mathbb{C}[\Theta_\xi]$ follows from pointwise convergence to a multivariate normal distribution and tightness on Θ_ξ, cf. Theorem 7.1 of Billingsley (1999). Each property is established in fundamental lemmas presented in Appendix 3: consult the line of proof of the subsequent theorem.

Theorem 2 *Under Assumptions A-B and H_0, $\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ converges weakly to a Gaussian element $s(\theta)$ of $\mathbb{C}[\Theta]$ with covariance function*

$$E[s(\theta_1)s(\theta_2)'] = V(\theta_1, \theta_2) = E\left[\epsilon_t^2 g_t(\theta_1) g_t(\theta_2)'\right]. \tag{13}$$

If additionally Assumption C holds, then $\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ converges weakly to a Gaussian element $z(\theta)$ of $\mathbb{C}[\Theta_\xi]$ with covariance function $E[z(\theta_1)'z(\theta_2)] = V(\theta_1)^{-1/2}V(\theta_1, \theta_2)V(\theta_2)^{-1/2}$.

---

[9] Throughout, for any real symmetric positive definite matrix X we define $X^{1/2} \equiv C\Lambda^{1/2}C'$, with C the orthogonal matrix of eigenvectors, and Λ the diagonal matrix of eigenvalues. Thus $X^{1/2}$ and $X^{-1/2}$ are trivially symmetric and positive definite.

Moreover, under Assumptions A-C and H_1*, there exists some function* $\eta : \Theta_\xi \to \mathbb{R}^{p_2 k+1}$ *such that*

$$\sup_{\theta \in \Theta_\xi} \left| \hat{V}(\theta)^{-1/2} \hat{s}_n(\hat{\phi}, 0, \theta) - V(\theta)^{-1/2} \eta(\theta) \right| = o_p(1), \tag{14}$$

where $V(\theta)^{-1/2}\eta(\theta) \neq 0$ *for all* $\theta = (\delta', \tau')' \in \Theta_\xi$ *except possibly for* τ *in a set* S *with Lebesgue measure zero.*

4.2 Test Statistics

Define a standard LM statistic

$$T_n(\theta) = n\hat{s}_n(\hat{\phi}, 0, \theta)' \hat{V}(\theta)^{-1} \hat{s}_n(\hat{\phi}, 0, \theta). \tag{15}$$

Under H_0, Theorem 2 and the continuous mapping theorem suffice to show $T_n(\theta) \Rightarrow z(\theta)'z(\theta)$ a chi-squared process (cf. Hansen, 1991), and $T_n(\theta)/n \to \eta(\theta)'V(\theta)^{-1}\eta(\theta)$ in probability under H_1, where $\eta(\theta)'V(\theta)^{-1}\eta(\theta) > 0$ for every $\theta \in \Theta_\xi$ except $\tau \in S = \cap S_i$. Two fundamental methods for handling the nuisance parameter set θ are randomization on an arbitrary compact subset of $\mathbb{R}^{p_1 k+1}$; or a data dependent selection method. Randomization will sacrifice power (see the simulation study of Section 6), in general, however $T_n(\theta)$ will have for each point θ a chi-squared limiting null distribution. See White (1989) and Lee *et al* (1996).

Popular approaches to the latter method include the supremum $\sup_{\theta \in \Theta_\xi} T_n(\theta)$, and the average $\text{ave}_{\Theta_\xi} T_n(\theta) = \int_{\Theta_\xi} T_n(\theta) dw(\theta)$ for some measure $w(\theta)$ absolutely continuous with respect to Lebesgue measure, with compact support with positive Lebesgue measure (e.g. uniform on Θ_ξ): see Davies (1977, 1987), Andrews (1993) and King and Shively (1993), and Andrews and Ploberger (1994)[10]. Each statistic may simply be represented as functions $g_n \equiv g(T_n(\theta))$ continuous with respect to the uniform metric, $g : \mathbb{C}[\Theta_\xi] \to \mathbb{C}[\Theta_\xi]$. Theorem 2 and the continuous mapping theorem therefore suffice to prove the following.

Theorem 3 *Under Assumptions A-C and* H_0*, there exists a Gaussian element* $z(\theta)$ *of* $\mathbb{C}[\Theta_\xi]$ *with covariance function* $E[z(\theta_1)'z(\theta_2)] = V(\theta_1)^{-1/2}V(\theta_1, \theta_2)V(\theta_2)$*,* $T(\theta) \equiv z(\theta)'z(\theta)$ *and* $g_0 \equiv g(T(\theta))$*, such that* $T_n(\theta) \Rightarrow T(\theta)$ *and* $g_n \Rightarrow g_0$*.*

Moreover, under H_1

$$g(T_n(\theta)/n) \to g\left(\eta(\theta)'V(\theta)^{-1}\eta(\theta)\right) \tag{16}$$

in probability where the function η *is defined in Theorem 2, and* $\eta(\theta)'V(\theta)^{-1}\eta(\theta) > 0$ *for all* $\theta = (\delta', \tau')' \in \Theta_\xi$ *except for* τ *in a set* S *with Lebesgue measure zero. In particular* $g\left(\eta(\theta)'V(\theta)^{-1}\eta(\theta)\right) > 0$ *and* $g(T_n(\theta)) \to \infty$ *with probability one.*

---

[10] See Bierens (1982) and Bierens and Ploberger (1997) for the integrated conditional moment (ICM) test in which $\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ itself is integrated.

5. Non-Singular Covariance Matrices There exist trivial cases in which $V(\theta)$ is singular: if $\theta = 0$, then $\tilde{z}_{2t}(0) = \tilde{z}_{2t}$ and $F(0'\tilde{z}_{1t}) = F(0)$ is a constant with probability one. This implies $\hat{s}_n(\hat{\phi}, 0, 0) = 0$ by the least-squares first-order conditions, therefore $V(0) = 0$, a zero-matrix. In this section, we analyze the set of all τ for which $V(\theta)$ is singular for any δ.

Assumption D For each t, $P(E[\epsilon_t^2|\tilde{z}_{1t}] \geq \varsigma) = 1$ for some time invariant real number $\varsigma > 0$.

Assumption E The matrix $b(\theta)$ defined in (10) has full column and row rank for each point $\theta \in \Theta$. The matrix $E[\tilde{z}_{2t}(\delta)\tilde{z}_{2t}(\delta)']$ is uniformly positive definite in Δ: $\inf_{\delta \in \Delta} \lambda_0(E[\tilde{z}_{2t}(\delta)\tilde{z}_{2t}(\delta)']) > 0$.

Remark: Assumption D is standard. Assumption E will be easily satisfied in many cases of interest by invoking Assumption A, in particular for the functions typically employed in the smooth transition literature. For example, let $f_t(\phi) = \phi'\tilde{z}_{1t}$, $q = 0$, $p_1 = p_2$ and $P(\inf_{\tau \in T} F_t(\tau) > 0) = 1$[11]. Then $E[\tilde{z}_{2t}(\delta)\tilde{z}_{2t}(\delta)'] = E[\tilde{z}_{2t}\tilde{z}_{2t}']$, a positive definite matrix by Assumption A (and the infimum operation in Assumption E is irrelevant). Similarly, $b(\theta) = E\left[F_t(\tau)\tilde{z}_{2t}(\delta)\partial' f_t(\phi_0)\right] = E\left[F_t(\tau)\tilde{z}_{1t}\tilde{z}_{1t}'\right]$, is positive definite because $F_t(\tau)^{1/2} r'\tilde{z}_{1t} = 0$ with positive probability for any $r \in \mathbb{R}^{p_1 k+1}$, $r \neq 0$, *if and only if* $r'\tilde{z}_{1t} = 0$ with positive probability, which is ruled out by Assumption A.

Define the following set for any $\xi \geq 0$ and any $\delta \in \Delta$

$$S_\xi^* = \left\{\tau \in \mathbb{R}^{p_1 k+1},\ |\tau| \geq \xi : \lambda_0(V(\theta)) = 0 \text{ and } P\left(\tau'\tilde{z}_{1t} \in R_0\right) = 1\right\}. \tag{17}$$

Thus, S_ξ^* contains bounded τ such that $V(\theta)$ is singular and $\tau'\tilde{z}_{1t}$ almost surely lies within the domain on which F is non-polynomial and analytic. Notice $S_0^* = \{\tau \in \mathbb{R}^{p_1 k+1} : \lambda_0(V(\theta)) = 0 \text{ and } P\left(\tau'\tilde{z}_{1t} \in R_0\right) = 1\}$.

Lemma 4 *For any $\delta \in \Delta$ define the following sets for all $r \in \mathbb{R}^{p_2 k+1}$, $r'r = 1$,*

$$S(r) = \{\tau \in \mathbb{R}^{p_1 k+1} : E[\epsilon_t r'\tilde{z}_{2t}(\delta)F(\tau'\tilde{z}_{1t})] = 0, \\ \text{and } P(\tau'\tilde{z}_{1t} \in R_0) = 1\}. \tag{18}$$

Under Assumptions A, B, D and E, for any $\delta \in \Delta$ the set $S_0^* \subseteq S(r)$ for some $r \in \mathbb{R}^{p_2 k+1}$, $r'r = 1$. If $P(E[\epsilon_t|\tilde{z}_{1t}] = 0) < 1$, then this set $S(r)$ has for any $\delta \in \Delta$ Lebesgue measure zero and is nowhere dense in $\mathbb{R}^{p_1 k+1}$.

Any $\tau \in \mathbb{R}^{p_1 k+1}$ that renders $V(\theta)$ singular must also render $E[\epsilon_t r'\tilde{z}_{2t}(\delta)F(\tau'\tilde{z}_{1t})] = 0$ for some $r \in \mathbb{R}^{p_2 k+1}$, $r'r = 1$, and any $\delta \in \Delta$. This is trivial under the null hypothesis[12], but gains importance when the alternative is true by Lemma 1. Thus, under H_1 we always know *something* about the contents of S_0^*.

---

[11] In the exponential and logistic cases $P(\inf_{\tau \in T} F(\tau'\tilde{z}_{1t}) > 0) = 1$ trivially holds given $\tilde{z}_{1t}$ is bounded, cf. Assumption A.

[12] In which case $E[\epsilon_t r'\tilde{z}_{2t}(\delta)F(\tau'\tilde{z}_{1t})] = 0$ is true for by the martingale difference property for any r, δ and τ.

Indeed, in the scalar case we can go ever further. Let $\tilde{z}_{2t}(\delta) = 1$ (hence $\theta = \tau$). Then $S_0^* = \{\tau \in \mathbb{R}^{p_1 k+1} : V(\tau) = 0 \text{ and } P(\tau'\tilde{z}_{1t} \in R_0) = 1\}$ is identically the set S^* considered in Bierens (1991) and de Jong (1996)[13]. Likewise, $S(r)$ is simply the set S defined in Bierens (1991), de Jong (1996) and Stinchcombe and White (1998).

Corollary 5 *Define the set* $S_0^* = \{\tau \in \mathbb{R}^{p_2 k+1} : V(\tau) = 0$*, and* $P(\tau'\tilde{z}_{1t} \in R_0) = 1\}$*. Then under Assumptions A, B, D and E*

$$S_0^* \subseteq S = \{\tau \in \mathbb{R}^{p_1 k+1} : E[\epsilon_t F(\tau'\tilde{z}_{1t})] = 0\}, \text{ and } P(\tau'\tilde{z}_{1t} \in R_0) = 1\}. \quad (19)$$

Under H_1, any nuisance vector τ that renders $V(\tau) = 0$ must also foil the moment condition, $E[\epsilon_t F(\tau'\tilde{z}_{1t})] = 0$. Trivially, therefore, the set S_0^* has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^{p_1 k+1}$. This implies Lemma 2 of Bierens (1990) and Lemma 2 of de Jong are useful *only under the null*: under H_0 and Assumption B of Bierens (1990) or Assumption 3 of de Jong (1996), the set S_0^* has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^{p_1 k+1}$; however, under Assumptions *D-E* and H_1, $S_0^* \subseteq S$.

We can go further still for specifications $f_t(\phi)$ that are linear under H_0 and $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t} = \tilde{z}_{1t}$, both in terms of characterizing the *exact* contents set S_ξ^*, and characterizing the contents under *both* hypotheses. Notice that Assumption E is now superfluous provided $P(\inf_{\tau \in T} F_t(\tau) > 0) = 1$ by the remark following Assumption E.

Theorem 6 *Let* $f_t(\phi) = \phi'\tilde{z}_{1t}$ *and* $\tilde{z}_{2t}(\delta) = \tilde{z}_{1t}$ *(hence* $\theta = \tau$*). Let either Assumption E or* $P(\inf_{\tau \in T} F_t(\tau) > 0) = 1$ *hold. Under Assumptions A, B, and D the set* S_ξ^* *is empty for any* $\xi > 0$ *and any* $\delta \in \Delta$*. In particular,* $S_0^* = \{0\}$ *if* $0 \in R_0$*, and* S_0^* *is empty otherwise.*

Remark 1: If the null hypothesis is linearity and $\tilde{z}_{2t}(\delta) = \tilde{z}_{1t}$ then there does not exist a non-zero nuisance vector τ that renders $V(\theta)$ singular. This corresponds to a traditional STAR or AR-ANN framework where we test $y_t = \phi_0'\tilde{z}_{1t} + \epsilon_t$ against $y_t = \phi_0'\tilde{z}_{1t} + \phi_1'\tilde{z}_{2t}F_t(\tau) + \epsilon_t$. The reader can verify from the line of proof that the key argument is if there exists some $\tau \in S_0^*$, $\tau \neq 0$, then $F(\tau'\tilde{z}_{1t})$ cannot be "non-polynomial and analytic" on the interval on which $\tau'\tilde{z}_{1t}$ takes its values, therefore $\tau'\tilde{z}_{1t} \in R_0$ with probability zero, hence $\tau \in S_0^*$ is impossible.

Remark 2: Observe that we are able to prove the result without reverting to additional assumptions, a la Bierens' (1991) Assumption B and de Jong's (1996) Assumption 3. In the latter cases the set S_0^* is only shown to have Lebesgue measure zero.

Remark 3: The result can be extended to other specifications for $\tilde{z}_{2t}(\delta)$ and $f_t(\phi)$ under appropriate assumptions and modifications to the line of proof.

Remark 4: Because for any $\delta \in \Delta$ there do not exist points $|\tau| \geq \xi > 0$ that render $\lambda_0(V(\theta)) = 0$ for tests of linearity with $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t}$, Assumption C

---

[13] Note that Bierens (1990) and de Jong (1996) consider only the exponential weight $F(\cdot)$. Their results easily extend to the case where $F(\cdot)$ satisfies Assumption B.

is superfluous in this case and may be replaced with Assumption D in Theorems 2 and 3.

Remark 5: In a maximum likelihood setting the statistic ave-LM(θ) can be interpreted as the limit of a (Gaussian) weighted average power optimal test, where power is directed toward alternatives near the null: see Andrews and Ploberger (1994). Similarly, sup-LM(θ) directs power toward distant alternatives, but is only known to be asymptotically admissible (Andrews and Ploberger, 1995). In both cases the information matrix is required to be "uniformly positive definite" in the nuisance parameter space. Thus, tests of linear autoregression against a smooth transition alternative provide a natural setting for an application of Andrews and Ploberger's (1994, 1995) optimal tests.

6. Approximating the Limit Distribution

The statistics $\text{ave}_{\Theta_\xi} T_n(\theta)$ and $\sup_{\theta\in\Theta_\xi} T_n(\theta)$ have limit null distributions that depend upon the covariance function $V(\theta_1,\theta_2)$ and therefore upon the underlying distribution. In this section, we demonstrate that a simulation technique for approximating the asymptotic p-value, a la Hansen (1991, 1996), cf. Giné and Zinn (1990), is asymptotically valid in the present context.

6.1 Asymptotics

Let $\tilde{z}_n$ denote the sample $(\tilde{z}_{11},...,\tilde{z}_{1n})'$, and let $(v_t)_{t=1}^n$ be *iid* standard normal random variables. Define the following processes, letting $g(\cdot)$ denote any continuous function on Θ_ξ:

$$\begin{aligned}
&\hat{z}_n(\hat{\phi},0,\theta) = \hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi},0,\theta) \\
&\hat{S}_n(\hat{\phi},0,\theta) = 1/n\sum\nolimits_{t=1}^{n}\hat{\epsilon}_t\tilde{z}_{2t}(\delta)F_t(\tau)v_t \\
&\hat{Z}_n(\hat{\phi},0,\theta) = \hat{V}(\theta)^{-1/2}\sqrt{n}\hat{S}_n(\hat{\phi},0,\theta) \\
&\hat{T}_n(\theta) = \hat{Z}_n(\hat{\phi},0,\theta)'\hat{Z}_n(\hat{\phi},0,\theta) \\
&\hat{g}_n = g(\hat{T}_n(\theta)).
\end{aligned} \tag{20}$$

It is easy to simulate $\hat{S}_n(\hat{\phi},0,\theta)$ and compute a larger number (say J) of statistics $\hat{T}_n(\theta)$ by which an approximate p-value, $\hat{p}_n^J$, can be computed: see Section 6.2. In order to prove $\hat{p}_n^J$ converges to the true p-value under H_0, we must show $\hat{Z}_n(\hat{\phi},0,\theta)$ "converges weakly in probability" to the same Gaussian process to which $\hat{z}_n(\hat{\phi},0,\theta)$ converges, $z(\theta)$ defined in Theorem 2, cf. Giné and Zinn (1990).

Let $F_0(\cdot)$ denote the asymptotic null distribution of g_0, and define $p_0 \equiv 1 - F_0(g_0)$, and the asymptotic p-value $p_n \equiv 1 - F_0(g_n)$. Similarly, let $\hat{F}_n(\cdot)$ denote the conditional distribution of $\hat{g}_n$[14], and define $\hat{p}_n \equiv 1 - \hat{F}_n(g_n)$.

Denote by $\Rightarrow_p$ weak convergence in probability (see Giné and Zinn, 1990: Sections 2 and 3).

Theorem 7 *Let $\xi > 0$. Under Assumptions A-C, $\hat{Z}_n(\hat{\phi},0,\theta) \Rightarrow_p z(\theta)$. That is, cf. Giné and Zinn (1990: eq. 3.4),*

$$\sup\nolimits_{x\in\mathbb{R}} |P(\sup\nolimits_{\theta\in\Theta_\xi} |\hat{Z}_n(\hat{\phi},0,\theta)| \leq x|\tilde{z}_n) - P(\sup\nolimits_{\theta\in\Theta_\xi} |z(\theta)| \leq x)| \to 0. \tag{21}$$

---

[14] That is $\hat{F}_n(x) = P(\hat{g}_n \leq x|\tilde{z})$.

Remark 1: Assumption C may be replaced with Assumption D (and Assumption E or $P(\inf_{\tau \in T} F_t(\tau) > 0) = 1$) if the null is linearity and $\tilde{z}_{2t}(\delta) = \tilde{z}_{1t}$.

Remark 2: By appealing to (21) and the continuous mapping theorem, we deduce $\hat{T}_n(\theta) \Rightarrow_p T(\theta)$ and $\hat{g}_n \Rightarrow_p g(T(\theta)) = g_0$. We immediately conclude $\hat{F}_n(x) \to F_0(x)$ in probability uniformly in $x \in \mathbb{R}$, and therefore $\hat{p}_n - p_n = o_p(1)$. Because F_0 is a continuous function on the real line, $p_n \Rightarrow p_0$ under H_0 by Theorem 3 and the continuous mapping theorem, where p_0 is uniformly distributed on $[0, 1]$. Moreover, $p_n \to 0$ in probability under H_1 by Theorem 3 ($g_n \to \infty$ with probability one). Hence the asymptotic p-value satisfies $\hat{p}_n \Rightarrow p_0$ under H_0 and $\hat{p}_n \to 0$ under H_1.

6.2 Algorithm The p-value algorithm is identical to that of Hansen (1991, 1996), although Giné and Zinn (1984, 1990) detail a generic procedure. Generate a double array of *iid* standard normal variables, $[v_{t,j}]_{t=1,j=1}^{n,J}$. For each $j = 1...J$, compute $\hat{S}_{n,j}(\hat{\phi}, 0, \theta) = 1/n \sum_{t=1}^{n} \hat{\epsilon}_t \tilde{z}_{2t}(\delta) F_t(\tau) v_{t,j}$, $\hat{T}_{n,j}(\theta) = n\hat{S}_{n,j}(\hat{\phi}, 0, \theta)' \; \hat{V}(\theta)^{-1} \hat{S}_{n,j}(\hat{\phi}, 0, \theta)$, and $\hat{g}_{n,j} = g(\hat{T}_{n,j}(\theta))$. The approximate asymptotic p-value $\hat{p}_n^J$ is simply the sample frequency $J^{-1} \sum_{j=1}^{J} (g(\hat{T}_{n,j}(\theta)) > g(T_n(\theta)))$. Because the J-samples $[v_{t,j}]_{t=1}^{n}$ are independent of each other, an appeal to the Glivenko-Cantelli Theorem guarantees $\hat{p}_n^J \to \hat{p}_n$ as $J \to \infty$, and Theorem 7 guarantees $\hat{p}_n \Rightarrow p_0$ under H_0, etc.

7. Simulation Study We now investigate the empirical size and power properties of $\text{ave}_{\Theta_\xi} T_n(\theta)$ and $\sup_{\theta \in \Theta_\xi} T_n(\theta)$ under a null of linearity, and under STAR and bilinear alternatives. Let $\tilde{z}_{it} = (1, y_{t-1}, ..., y_{t-p_i})'$ for some $1 \leq p_2 \leq p_1$. Our simulations are based on the following models:

$$\begin{aligned}
&H_0 : y_t = \phi_0' \tilde{z}_{1t} + \epsilon_t \qquad (22)\\
&H_1^L : y_t = \phi_0' \tilde{z}_{1t} + \phi_1' \tilde{z}_{2t} [1 + \exp\{-\gamma' \tilde{z}_{1t}\}]^{-1} + \epsilon_t\\
&H_1^E : y_t = \phi_0' \tilde{z}_{1t} + \phi_1' \tilde{z}_{2t} \exp\left\{ -\sum\nolimits_{i=1}^{p_1} \gamma_i (\tilde{z}_{1t,i} - c_i)^2 \right\} + \epsilon_t\\
&H_1^{AN} : y_t = \phi_0' \tilde{z}_{1t} + \phi_1 [1 + \exp\{-\gamma' \tilde{z}_{1t}\}]^{-1} + \epsilon_t\\
&H_1^{SE} : y_t = \phi_0' \tilde{z}_{1t} + \phi_1' \tilde{z}_{2t} I(y_{t-1} > c_1) + \epsilon_t\\
&H_1^{BL} : y_t = \phi_0' \tilde{z}_{1t} + \phi_1 y_{t-1} \epsilon_{t-1} + \epsilon_t, \;\; |\phi_2| < 1
\end{aligned}$$

where ϵ_t are *iid* standard normal. Under H_0 the true data generating process is a linear autoregression; under H_1^L and H_1^E the true process is a 2-regime LSTAR and ESTAR, respectively; under H_1^{AN} the time series is governed by a logistic AR-ANN; under H_1^{SE} the process is a self-exciting autoregression (SETAR), equivalent to the LSTAR $y_t = \phi_0' \tilde{z}_{1t} + \phi_1' \tilde{z}_{2t} [1 + \exp\{-\gamma_1 (y_{t-1} - c_1)\}]^{-1} + \epsilon_t$ with $\gamma_1 \to \infty$; under H_1^{BL}, the process is bilinear.

7.1 Set-up We use sample sizes $n = 100$ and 500, generate $3n$ observations and retain the last n. For each simulated series we randomly select the orders $p_i \in \{1, ..., 10\}$, and the parameter vectors $\phi_i \in [-.95, .95]^{p_i}$, $\gamma \in [.5, 5]^{p_1}$, and $c \in [-.5, .5]^{p_1}$. Because we require the null model to be covariance stationary, only vectors ϕ_0 with characteristic roots outside the unit circle are considered.

We generate 1000 replications of each series above. A linear model is estimated and the resulting residuals are tested at the 5%-level. In order to specify the linear model, we employ a minimum AIC model selection criterion for the order p_1 over the integer set $\{1, ..., 10\}$[15].

7.2 General Tests Fix $p_1 = p_2 = p$. We set $\tilde{z}_t(\delta) = \tilde{z}_t$ and use $F_t(\tau) = [1 + \exp\{\tau' \tilde{z}_t\}]^{-1}$ and $F_t(\tau) = \exp\{\tau' \tilde{z}_t\}$. These are the STAR$_L$ and STAR$_E$ tests. The nuisance parameter space is $T = [.5, 5]^p$. For each STAR test the weight $F_t(\tau)$ is constructed from the standardized $\tilde{z}_t$: $F_t(\tau) = F(\sum_{i=1}^{p} \tau_i(\tilde{z}_{t,i} - \bar{\tilde{z}}_i)/s_i)$, where s_i denotes the sample standard deviation of $\tilde{z}_{t,i}$. The test functionals are computed over $[n/2]$ randomly selected nuisance parameters $\{\tau_i\}_{i=1}^{[n/2]} \in T$. The average statistic is computed with a uniform measure. Asymptotic p-values are computed according to Section 6.2 with $J = 1000$.

We perform the neural test of neglected nonlinearity (Lee *et al,* 1996), the Bierens test, the McLeod-Li test, the RESET test, and the polynomial regression test of Luukkonen *et al* (1988) and Teräsvirta (1994). The Bierens test is simply the STAR test with $\tilde{z}_{2t}(\delta) = 1$ (denoted BIER). The neural test is equivalent to a randomized Bierens test where τ is randomly selected from T.

For the STAR polynomial test, we estimate models of the form $y_t = \phi_0' \tilde{z}_t + \sum_{i=1}^{L} \vartheta_i' \tilde{z}_t y_{t-d}^i + u_t$ for $d = 1...p$. Under a null of linearity against an LSTAR (ESTAR) alternative, $L = 3$ (4) and $\vartheta_i = 0$, $i = 1..3$ (4). LM tests for each d is performed, and the test statistic with the smallest p-value based on the chi-squared distribution is selected. See Luukkonen *et al* (1988) and Teräsvirta (1994). These are the $POLY_L$ and $POLY_E$ tests, respectively. For the McLeod-Li test, we perform a standard portmanteau test on the squared null residuals for lags 1...3. For the RESET test, we follow the procedure detailed in Thursby and Schmidt (1977), and use three lags.

For all LM tests employed in this study, covariance matrix estimators robust to unknown forms of conditional heteroscedasticity are used.

7.3 Most Powerful Tests By appealing to the Neyman-Pearson lemma most powerful tests against STAR and ANN alternatives are easy to generate, and will help gauge the strength of the proposed STAR tests. Because ϕ_0 and $\sigma = 1$ are known, for any ϕ_0 and θ each STAR and ANN model in (22) can be represented as

$$y_t(\phi_0) = \phi_1' z_t(\theta) + \epsilon_t, \tag{23}$$

where $y_t(\phi_0) = y_t - \phi_0' \tilde{z}_{1t}$ and $z_t(\theta) = \tilde{z}_{2t}(\delta) F_t(\tau)$. For an arbitrary point (ϕ_0, θ) the least squares estimator of ϕ_1 is $\hat{\phi}_1(\theta, \phi_0) = (z(\theta)' z(\theta))^{-1} z(\theta)' y(\phi_0)$. The

---

[15] Simulations based on choosing the smallest order such that a Ljung-Box Q-test p-value is above 5% generates essentially identical results.

best test is simply the likelihood ratio, which in the present known standard normal setting reduces to

$$\begin{aligned}&\exp\{.5 \times \hat{\phi}_1(\theta, \phi_0)' \left(z(\theta)' z(\theta)\right) \hat{\phi}_1(\theta, \phi_0)\} \\ &\quad = \exp\{.5 \times y(\phi_0)' z(\theta) \left(z(\theta)' z(\theta)\right)^{-1} z(\theta)' y(\phi_0)\} \\ &\quad = \exp\{.5 \times T_n(\phi_0, \theta)\},\end{aligned} \tag{24}$$

say. Evaluated under the null, $T_n(\phi_0, \theta)$ is identically $T_n(\theta)$ when ϵ_t is iid, $\hat{\phi}$ is replaced by ϕ_0 and the covariance matrix $\hat{V}(\theta)$ is appropriately simplified given the known independence of ϵ_t. Thus, $\text{ave}_\theta\, T_n(\phi_0, \theta)$ and $\sup_{\theta \in \Theta_\xi} T_n(\phi_0, \theta)$ should provide better empirical size (by using the true ϕ_0) and at least as much power as the respective tests $\text{ave}_\theta\, T_n(\theta)$ and $\sup_{\theta \in \Theta_\xi} T_n(, \theta)$ against STAR and ANN alternatives. These are the "most powerful STAR" tests, denoted MPSTAR. Moreover, because the SETAR process is simply an LSTAR with $\tau = \infty$, the logistic sup-MPSTAR test (which directs power toward distant alternatives, cf. Andrews and Ploberger, 1994) should come close to a most powerful test against a SETAR alternative.

7.4 Results Test results under the null are contained in Table 1, and Table 2 contains empirical powers. We only display results for sup-tests[16]. For each test statistic empirical size is comparable to the nominal size.

In the present general environment in which all parameters are randomly selected, the popular polynomial regression test is dominated by every test (except the McLeod-Li test in some cases) against each alternative. Indeed, the STAR tests massively out-performs the conventional test. Impressively, against an LSTAR alternative with $n = 500$, the sup-STAR tests obtain *empirical power nearly identical to the most powerful* sup-MPSTAR tests (within .006), with a rejection rate above 90%. Similarly, the sup-STAR tests are comparable to most powerful MPSTAR tests against AR-ANN and SETAR alternatives (in particular, the sup-STAR test with exponential smooth transition weight).

Excluding the most powerful tests, the sup-STAR tests dominate every test against every alternative, except the McLeod-Li test against the bilinear alternative (which is not surprising)[17]. Finally, because the smooth transition vector moment condition includes the moment condition studied in Bierens (1990), it is not surprising that the sup-STAR tests out-perform the sup-BIER tests.

---

[16] The average STAR and BIER statistical functionals are qualitatively similar to the sup-tests in terms of power, and in terms of test performance relative to each other, and all other tests.

[17] Like the sup-STAR test, the ave-STAR test dominates all other tests, with power approaching that of the ave-BESTAR test against STAR, ANN and SETAR alternatives.

Appendix 1: Assumptions

Assumption A.1 The data-generating process $\{z_t\} = \{y_t, x_t\} \in \mathbb{R} \times \mathbb{R}^{k-1}$ exists on $L_2(\Omega, P, \mathfrak{F}_t)$ with $\mathfrak{F}_t$ a sequence of σ-algebras induced by (z_{t-i}), $i = 0, 1, ...$, such that $\mathfrak{F}_{t-1} \subset \mathfrak{F}_t$. The process $\{z_t\}$ is strictly stationary, ergodic, bounded, governed by a non-degenerate joint distribution function with non-degenerate marginal distributions, and for some $\kappa > 0$ and each $t \in \mathbb{Z}$, $\|z_t\|_{4+\kappa} < \infty$. The process $\{z_t\}$ is υ-stable in L_1 on an α-mixing base with coefficients $\sum_{i=1}^{\infty} \alpha_i < \infty$: see Bierens (1981, 1984, 1987, 1991, 1994).

Assumption A.2 The function $f : \mathbb{R}^{p_1 k+1} \times \Phi \to \mathbb{R}$ is for each $\tilde{z}_{1t} \in \mathbb{R}^{p_1 k+1}$ a continuous real function and twice continuously differentiable on Φ, a compact subset of $\mathbb{R}^m$. Moreover, $f(\tilde{z}_{1t}, \phi)$ is for each $\phi \in \Phi$ a Borel measurable function on $\mathbb{R}^{p_1 k+1}$. Define $\partial f_t(\phi) \equiv (\partial/\partial\phi) f_t(\phi)$. For each $t \in \mathbb{Z}$ the following bounds hold: $\left\|\sup_{\phi\in\Phi} |f_t(\phi)|\right\|_{4+\kappa} < \infty$, $\left\|\sup_{\phi\in\Phi} |\partial f_t(\phi)|\right\|_{4+\kappa} < \infty$, $\left\|\sup_{\phi\in\Phi} |\partial f_t(\phi)\partial' f_t(\phi)|\right\|_{4+\kappa} < \infty$, and $\left\|\sup_{\phi\in\Phi} |\partial\partial' f_t(\phi)|\right\|_{4+\kappa}$ for some $\kappa > 0$.

Let $\tilde{z}_{2t}(\delta)$ be a bounded mapping from $\mathbb{R}^{p_2 k+1} \times \Delta$ to $\mathbb{R}^{p_2 k+1}$, measurable with respect to $\mathfrak{F}_{t-1}$, such that $P(\inf_{\delta\in\Delta} |\tilde{z}_{2t}(\delta)| > a_0) = 1$ and $P(\sup_{\delta\in\Delta} |\tilde{z}_{2t}(\delta)| < a_1) = 1$ for some constants $0 < a_0 < a_1 < \infty$, and $\tilde{z}_{2t}(0) = \tilde{z}_{2t}$. Δ is a compact subset of $\mathbb{R}^{(p_2 k+1)q}$.

Assumption A.3 $\{F_t(\tau) = F(\tau' \tilde{z}_{1t})\}$ is a stationary, ergodic sequence of real-valued functions on $\mathbb{R}^{p_1 k+1} \times \mathbb{R}^{p_1 k+1}$ measurable with respect to $\Im_{t-1}$. Moreover, $F_t(0) = c$ with probability one for some finite constant $c \in \mathbb{R}$.

Assumption A.4 There exists a unique element $\phi_0 = \arg\inf_{\phi\in\Phi} E(y_t - f_t(\phi))^2$ where ϕ_0 is in the interior of Φ. Under either hypothesis $E[\epsilon_t (\partial/\partial\phi) f_t(\phi_0)] = E[\epsilon_t \tilde{z}_{1t}] = 0$, where $\tilde{z}_{1t} = (1, z'_{t-1}, ..., z'_{t-p_1})'$.

Assumption A.5 The matrix $A = E[\partial f_t(\phi_0) \partial f_t(\phi_0)]$ is positive definite and non-stochastic. Recall $\theta = (\delta', \tau')' \in \Theta$. The following uniform moment bounds hold for each t where in all cases $\kappa > 0$. Let T denote an arbitrary compact subset of $\mathbb{R}^{p_1 k+1}$, and let $\{M_i\}$ be a sequence of positive, finite constants:

$$\left\|\sup_{\theta\in\Theta} |F_t(\tau)^2 \tilde{z}_{2t}(\delta) \tilde{z}_{2t}(\delta)'|\right\|_{4+\kappa} \leq M_1$$

$$\left\|\sup_{\phi\in\Phi} \sup_{\tau\in T} |F_t(\tau) \partial f_t(\phi) \partial' f_t(\phi)|\right\|_{4+\kappa} \leq M_2$$

$$\left\|\sup_{\phi\in\Phi} \sup_{\theta\in\Theta} |F_t(\tau) \tilde{z}_{2t}(\delta) \partial' f_t(\phi)|\right\|_{4+\kappa} \leq M_3$$

$$\|\sup_{\theta\in\Theta} |F_t(\tau) \times \tilde{z}_{2t}(\delta)|\|_{4+\kappa} \leq M_4$$

$$\|\sup_{\theta\in\Theta} |(\partial/\partial\theta) F_t(\tau) \tilde{z}_{2t}(\delta)|\|_{4+\kappa} \leq M_5.$$

Appendix 2: Tables

Table 1: H_0

n	100	500	n	100	500
STAR^{*}_{L}	.057	.029	MPSTAR_L	.042	.016
STAR_E	.061	.029	MPSTAR_E	.024	.043
BIER_L	.011	.024	POLY_L	.001	.003
BIER_E	.040	.058	POLY_E	.001	.003
NEUR_L	.039	.046	ML-1**	.052	.070
NEUR_E	.039	.059	ML-2	.057	.088
RESET	.045	.038	ML-3	.064	.105

Notes: * Each STAR and BIER test is a sup-test.
** ML-h denotes the ML test with h-lags.

Table 2

n	100					500				
	H_1^L	H_1^E	H_1^{AN}	H_1^{SE}	H_1^{BL}	H_1^L	H_1^E	H_1^{AN}	H_1^{SE}	H_1^{BL}
STAR_L	.478	.303	.281	.387	.301	.912	.822	.803	.825	.766
STAR_E	.426	.336	.302	.325	.296	.915	.824	.872	.831	.729
MPSTAR_L	.635	.801	.466	.570	.165	.918	.901	.822	.980	.565
MPSTAR_E	.536	.698	.359	.421	.129	.921	.926	.881	.899	.572
BIER_L	.325	.325	.450	.213	.208	.722	.617	.784	.588	.624
BIER_E	.297	.376	.406	.211	.226	.786	.653	.759	.792	.624
NEUR_L	.357	.356	.365	.228	.211	.635	.606	.736	.589	.469
NEUR_E	.342	.392	.376	.261	.206	.622	.622	.777	.593	.479
POLY_L	.043	.004	.023	.002	.018	.433	.180	.019	.316	.029
POLY_E	.043	.004	.023	.002	.018	.433	.180	.019	.316	.029
RESET	.261	.021	.027	.112	.092	.411	.109	.039	.571	.006
ML-1	.113	.031	.182	.078	.516	.306	.065	.058	.254	.975
ML-2	.124	.050	.177	.094	.517	.366	.127	.058	.355	.987
ML-3	.151	.078	.163	.113	.524	.409	.151	.058	.371	.996

Appendix 3: Formal Proofs

Proof of Lemma 1. By Assumption A the proof follows directly from Stinchcombe and White (1994: Theorem 2.3). Indeed, if $F(u) = \exp\{u\}$, the result follows trivially from Lemma 1 of Bierens (1990). $\blacksquare$

Proof of Theorem 2. The result under H_0 follows from Lemmas A.1-A.6, below. Lemma A.1 proves consistency of the sample statistics in (12); Lemma A.2 proves $\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ and $V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta)$ have the same pointwise limiting distribution; Lemma A.3 demonstrates the finite distributions of the process $V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta)$ converge to normal distributions under the null hypothesis; Lemmas A.4 and A.5 prove $\sqrt{n}s_n(\phi_0, 0, \theta)$ is tight in $\mathbb{C}[\Theta]$, and $V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta)$ is tight in $\mathbb{C}[\Theta_\xi]$ by appealing to a general result due to Bierens and Ploberger (1997).

Finally, the result under H_1 follows from Lemma A.6. $\blacksquare$

Lemma A.1 *Under Assumptions A-B and under both H_0 and H_1,*

$$\left|\hat{A} - A\right| = o_p(1), \tag{25}$$

$$\sup_{\theta\in\Theta} |\hat{b}(\theta) - b(\theta)| = o_p(1),$$

$$\sup_{\theta\in\Theta} |\hat{V}(\theta) - V(\theta)| = o_p(1).$$

Lemma A.2 *Under Assumptions A-C and H_0,*

$$\sup_{\theta\in\Theta_\xi} |\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta) - V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta)| = o_p(1). \tag{26}$$

Remark 1: The result implies the random vectors $\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta)$ and $V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta)$ have the same pointwise limit distribution, cf. Theorem 3.1 Billingsley (1999).

Lemma A.3 *Under Assumptions A-C and H_0,*

$$V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta) \rightarrow N(0, I_{p_2 k+1}) \tag{27}$$

in distribution pointwise in Θ_ξ.

Lemma A.4 *Under Assumptions A-B and H_0, $\sqrt{n}s_n(\phi_0, 0, \theta)$ is tight on Θ.*

Lemma A.5 *Under Assumptions A-C and H_0, $V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta)$ is tight on Θ_ξ.*

Lemma A.6 *Under H_1 there exists a function $\eta : \Theta_\xi \rightarrow \mathbb{R}^{p_2 k+1}$ such that*

$$\sup_{\theta\in\Theta_\xi} \left|\hat{V}(\theta)^{-1/2}\hat{s}_n(\hat{\phi}, 0, \theta) - V(\theta)^{-1/2}\eta(\theta)\right| = o_p(1), \tag{28}$$

where $V(\theta)^{-1/2}\eta(\theta) \neq 0$ except for τ in a set S with Lebesgue measure zero.

Proof of Lemma A.1.

Step 1 ($|\hat{A} - A| = o_p(1)$): Under either hypothesis $\hat{\phi}$ is a consistent estimator of $\phi_0 \in \Phi$ under Assumption A and Theorem 8.2.2 of Bierens (1994). Thus, under either hypothesis $|\hat{A} - A| = o_p(1)$ follows from

$$\sup_{\phi \in \Theta} \left| n^{-1} \sum\nolimits_{t=1}^{n} \partial f_t(\phi) \partial' f_t(\phi) - E\left[\partial f_t(\phi) \partial' f_t(\phi)\right] \right| = o_p(1), \tag{29}$$

which follows easily from Assumption A and Bierens' (1991) Theorem 17 uniform law of large numbers for dependent, heterogenous processes.

Step 2 ($\sup_{\theta \in \Theta} |\hat{b}(\theta) - b(\theta)| = o_p(1)$): By the consistency of $\hat{\phi}$, $\sup_{\theta \in \Theta} |\hat{b}(\theta) - b(\theta)| = o_p(1)$ follows from the uniform bound

$$\sup_{\theta \in \Theta} \sup_{\phi \in \Phi} |b_n(\theta, \phi) - b(\theta)| = o_p(1), \tag{30}$$

where $b_n(\theta, \phi) \equiv n^{-1} \sum_{t=1}^{n} F_t(\tau) \tilde{z}_{2t}(\delta) \partial' f_t(\phi)$ and $b(\theta) \equiv E[F_t(\tau) \tilde{z}_{2t}(\delta) \partial' f_t(\phi)]$. The latter property follows upon application of Theorem 17 of Bierens (1991) by writing for each $i = 1...p_2 k + 1$, each $j = 1...m$, and any $\phi \in \Phi$,

$$n^{-1} \sum\nolimits_{t=1}^{n} \psi^{(i,j)}(\Gamma^{(i,j)}(\tilde{z}_t, \phi, \theta)) = n^{-1} \sum\nolimits_{t=1}^{n} F_t(\tau) \tilde{z}_{2t}(\delta) \partial' f_t(\phi) \tag{31}$$

where the real continuous functions $\Gamma^{(i,j)}(\cdot)$ are defined as

$$\begin{aligned} \Gamma^{(i,j)}(\tilde{z}_t, \phi, \theta) &= \left(\gamma_1^{(i,j)}(\tilde{z}_t, \phi, \theta), \gamma_2^{(i,j)}(\tilde{z}_t, \phi, \theta) \right)' \\ &= \left(F_t(\tau) \tilde{z}_{2t,i}(\delta),\ \partial_j f_t(\phi) \right)', \end{aligned} \tag{32}$$

and $\psi^{(i,j)}(\xi_1, \xi_2) = \xi_1 \times \xi_2$ for every i, j. We require $\sup_{|\xi| \le d} \left| (\partial / \partial \xi) \psi^{(i,j)}(\xi) \right| = O(d^{\mu})$ as $d \to \infty$, which easily follows by Assumption A for $\mu = 1$. Moreover, we need

$$\max_k \left\{ \sup_{\phi \in \Phi} \left| \gamma_k^{(i,j)}(\phi, \tilde{z}_{1t}) \right| \right\} \le \rho \bar{d}(\tilde{z}_{1t}) \tag{33}$$

for some non-negative continuous real function $\bar{d}(\tilde{z}_{1t})$ on $\mathbb{R}^{p_1 k + 1}$ such that for some $\kappa > 0$, $\sup_t E\left[\bar{d}(\tilde{z}_{1t})^{1+\mu+\kappa}\right] < \infty$, where μ is defined above, and $\rho = 1$[18]. Simply define $\bar{d}(\tilde{z}_{1t}) = \sup_{\phi \in \Phi} |(\partial / \partial \phi) f_t(\phi)|$ where continuity is ensured by Assumption A. Define

$$\bar{d}(\tilde{z}_{1t}) = \sup_{\theta \in \Theta} |F_t(\tau) \tilde{z}_{2t}(\delta)| + \sup_{\phi \in \Phi} |\partial f_t(\phi)|, \tag{34}$$

and $\sup_t E\left[\bar{d}(\tilde{z}_{1t})^{1+\mu+\kappa}\right] < \infty$ follows from Assumption A with $\mu = 1$.

Step 3 ($\sup_{\theta \in \Theta} |\hat{V}(\theta) - V(\theta)| = o_p(1)$): Using the fact that each $|\hat{\phi} - \phi_0|$, $\sup_{\theta \in \Theta} |\hat{b}(\theta) - b(\theta)|$ and $|\hat{A} - A|$ is $o_p(1)$, after some work we may write under either hypothesis $\hat{V}(\theta) = n^{-1} \sum_{t=1}^{n} \epsilon_t^2 g_t(\theta) g_t(\theta)' + o_p(1)$. Again, apply Theorem 17 of Bierens (1991) to each element $\epsilon_t^2 g_{t,i}(\theta) g_{t,j}(\theta)$, using the uniform bounds stated in Assumption A and derived in Lemma A.8 for $g_t(\theta)$. ■

---

[18] The identity $\rho = 1$ is trivial here because we do not include moving average terms: see Bierens (1991).

Proof of Lemma A.2. Using properties of the l_1-norm and Minkowski's inequality

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\hat{V}(\theta)^{-1/2}\sqrt{n}\hat{s}_n(\hat{\phi},0,\theta)-V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0,0,\theta)\right| \qquad (35)\\
&\quad\leq \sup_{\theta\in\Theta}\left|V(\theta)^{-1/2}\right|\sup_{\theta\in\Theta}\left|\sqrt{n}\hat{s}_n(\hat{\phi},0,\theta)-\sqrt{n}s_n(\phi_0,0,\theta)\right|\\
&\qquad+\sup_{\theta\in\Theta}\left|\hat{V}(\theta)^{-1/2}-V(\theta)^{-1/2}\right|\sup_{\theta\in\Theta}\left|\sqrt{n}s_n(\phi_0,0,\theta)\right|\\
&\qquad+\sup_{\theta\in\Theta}\left|\hat{V}(\theta)^{-1/2}-V(\theta)^{-1/2}\right|\sup_{\theta\in\Theta}\left|\sqrt{n}\hat{s}_n(\hat{\phi},0,\theta)-\sqrt{n}s_n(\phi_0,0,\theta)\right|,
\end{aligned}$$

By Lemma A.8, below, $\sup_{\theta\in\Theta}\left|V(\theta)^{-1/2}\right|$ is finite, and by Lemma A.1 and the Slutsky theorem $\sup_{\theta\in\Theta}|\hat{V}(\theta)^{-1/2}-V(\theta)^{-1/2}|=o_p(1)$. Thus $\sup_{\theta\in\Theta}|\hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2}|\sup_{\theta\in\Theta}|\sqrt{n}s_n(\phi_0,0,\theta)|=o_p(1)$ follows upon application of Lemmas A.1, A.3 and A.4: by Lemmas A.3 and A.4 we deduce $\sqrt{n}s_n(\phi_0,0,\theta)$ converges weakly to a Gaussian element $s(\theta)$ in $\mathbb{C}[\theta]$ with covariance function $V(\theta_1,\theta_2)$; hence by the continuous mapping theorem $\sup_{\theta\in\Theta}|\sqrt{n}s_n(\phi_0,0,\theta)|\Rightarrow\sup_{\theta\in\Theta}|s(\theta)|$; and the $o_p(1)$ rate follows from Lemma A.1 and Crámer's Theorem.

It remains to show $\sup_{\theta\in\Theta}|\sqrt{n}\hat{s}_n(\hat{\phi},0,\theta)-\sqrt{n}s_n(\phi_0,0,\theta)|$ is $o_p(1)$. By the mean-value-theorem, for some $\phi_*(\theta)\in\mathbb{R}^{p_1k+1}$ satisfying $|\phi_*(\theta)-\phi_0|\leq|\hat{\phi}-\phi_0|$, a.s.,

$$\sqrt{n}\hat{s}_n(\hat{\phi},0,\theta)-\sqrt{n}\hat{s}_n(\phi_0,0,\theta)=(\partial/\partial\phi)\hat{s}_n(\phi_*(\theta),0,\theta)\sqrt{n}(\hat{\phi}-\phi_0), \qquad (36)$$

where for any $\phi\in\Phi$

$$\begin{aligned}
(\partial/\partial\phi)\hat{s}_n(\phi,0,\theta) &= n^{-1}\sum\nolimits_{t=1}^{n}(\partial/\partial\phi)(y_t-f_t(\phi))\tilde{z}_{2t}(\delta)F_t(\tau) \qquad (37)\\
&= -n^{-1}\sum\nolimits_{t=1}^{n}\epsilon_t\tilde{z}_{2t}(\delta)\partial' f_t(\phi)F_t(\tau)\\
&= -b_n(\theta,\phi),
\end{aligned}$$

say. From the line of proof of Lemma A.1, cf. (30), $\sup_{\theta\in\Theta}\sup_{\phi\in\Phi}|b_n(\theta,\phi)-b(\theta,\phi)|=o_p(1)$ where $b(\theta,\phi)=E[\epsilon_t\tilde{z}_{2t}(\delta)\partial' f_t(\phi)F_t(\tau)]$. Because $|\phi_*(\theta)-\phi_0|\leq|\hat{\phi}-\phi_0|=o_p(1)$, using Assumption A it is easy to show $\sup_{\theta\in\Theta}|b_n(\theta,\phi_*(\theta))-b(\theta)|=o_p(1)$, where $b(\theta)=E[\epsilon_t\tilde{z}_{2t}(\delta)\partial' f_t(\phi_0)F_t(\tau)]$.

We deduce

$$\sup_{\theta\in\Theta}\left|\sqrt{n}\hat{s}_n(\hat{\phi},0,\theta)-\sqrt{n}\hat{s}_n(\phi_0,0,\theta)+b(\theta)\sqrt{n}(\hat{\phi}-\phi_0)\right|=o_p(1). \qquad (38)$$

Moreover, from standard nonlinear least squares algebra $|\sqrt{n}(\hat{\phi}-\phi_0)-\hat{A}^{-1}(1/\sqrt{n})\times\sum_{t=1}^{n}\partial f_t(\phi)\epsilon_t|=o_p(1)$, and $|\hat{A}-A|=o_p(1)$ by Lemma A.1. Hence,

$$\left|\sqrt{n}(\hat{\phi}-\phi_0)-A^{-1}(1/\sqrt{n})\sum\nolimits_{t=1}^{n}\partial f_t(\phi)\epsilon_t\right|=o_p(1). \qquad (39)$$

Substituting (39) into (38), and noting the identity

$$\sqrt{n}\hat{s}_n(\phi_0,0,\theta)-b(\theta)A^{-1}(1/\sqrt{n})\sum\nolimits_{t=1}^{n}\partial f_t(\phi)\epsilon_t=\sqrt{n}s_n(\phi_0,0,\theta), \qquad (40)$$

gives

$$\sup_{\theta\in\Theta} \left| \sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta) - \sqrt{n}s_n(\phi_0, 0, \theta) \right| = o_p(1). \tag{41}$$

■

Proof of Lemma A.3. For any $r \in \mathbb{R}^{p_2k+1}$ arbitrarily normalized to $r'r = 1$, define

$$w_t(\theta) = r'V(\theta)^{-1/2}g_t(\theta)\epsilon_t, \tag{42}$$

hence $\sqrt{n}r'V(\theta)^{-1/2}s_n(\phi_0, 0, \theta) = 1/\sqrt{n}\sum_{t=1}^n w_t(\theta)$. Under H_0, $1/\sqrt{n}\sum_{t=1}^n w_t(\theta) \rightarrow N(0,1)$ in distribution pointwise in Θ_ξ follows by a straightforward application of the Theorem 29 pointwise martingale difference central limit theorem of Bierens (1991), cf. McLeish (1974: Theorem of 2.3). Clearly $\{w_t(\theta)\}$ forms a martingale difference sequence under the null by Assumption A and the $\Im_{t-1}$-measurability of $g_t(\theta)$. The remaining conditions for the central limit theorem to apply follow from Lemma A.7, below. A Cramér-Wold device then suffices to establish (29), cf. Billingsley (1995: Theorem 29.4). ■

Lemma A.7 *Under the conditions of Lemma A.3, for each $\theta \in \Theta_\xi$*

$$\operatorname*{plim}_{n\rightarrow\infty} 1/n\sum_{t=1}^n w_t(\theta)^2 = \lim_{n\rightarrow\infty} 1/n\sum_{t=1}^n E[w_t(\theta)^2] = 1, \tag{43}$$

and for some $\kappa > 0$

$$\operatorname*{plim}_{n\rightarrow\infty} \sum_{t=1}^n E|w_t(\theta)/\sqrt{n}|^{2+\kappa} = 0. \tag{44}$$

Proof of Lemma A.7. By the normalization $r'r = 1$, and by stationarity and ergodicity for each t

$$\begin{aligned} E[w_t(\theta)^2] &= E\left[\epsilon_t r'V(\theta)^{-1/2}g_t(\theta)\right]^2 \\ &= r'V(\theta)^{-1/2}E\left[\epsilon_t^2 g_t(\theta)g_t(\theta)'\right]V(\theta)^{-1/2}r \\ &= r'V(\theta)^{-1/2}V(\theta)V(\theta)^{-1/2}r = r'r = 1. \end{aligned} \tag{45}$$

The weak limit $\operatorname{plim}_{n\rightarrow\infty} 1/n\sum_{t=1}^n w_t(\theta)^2 - 1 = 0$ then follows from

$$\sup_{\theta\in\Theta} \left| 1/n\sum_{t=1}^n \epsilon_t^2 g_t(\theta)g_t(\theta)' - V(\theta) \right| = o_p(1), \tag{46}$$

proved in Step 3 of the line of proof of Lemma A.1.

Limit (44) follows from the following bound: by Assumption A, standard l_1-norm inequalities and the envelope inequality[19], for some $\kappa > 0$

$$\begin{aligned} &E|w_t(\theta)|^{2+\kappa} \\ &\quad\le |r|^{2+\kappa}|V(\theta)^{-1/2}|^{2+\kappa}E|g_t(\theta)|^{2+\kappa} \\ &\quad\le |r|^{2+\kappa}\sup_{\theta\in\Theta_\xi}|V(\theta)^{-1/2}|^{2+\kappa}||\sup_{\theta\in\Theta_\xi}|g_t(\theta)|||_{2+\kappa}^{2+\kappa} \le M \end{aligned} \tag{47}$$

---

[19] Let $h : \mathbb{R}^{p_2k+1} \times \Theta$, and write $h_t(\theta)$. We refer to the property $E[h_t(\theta)] \le \sup_{\theta\in\Theta} E[h_t(\theta)] \le E[\sup_{\theta\in\Theta} h_t(\theta)]$ as the "envelope inequality", with $\sup_{\theta\in\Theta} h_t(\theta)$ the envelope of $h_t(\theta)$ for each t. See Theorem 21.3 of Davidson (1994). In (47), we use $h_t(\theta) = |g_t(\theta)| = \sum_{i=1}^{p_2k+1}|g_{t,i}(\theta)|$.

for some positive bounded M, where a bound on $\sup_{\theta\in\Theta_\xi}|V(\theta)^{-1/2}|$ and $||\sup_{\theta\in\Theta_\xi}|g_t(\theta)|||_{2+\kappa}$ are established in Lemma A.8, below[20]. Thus, $\sum_{t=1}^{r} E|w_t(\theta)|^{2+\kappa}/n^{1+\kappa/2} = o(1/n^{\kappa/2})$. ■

Proof of Lemma A.4. Define $\psi_t(\theta) = r'g_t(\theta)$ for any $r \in \mathbb{R}^{p_2k+1}$, $r'r = 1$. Thus $r'\sqrt{n}s_n(\phi_0, 0, \theta) = 1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t\psi_t(\theta)$. We will apply the tightness Lemma A.1 of Bierens and Ploberger (1997) to the sequence of random functions $1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t\psi_t(\theta)$ by verifying the sufficient conditions of that result. Clearly $\{\epsilon_t\psi_t(\theta)\}$ forms a martingale difference sequence under the null by Assumption A, with $\psi_t(\theta)$ measurable with respect to $\Im_{t-1}$, and $\theta \in \Theta$, where Θ is a compact Euclidean space with positive Lebesgue measure.

Define $K_t \equiv \sup_{\theta\in\Theta}|(\partial/\partial\theta)\psi_t(\theta)|$. It suffices to show

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E[\epsilon_t^2 {K_t}^2] < \infty, \quad \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E[\epsilon_t^2\psi_t(\theta_0)^2] < \infty, \tag{48}$$

where the latter inequality holds for at least one point $\theta_0 \in \Theta$. The first inequality follows from the Assumption A, standard l_1-norm inequalities the Cauchy-Schwartz inequality:

$$E\left[\epsilon_t^2 {K_t}^2\right] \leq \|\epsilon_t\|_4^2\, |r|^2\, \|\sup_{\theta\in\Theta}|(\partial/\partial\theta)g_t(\theta)|\|_4^2 \leq M < \infty, \tag{49}$$

where $||\sup_{\theta\in\Theta}|(\partial/\partial\theta)g_t(\theta)|\,||_4$ is bounded in Lemma A.8, below, and $|r|^2 < \infty$. The second inequality follows in a similar manner for any point $\theta \in \Theta$ by noting $||\sup_{\theta\in\Theta}|g_t(\theta)|||_4 \leq C$ for some positive constant C, cf. Lemma A.8. ■

Proof of Lemma A.5. Consider the normalized score: for any $r \in \mathbb{R}^{p_2k+1}$, $r'r = 1$, write

$$\begin{aligned} r'V(\theta)^{-1/2}\sqrt{n}s_n(\phi_0, 0, \theta) &= 1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t r'V(\theta)^{-1/2}g_t(\theta) \qquad (50)\\ &= 1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t r'\psi_t(\theta)\\ &= 1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t w_t(\theta), \end{aligned}$$

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

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E[\epsilon_t^2 {K_t}^2] < \infty, \quad \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E[\epsilon_t^2 w_t(\theta_0)^2] < \infty, \tag{51}$$

for at least one point $\theta_0 \in \Theta$, where we define $K_t = \sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta)w_t(\theta)|$. In this case, we must be careful to bound $V(\theta)^{-1/2}$ and $(\partial/\partial\theta)V(\theta)^{-1/2}$, a rather messy task. We place the laborious derivations in Lemma A.8, below.

---

[20] Trivially by Liaponov's inequality for some finite positive B, $|r|^{2+\kappa} \leq B|r|_2^{2+\kappa} = B(r'r)^{(2+\kappa)/2} = B < \infty$, given $r'r = 1$.

Moreover, it is uniquely here that the restrition $\xi > 0$ is imperative. Using $\xi = 0$ (i.e. $\Theta_0 = \Theta$), from Lemma A.8, below we obtain for some positive finite constant B, $\sup_{\theta\in\Theta}|V(\theta)^{-1/2}| \leq B(p_2k + 1)\sup_{\theta\in\Theta}[\lambda_0(V(\theta))^{-1}]$. The right-hand-side supremum does not exist if $0 \in \Theta$. Even if we "extend" the real line such that $\sup_{\theta\in\Theta}[\lambda_0(V(\theta))^{-1}] = +\infty$, a finite bound for (47) would still have to be established. Depending on the specification of $\tilde{z}_{2t}(\delta)$, $\lambda_0(V(\theta)) = 0$ for any $\theta = (\delta', 0')'$ is certainly possible. I would like to thank an anonymous referee for pointing out the issue of bounding $\xi > 0$.

The second inequality in (51) easily follows from Assumption A the Cauchy-Schwartz and envelope inequalities, and Lemma A.8: for any $\theta \in \Theta_\xi$

$$E[\epsilon_t^2 w_t(\theta)^2] \le \|\epsilon_t\|_4^2\,|r|^2|V(\theta)^{-1/2}|^2||\sup\nolimits_{\theta\in\Theta_\xi}|g_t(\theta)|||_4^2 \le M < \infty. \tag{52}$$

For the first inequality in (51), we will prove $E[\epsilon_t^2 {K_t}^2] \le M$ for some positive constant $M < \infty$ for all t. By the Cauchy-Schwartz inequality for all t

$$E[\epsilon_t^2 {K_t}^2] \le \|\epsilon_t\|_4^2\,|r|^2||\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta)\psi_t(\theta)|\,||_4^2. \tag{53}$$

The proof is complete upon bounding $||\sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta)\psi_t(\theta)|\,||_4^2$. The $(l,j)^{th}$-component $(\partial/\partial\theta_l)\psi_{t,j}(\theta)$ of the $s \times p_1k + 1$-matrix $(\partial/\partial\theta)\psi_t(\theta)$ ($s = q(p_2k + 1) + p_1k + 1$) is exactly

$$\begin{aligned}(\partial/\partial\theta_l)\psi_{t,j}(\theta) \quad = \quad & \sum\nolimits_{i=1}^{p_2k+1}(\partial/\partial\theta_l)V(\theta)_{j,i}^{-1/2} \times g_{t,i}(\theta) \\ & + \sum\nolimits_{i=1}^{p_2k+1} V(\theta)_{j,i}^{-1/2} \times (\partial/\partial\theta_l)g_{t,i}(\theta).\end{aligned} \tag{54}$$

Using (54) and Minkowski's inequality repeatedly, the moment $||\sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta)\psi_t(\theta)|\,||_4$ is bounded as

$$\begin{aligned}&\left\|\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta)\psi_t(\theta)|\right\|_4 \\ &\le \left\|\sum\nolimits_{l=1}^{s}\sum\nolimits_{j=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}\left|\sum\nolimits_{i=1}^{p_2k+1}(\partial/\partial\theta_l)V(\theta)_{j,i}^{-1/2} \times g_{t,i}(\theta)\right|\right\|_4 \\ &\quad + \left\|\sum\nolimits_{l=1}^{s}\sum\nolimits_{j=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}\left|\sum\nolimits_{i=1}^{p_2k+1}V(\theta)_{j,i}^{-1/2} \times (\partial/\partial\theta_l)g_{t,i}(\theta)\right|\right\|_4 \\ &\le (p_2k+1)\left\|\sum\nolimits_{l=1}^{s}\sup\nolimits_{\theta\in\Theta_\xi}\left|\sum\nolimits_{i,j=1}^{p_2k+1}\left|(\partial/\partial\theta_l)V(\theta)_{j,i}^{-1/2}\right|\sum\nolimits_{i=1}^{p_2k+1}|g_{t,i}(\theta)|\right|\right\|_4 \\ &\quad + s(p_2k+1)\left\|\sup\nolimits_{\theta\in\Theta_\xi}\left|\sum\nolimits_{i,j=1}^{p_2k+1}\left|V(\theta)_{j,i}^{-1/2}\right|\sum\nolimits_{l=1}^{s}\sum\nolimits_{i=1}^{p_2k+1}|(\partial/\partial\theta_l)g_{t,i}(\theta)|\right|\right\|_4 \\ &\le (p_2k+1)\sum\nolimits_{l=1}^{p_1k+1}\sup\nolimits_{\theta\in\Theta_\xi}\left|(\partial/\partial\theta_l)V(\theta)^{-1/2}\right|\left\|\sup\nolimits_{\theta\in\Theta_\xi}|g_t(\theta)|\right\|_4 \\ &\quad + s(p_2k+1)\sup\nolimits_{\theta\in\Theta_\xi}\left|V(\theta)^{-1/2}\right|\left\|\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta)g_t(\theta)|\right\|_4 .\end{aligned} \tag{55}$$

Lemma A.8 establishes bounds on $||\sup_{\theta\in\Theta_\xi}|g_t(\theta)|\,||_4$, $||\sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta)g_t(\theta)|\,||_4$, $\sup_{\theta\in\Theta_\xi}|V(\theta)^{-1/2}|$ and for each l, $\sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta_l)V(\theta)^{-1/2}|$, hence $||\sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta)\psi_t(\theta)|\,||_4$ $\le M$ for some finite M, which completes the proof. ■

Lemma A.8 *For some positive finite constants* B *and* C_i, $i = 1...7$, *and some*

$\kappa > 0$

i. $|A^{-1}|^2 \leq C_1,$ (56)
ii. $\sup_{\theta\in\Theta_\xi} |b(\theta)| \leq C_2,$
iii. $\sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta)b(\theta)| \leq C_3,$
iv. $|| \sup_{\theta\in\Theta_\xi} |g_t(\theta)| ||_{4+\kappa} \leq C_4,$
v. $|| \sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta)g_t(\theta)| \, ||_{4+\kappa} \leq C_5,$
vi. $\sup_{\theta\in\Theta_\xi} |V(\theta)^{-1/2}|^2 \leq B(p_2k+1)\left[\inf_{\theta\in\Theta_\xi} \lambda_0(V(\theta))\right]^{-1} \leq C_6$
vii. $\sup_{\theta\in\Theta_\xi} |(\partial/\partial\theta_l)V(\theta)^{-1/2}| \leq C_7, \; l = 1...p_1k+1$

Proof of Lemma A.8. Denote by $\lambda_i(X)$ the ordered eigenvalues of matrix $X \in \mathbb{R}^{m+1\times m+1}$: $\lambda_0(X) \leq ,..., \leq \lambda_m(X)$.

i. Because A is positive definite, each $0 < \lambda_i(A) < \infty$, hence by Liaponov's inequality for some B

$$\begin{aligned} |A^{-1}| &\leq |A^{-1/2}|^2 \leq B|A^{-1/2}|_2^2 \\ &= B \times Tr\left(A^{-1/2\prime}A^{-1/2}\right) = B \times Tr\left(A^{-1}\right) \\ &= B\sum\nolimits_i \lambda_i(A^{-1}) = B\sum\nolimits_i 1/\lambda_i(A) \leq C_1. \end{aligned} \tag{57}$$

ii. A bound on $\sup_{\theta\in\Theta_\xi} |b(\theta)|$ follows immediately from Assumption A and the envelope and Cauchy-Schwartz inequalities:

$$\begin{aligned} &\sup\nolimits_{\theta\in\Theta_\xi} |b(\theta)| \\ &\leq E\left[\sum\nolimits_{i,j} \sup\nolimits_{\theta\in\Theta_\xi} |F_t(\tau)\tilde{z}_{2t,i}(\delta)| \sup\nolimits_{\phi\in\Phi} |\partial_j f_t(\phi)|\right] \\ &\leq \left(E\left[\sum\nolimits_i \sup\nolimits_{\theta\in\Theta_\xi} |F_t(\tau)\tilde{z}_{2t,i}(\delta)|\right]^2\right)^{1/2} \left(E\left[\sum\nolimits_j \sup\nolimits_{\phi\in\Phi} |\partial_j f_t(\phi)|\right]^2\right)^{1/2} \\ &= \left\|\sup\nolimits_{\theta\in\Theta_\xi} |F_t(\tau)\tilde{z}_{2t}(\delta)|\right\|_2 \left\|\sup\nolimits_{\phi\in\Phi} |\partial f_t(\phi)|\right\|_2 \leq C_2. \end{aligned} \tag{58}$$

iii. Similarly, using Assumption A,

$$\begin{aligned} &\sup\nolimits_{\theta\in\Theta_\xi} |(\partial/\partial\theta)b(\theta)| \\ &\leq \left\|\sup\nolimits_{\theta\in\Theta_\xi} |(\partial/\partial\theta)F_t(\tau)\tilde{z}_{2t}(\delta)|\right\|_2 \left\|\sup\nolimits_{\phi\in\Phi} |\partial f_t(\phi)|\right\|_2 \leq C_3. \end{aligned} \tag{59}$$

iv. Next, by Assumption A, Minkowski's inequalities, and properties of the

l_1-norm, and (i) and (ii), for some $0 < B < \infty$

$$\begin{aligned}
&\left\|\sup\nolimits_{\theta\in\Theta_\xi}|g_t(\theta)|\right\|_{4+\kappa} \qquad (60)\\
&\quad \leq \left\|\sum\nolimits_{i=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}|F_t(\tau)\tilde{z}_{2t,i}(\delta)|\right\|_{4+\kappa}\\
&\qquad + \sup\nolimits_{\theta\in\Theta_\xi}|b(\theta)|\times|A^{-1}|\times\left\|\sum\nolimits_{i=1}^{p_2k+1}\sup\nolimits_{\phi\in\Phi}|\partial_i f_t(\phi)|\right\|_{4+\kappa}\\
&\quad = \left\|\sup\nolimits_{\theta\in\Theta_\xi}|F_t(\tau)\tilde{z}_{2t}(\delta)|\right\|_{4+\kappa}\\
&\qquad + \sup\nolimits_{\theta\in\Theta_\xi}|b(\theta)|\times|A^{-1}|\times\left\|\sup\nolimits_{\phi\in\Phi}|\partial f_t(\phi)|\right\|_{4+\kappa}\\
&\quad \leq C_6.
\end{aligned}$$

v. Similarly, by Assumption A, (i) and (iii)

$$\begin{aligned}
&\left\|\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta)g_t(\theta)|\right\|_{4+\kappa} \qquad (61)\\
&\quad \leq \left\|\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta)F_t(\tau)\tilde{z}_{2t}(\delta)|\right\|_{4+\kappa}\\
&\qquad + \sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta)b(\theta)|\times|A^{-1}|\times\left\|\sup\nolimits_{\phi\in\Phi}|\partial f_t(\phi)|\right\|_{4+\kappa}\\
&\quad \leq C_5.
\end{aligned}$$

vi. Next, by Liaponov's inequality for some B

$$\begin{aligned}
|V(\theta)^{-1/2}|^2 &\leq B|V(\theta)^{-1/2}|_2^2 = B\times Tr\left(V(\theta)^{-1/2\prime}V(\theta)^{-1/2}\right) \qquad (62)\\
&= B\times Tr\left(V(\theta)^{-1}\right) = B\times\sum\nolimits_{i=0}^{p_2k}\lambda_i(V(\theta)^{-1})\\
&= B\times\sum\nolimits_{i=0}^{p_2k}1/\lambda_i(V(\theta))\\
&\leq B\times(p_2k+1)\,\lambda_0(V(\theta))^{-1}
\end{aligned}$$

hence

$$\sup\nolimits_{\theta\in\Theta_\xi}|V(\theta)^{-1/2}|^2 \leq B\times(p_2k+1)\left[\inf\nolimits_{\theta\in\Theta_\xi}\lambda_0(V(\theta))\right]^{-1} \leq C_6, \qquad (63)$$

which is guaranteed for some finite C_6 by Assumption C: $\inf_{\theta\in\Theta_\xi}\lambda_0(V(\theta)) > 0$.

vii. Finally, consider $(\partial/\partial\theta_l)V(\theta)^{-1/2}$ for each $l = 1...p_1k+1$. By standard properties of matrix differentiation (e.g. Harville, 1997)

$$(\partial/\partial\theta_l)V(\theta)^{-1/2} = -(1/2)[V(\theta)^{-1/2}\times(\partial/\partial\theta_l)V(\theta)\times V(\theta)^{-1}]. \qquad (64)$$

Hence, for some positive constant B, by Liaponov's inequality and the derivations in (vi),

$$\begin{aligned}
&\sup\nolimits_{\theta\in\Theta_\xi}\left|(\partial/\partial\theta_l)V(\theta)^{-1/2}\right| \qquad (65)\\
&\quad \leq \sup\nolimits_{\theta\in\Theta_\xi}\left|V(\theta)^{-1/2}\times(\partial/\partial\theta_l)V(\theta)\times V(\theta)^{-1}\right|\\
&\quad \leq \sup\nolimits_{\theta\in\Theta_\xi}|V(\theta)^{-1/2}|^3|(\partial/\partial\theta_l)V(\theta)|\\
&\quad \leq B(p_2k+1)^{3/2}\left[\inf\nolimits_{\theta\in\Theta_\xi}\lambda_0(V(\theta))\right]^{-3/2}\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta_l)V(\theta)|
\end{aligned}$$

where $\inf_{\theta\in\Theta_\xi}\lambda_0(V(\theta)) > 0$ by Assumption C. The proof is complete when we show the l_1-normed $|(\partial/\partial\theta_l)V(\theta)|$ is uniformly bounded by some positive finite M.

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

$$\begin{aligned}(\partial/\partial\theta_l)V(\theta) &= (\partial/\partial\theta_l)E\left[\epsilon_t^2 g_t(\theta)g_t(\theta)'\right] \\ &= \left(E\left[\epsilon_t^2(\partial/\partial\theta_l)g_{t,i}(\theta)g_{t,j}(\theta)\right]\right)_{i,j} \\ &\quad + \left(E\left[\epsilon_t^2 g_{t,i}(\theta)(\partial/\partial\theta_l)g_{t,j}(\theta)\right]\right)_{i,j}.\end{aligned} \tag{66}$$

By the envelope and repeated Cauchy-Schwartz inequalities,

$$\begin{aligned}&\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta_l)V(\theta)| \\ &\le 2\sum\nolimits_{i,j=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}\left|E\left[\epsilon_t^2(\partial/\partial\theta_l)g_{t,i}(\theta)g_{t,j}(\theta)\right]\right| \\ &\le 2E\left[\epsilon_t^2\sum\nolimits_{i=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta_l)g_{t,i}(\theta)|\sum\nolimits_{i=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}|g_{t,i}(\theta)|\right] \\ &\le 2\left\|\epsilon_t\right\|_4^2\left\|\sum\nolimits_{i=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta_l)g_{t,i}(\theta)|\right\|_4\left\|\sum\nolimits_{i=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}|g_{t,i}(\theta)|\right\|_4 \\ &= 2\left\|\epsilon_t\right\|_4^2\left\|\sup\nolimits_{\theta\in\Theta_\xi}|(\partial/\partial\theta_l)g_t(\theta)|\right\|_4\left\|\sup\nolimits_{\theta\in\Theta_\xi}|g_t(\theta)|\right\|_4 \le M,\end{aligned} \tag{67}$$

where the last line follows from Assumption A, and (iv) and (v) above. ■

Proof of Lemma A.6. Using Minkowski's inequality and properties of the l_1-norm,

$$\begin{aligned}&\sup\nolimits_{\theta\in\Theta_\xi}\left|\hat{V}(\theta)^{-1/2}\hat{s}_n(\hat{\phi},0,\theta) - V(\theta)^{-1/2}\eta(\theta)\right| \\ &\le \left|V(\theta)^{-1/2}\right|\left|\hat{s}_n(\hat{\phi},0,\theta)-\eta(\theta)\right| + \left|\hat{V}(\theta)^{-1/2}-V(\theta)^{-1/2}\right||\eta(\theta)| \\ &\quad + \left|\hat{V}(\theta)^{-1/2}-V(\theta)^{-1/2}\right|\left|\hat{s}_n(\hat{\phi},0,\theta)-\eta(\theta)\right|.\end{aligned} \tag{68}$$

By the consistency of $\hat{\phi}$, the mean-value-theorem and Lemma A.1 (see Step 2 in the line of proof),

$$\begin{aligned}&\sup\nolimits_{\theta\in\Theta_\xi}\left|\hat{s}_n(\hat{\phi},0,\theta) - n^{-1}\sum\nolimits_{t=1}^n \epsilon_t\tilde{z}_{2t}(\delta)F_t(\tau)\right| \\ &\le \sup\nolimits_{\theta\in\Theta_\xi}\sup\nolimits_{\phi\in\Phi}|b_n(\theta,\phi)-b(\theta)|\,|\hat{\phi}-\phi_0| = o_p(1)\end{aligned} \tag{69}$$

where $b_n(\phi,\theta) = n^{-1}\sum_{t=1}^n \tilde{z}_{2t}(\delta)F_t(\tau)\partial' f_t(\phi)$. Notice,

$$\sup\nolimits_{\theta\in\Theta_\xi}\left|n^{-1}\sum\nolimits_{t=1}^n \epsilon_t\tilde{z}_{2t}(\delta)F_t(\tau) - \eta(\theta)\right| = o_p(1) \tag{70}$$

follows easily from Theorem 17 of Bierens (1991) and the bounds stated in Assumption A, where $\eta(\theta) = E\left[\epsilon_t\tilde{z}_{2t}F_t(\theta)\right]$. Therefore

$$\sup\nolimits_{\theta\in\Theta_\xi}\left|\hat{s}_n(\hat{\phi},0,\theta)-\eta(\theta)\right| = o_p(1), \tag{71}$$

implying the first term of the right-hand-side of (68) is $o_p(1)$. Moreover, because $\hat{V}(\theta)$ is consistent for $V(\theta)$, and $|\eta(\theta)| < \infty$ by Assumption A and the Cauchy-Schwartz inequality, the second and third terms are $o_p(1)$.

Finally, Lemma 1 guarantees for any δ the set $S = \{\tau \in \mathbb{R}^{p_1 k+1} : \eta(\theta) \neq 0$ and $P(\tau' \tilde{z}_{1t} \in R_0) = 1\}$ has Lebesgue measure and is nowhere dense in $\mathbb{R}^{p_1 k+1}$. Because $V(\theta)$ is uniformly positive definite in Θ_ξ, so are $V(\theta)^{-1}$ and $V(\theta)^{-1/2}$, hence $V(\theta)^{-1/2}\eta(\theta) \neq 0$ for every $\theta \in \Theta_\xi$ except $\tau \in S$. ■

Proof of Lemma 4. Consider any $\tau \in S_0^*$. If the null hypothesis is true then $E[\epsilon_t r' \tilde{z}_{2t}(\delta) F_t(\tau)] = 0$ is trivial for any $r \in \mathbb{R}^{p_2 k+1}$ and every $\theta \in \Theta$ by Assumption A. Hence, assume $P[E[\epsilon_t|\tilde{z}_{1t}] = 0] < 1$. Let $\delta \in \Delta$ be arbitrary.

From the Rayleigh-Ritz Theorem (e.g. Horn and Johnson, 1985: Theorem 4.2.2), for any $\tau \in S_0^*$,

$$\lambda_0(\theta) = \inf\nolimits_{r \in \mathbb{R}^{p_2 k+1} : r'r = 1} r' V(\theta) r = 0. \tag{72}$$

Therefore $\lambda_0(V(\theta)) = 0$ *if and only if* there exists some $r \in \mathbb{R}^{p_2 k+1}$, $r'r = 1$, such that $r'V(\theta)r = 0$, where r need not be unique.

For any such $r \in \mathbb{R}^{p_2 k+1}$ normalized to $r'r = 1$, $r'V(\theta)r = 0$ implies

$$r' g_t(\theta) g_t(\theta)' r E[\epsilon_t^2 | \tilde{z}_{1t}] = 0, \ a.s. \tag{73}$$

By Assumption D, $E[\epsilon_t^2|\tilde{z}_{1t}] \geq \varsigma$ with probability one for some constant $\varsigma > 0$, hence $r'V(\theta)r = 0$ *if and only if* $r'g_t(\theta) = 0$ for all t with probability one.

Separating $g_t(\theta)$, $\tau \in S_0^*$ implies for all t

$$\begin{aligned} & r' g_t(\theta) = 0, \ a.s., \\ & r' \tilde{z}_{2t}(\delta) F_t(\tau) - r' b(\theta) A^{-1} \partial f_t(\phi_0) = 0, \ a.s., \end{aligned} \tag{74}$$

hence

$$r' \tilde{z}_{2t}(\delta) F_t(\tau) = \alpha(r, \theta)' \partial f_t(\phi_0), \ a.s., \tag{75}$$

where $\alpha(r, \theta) \equiv A^{-1} b(\theta) r$.

Because $r'r = 1$, $r = 0$ is ruled out. Moreover, $E[\tilde{z}_{2t}(\delta)\tilde{z}_{2t}(\delta)']$ is uniformly positive definite in Δ by Assumption E. Thus, using Assumption A, $r'\tilde{z}_{2t}(\delta)$ is bounded and non-zero with probability one for any r, $r'r = 1$. Therefore $r'\tilde{z}_{2t}(\delta)$ satisfies the requirements of Lemma 1 (simply substitute $\tilde{z}_{2t}(\delta)$ for $r'\tilde{z}_{2t}(\delta)$), which, along with Assumption B, implies the set $S(r)$ has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^{p_1 k+1}$.

Moreover, by Assumption A, $E[\epsilon_t \partial f_t(\phi_0)] = 0$ under either hypothesis. By the almost sure identity in (78), we have for every $\tau \in S_0^*$,

$$E[\epsilon_t r' \tilde{z}_{2t}(\delta) F_t(\tau)] = \alpha(r, \theta)' E[\epsilon_t \partial f_t(\phi_0)] = 0, \tag{76}$$

hence $\tau \in S(r)$ for any $r \in \mathbb{R}^{p_2 k+1}$, $r'r = 1$, such that $r'V(\theta)r = 0$. This implies $S_0^* \subseteq S(r)$ for some r, as claimed. ■

Proof of Theorem 6. Recall $\partial f_t(\phi) = \tilde{z}_{1t}$, $\tilde{z}_{2t}(\delta) = \tilde{z}_{2t} = \tilde{z}_{1t} = \tilde{z}_t$, say (i.e. $\theta = \tau$ and $p_1 = p_2$), and

$$S_0^* = \{\tau \in \mathbb{R}^{p_1 k+1} : \lambda_0(V(\theta)) = 0 \text{ and } P(\tau' \tilde{z}_t \in R_0) = 1\}. \tag{77}$$

Define $\bar{R}_0$, the closed compliment interval of R_0: $F(u)$ is either polynomial or non-analytic on the domain $\bar{R}_0$ (clearly not both!).

Clearly $0 \in S_0^*$ provided $0 \in R_0$: $F(0)$ is a constant, $\tilde{z}_{2t}(\delta) = \tilde{z}_t$, hence $\sqrt{n}\hat{s}_n(\hat{\phi}, 0, 0) = 0$ by the first order conditions, implying $V(0) = 0$. If $0 \notin R_0$, then $0'\tilde{z}_t \in R_0$ with zero probability, hence $0 \notin S_0^*$.

Now assume there exists any other $\tau_0 \in S_0^*$, $\tau_0 \neq 0$. By Lemma A.9, below, we deduce every s^{th}-derivative evaluated at $\tau'\tilde{z}_t$ is zero with probability one: $(\partial/\partial u)^s F(u)|_{u=\tau_0'\tilde{z}_t} = 0$, *a.s.*, $\forall s \in \mathbb{N}$. In other words, for a given $\tau_0 \in S_0^*$, $\tau_0 \neq 0$, it must be the case that $P\left(\tau_0'\tilde{z}_t \in \bar{R}_0\right) = 1$. Therefore $F(\cdot)$ is not "*non-polynomial and analytic*" on the domain on which $\tau_0'\tilde{z}_t$ takes its values (Theorem 3.3 of Evgrafov, 1978). Because $F(\cdot)$ is non-polynomial and analytic on R_0 by Assumption A, we deduce $P(\tau_0'\tilde{z}_t \notin R_0) = 1$, and therefore $\tau_0 \notin S_0^*$, a contradiction of the assumption $\tau_0 \in S_0^*$.

Therefore $S_0^* = \{0\}$ if $0 \in R_0$, and S_0^* is empty otherwise. We deduce that S_ξ^* is empty for any $\xi > 0$ by noting $S_\xi^* \subseteq S_0^*$ by construction, and every $\theta = (\delta', \tau')' \in \Theta_\xi$ satisfies $|\tau| \geq \xi > 0$ hence $0 \notin S_\xi^*$. ■

Lemma A.9 *Under the conditions of Theorem 6, if there exists some $\tau \neq 0$ such that $\tau \in S_0^*$, then $(\partial/\partial u)^s F(u)|_{u=\tau'\tilde{z}_t} = 0$, a.s., for every $s \in \mathbb{N}$.*

Proof of Lemma A.9. Assume there exists a point $\tau \in S_0^*$ such that $\tau \neq 0$. By construction any $\tau \in S_0^*$ satisfies $\lambda_0(V(\theta)) = 0$. From the line of proof of Lemma 4, cf. (75), we deduce for some $r \in \mathbb{R}^{p_2 k+1}$, $r'r = 1$, (given $\partial f_t(\phi) = \tilde{z}_t$)

$$r'\tilde{z}_t F(\tau'\tilde{z}_t) = \alpha(r, \theta)'\tilde{z}_t, \text{ a.s.}, \tag{78}$$

where $\alpha(r, \theta) \equiv A^{-1}b(\theta)r$.

Differentiate and multiply both sides of (78) by $\tilde{z}_t$. Denoting by $F^s(\tau'\tilde{z}_t)$ the s^{th} derivative evaluated at $\tau'\tilde{z}_t$, $(\partial/\partial u)^s F(u)|_{u=\tau'\tilde{z}_t}$, we obtain

$$r'\tilde{z}_t F(\tau'\tilde{z}_t) + r'\tilde{z}_t F^1(\tau'\tilde{z}_t)\tau'\tilde{z}_t = \alpha(r, \theta)'\tilde{z}_t, \text{ a.s.} \tag{79}$$

From the identity $r'\tilde{z}_t F_t(\tau) = \alpha(r, \theta)'\tilde{z}_t$, *a.s.*, in (78), cancelling in (79) renders

$$r'\tilde{z}_t F^1(\tau'\tilde{z}_t)\tau'\tilde{z}_t = 0, \text{ a.s.} \tag{80}$$

Because $r'\tilde{z}_t \neq 0$ and $\tau'\tilde{z}_t \neq 0$ each with probability one due to $r \neq 0$, $\tau \neq 0$, and the non-singularity of A, (80) holds *if and only if*

$$F^1(\tau'\tilde{z}_t) = 0, \text{ a.s.} \tag{81}$$

Differentiate and multiply both sides of (81) by $\tilde{z}_t$: we obtain

$$F^2(\tau'\tilde{z}_t)\tau'\tilde{z}_t = 0, \text{ a.s.}, \tag{82}$$

hence $F^2(\tau'\tilde{z}_{1t}) = 0$, *a.s.*, due to $\tau \neq 0$, and so on. Repeating, for any $\tau \in S_0^*$, $\tau \neq 0$, and for every $s \in \mathbb{N}$

$$F^s(\tau'\tilde{z}_t) = (\partial/\partial u)^s F(u)|_{u=\tau'\tilde{z}_t} = 0, \text{ a.s.} \tag{83}$$

■

Proof of Theorem 7. From Theorem 2, $\hat{z}_n(\hat{\phi}, 0, \theta)$ converges weakly to the Gaussian element $z(\theta)$ on $\mathbb{C}[\Theta_\xi]$, with covariance function $E[z(\theta_1)z(\theta_2)']$ $= V(\theta_1)^{-1/2}V(\theta_1, \theta_2)V(\theta_2)^{-1/2}$. In order to show $\hat{Z}_n(\hat{\phi}, 0, \theta)$, conditioned on the sample $\tilde{z}_n = (\tilde{z}_{11}, ..., \tilde{z}_{1n})$, converges weakly to the same Gaussian element $z(\theta)$, we must demonstrate $\hat{Z}_n(\hat{\phi}, 0, \theta)$, conditioned on $\tilde{z}_n$, and $\hat{z}(\hat{\phi}, 0, \theta)$ converge pointwise to the same multivariate normal distribution; and $\hat{Z}_n(\hat{\phi}, 0, \theta)$, conditioned on $\tilde{z}_n$, is stochastically equicontinuous in $\mathbb{C}[\Theta_\xi]$, cf. Giné and Zinn (1990: Theorem 3.1).

Step 1 ($\hat{Z}_n(\hat{\phi}, 0, \theta) \rightarrow N(0, I_{p_2k+1})$): From now on operate conditionally on the sample $\tilde{z}_n$, and define

$$S_n(\theta) = 1/n\sum\nolimits_{t=1}^{n} \epsilon_t g_t(\theta)v_t, \quad Z_n(\theta) = V(\theta)^{-1/2}\sqrt{n}S_n(\phi_0, 0, \theta). \tag{84}$$

Using steps identical to the line of proof of Lemma A.2, noting that $\hat{\phi}$ and $\hat{V}(\theta)$ are consistent under either hypothesis, and recalling v_t is $(0,1)$-*iid*, it is easy to show

$$\sup\nolimits_{\theta\in\Theta_\xi} \left|\hat{Z}_n(\hat{\phi}, 0, \theta) - Z_n(\theta)\right| = o_p(1). \tag{85}$$

Thus, it suffices to consider only $Z_n(\theta)$ in the following.

Because v_t is $(0,1)$-*iid*, the covariance function of $Z_n(\theta)$, conditioned on $\tilde{z}_n$, is (recalling $V(\theta)^{-1/2}$ is symmetric)

$$\begin{aligned} &E[Z_n(\theta_1)Z_n(\theta_2)'|\tilde{z}_n] \\ &\quad = V(\theta_1)^{-1/2}1/n\sum\nolimits_{t=1}^{n} E[\epsilon_t^2 v_t^2 g_t(\theta_1)g_t(\theta_2)'|\tilde{z}_n]V(\theta_2)^{-1/2} \\ &\quad = V(\theta_1)^{-1/2}1/n\sum\nolimits_{t=1}^{n} \epsilon_t^2 g_t(\theta_1)g_t(\theta_2)'V(\theta_2)^{-1/2}. \end{aligned} \tag{86}$$

Using the bounds in Lemma A.8 and Theorem 17 of Bierens (1991), it is straightforward to show

$$\sup\nolimits_{(\theta_1,\theta_2)\in\Theta_\xi} \left|1/n\sum\nolimits_{t=1}^{n} \epsilon_t^2 g_t(\theta_1)g_t(\theta_2)' - V(\theta_1, \theta_2)\right| = o_p(1), \tag{87}$$

where $V(\theta_1, \theta_2) = E[\epsilon_t^2 g_t(\theta_1)g_t(\theta_2)']$. Thus, mimicking the line of proof of Lemma A.3 and recalling $v_t \overset{iid}{\sim} N(0,1)$, we obtain under H_0

$$Z_n(\theta) \rightarrow N(0, I_{p_2k+1}) \tag{88}$$

in distribution pointwise in Θ_ξ. This implies $Z_n(\theta)$ and $z_n(\phi_0, 0, \theta)$ have the same pointwise limit distribution because mean-zero multivariate normal distribution are fully characterized by their covariance functions. Therefore, from (85), (88) and Lemma A.2 $\hat{Z}_n(\hat{\phi}, 0, \theta)$ and $\hat{z}(\hat{\phi}, 0, \theta)$ have the same pointwise limit distribution.

Step 2 (Stochastic Equicontinuity in Θ_ξ**):** From Theorem 3.1 of Giné and Zinn (1990) and the Cauchy-Schwartz inequality it suffices to show

$$\lim_{\varpi\rightarrow 0} \limsup_{n\rightarrow\infty} P\left(E_v\left[\sup_{\theta_i\in\Theta_\xi, |\theta_1-\theta_2|<\varpi} |Z_n(\theta_1) - Z_n(\theta_2)|^2\right] > \varepsilon\right) = 0. \tag{89}$$

for all $\varepsilon > 0$, where E_v denotes the expectation only with respect to $(v_t)_{t=1}^{n}$, and

$$Z_n(\theta_1) - Z_n(\theta_2) = 1/\sqrt{n}\sum\nolimits_{t=1}^{n} \epsilon_t v_t \psi_t(\theta_1,\theta_2), \tag{90}$$

say, where $\psi_t(\theta_1,\theta_2) = V(\theta_1)^{-1/2}g_t(\theta_1) - V(\theta_2)^{-1/2}g_t(\theta_2)$. By appealing to Chebychev's inequality, in order to prove (89) it suffices first to bound $E_v[\sup_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|Z_n(\theta_1)-Z_n(\theta_2)|^2]$, and then bound $E(E_v[\sup_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|Z_n(\theta_1) - Z_n(\theta_2)|^2])$ by some $C \times \varpi$, $0 < C < \infty$.

Step 2.1 ($E_v[\sup_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|Z_n(\theta_1) - Z_n(\theta_2)|^2]$): From Liaponov's inequality, there exists some finite $B > 0$ such that

$$\begin{aligned}
&E_v\left[\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|Z_n(\theta_1)-Z_n(\theta_2)|^2\right] \\
&\quad\le B\times E_v\left[\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|Z_n(\theta_1)-Z_n(\theta_2)|_2^2\right] \\
&\quad= B\times E_v\left[\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}\sum\nolimits_{l=1}^{p_2k+1} n^{-1}\sum\nolimits_{s,t}^{n}\epsilon_s\epsilon_t v_s v_t \psi_{s,l}(\theta_1,\theta_2)\psi_{t,l}(\theta_1,\theta_2)\right] \\
&\quad\le Bn^{-1}\sum\nolimits_{t=1}^{n}\epsilon_t^2\sum\nolimits_{l=1}^{p_2k+1}\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}\psi_{t,l}(\theta_1,\theta_2)^2 \\
&\quad\le Bn^{-1}\sum\nolimits_{t=1}^{n}\epsilon_t^2\left(\sum\nolimits_{l=1}^{p_2k+1}\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}\left|\psi_{t,l}(\theta_1,\theta_2)\right|\right)^2 \\
&\quad= Bn^{-1}\sum\nolimits_{t=1}^{n}\epsilon_t^2\left(\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|\psi_t(\theta_1,\theta_2)|\right)^2
\end{aligned}\tag{91}$$

where the second inequality follows from $E_v[v_s v_t] = 0$, $\forall s \neq t$, and 1 otherwise.

Step 2.2 ($E(E_v[\sup_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|Z_n(\theta_1) - Z_n(\theta_2)|^2]) < \infty$): Using (91), the Cauchy-Schwartz inequality and stationarity,

$$\begin{aligned}
&E\left(E_v\left[\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|Z_n(\theta_1)-Z_n(\theta_2)|^2\right]\right) \\
&\quad\le Bn^{-1}\sum\nolimits_{t=1}^{n}E\left[\epsilon_t^2\left(\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|\psi_t(\theta_1,\theta_2)|\right)^2\right] \\
&\quad\le B\,\|\epsilon_t\|_4^2\,||\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|\psi_t(\theta_1,\theta_2)|||_4^2.
\end{aligned}\tag{92}$$

Now, by the mean-value-theorem and the definition of $\psi_t(\theta_1,\theta_2)$, for each ϖ

$$\begin{aligned}
&\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|\psi_t(\theta_1,\theta_2)| \\
&\quad\le \sum\nolimits_{l=1}^{p_2k+1}\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}\left|(\partial/\partial\theta_l)\left[V(\theta)^{-1/2}g_t(\theta)\right]\right|\times|\theta_{1,l}-\theta_{2,l}| \\
&\quad\le \sum\nolimits_{l=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}\left|(\partial/\partial\theta_l)\left[V(\theta)^{-1/2}g_t(\theta)\right]\right| \\
&\qquad\times\sum\nolimits_{l=1}^{p_2k+1}\sup\nolimits_{\theta_i\in\Theta_\xi,|\theta_1-\theta_2|<\varpi}|\theta_{1,l}-\theta_{2,l}| \\
&\quad\le \sum\nolimits_{l=1}^{p_2k+1}\sup\nolimits_{\theta\in\Theta_\xi}\left|(\partial/\partial\theta_l)\left[V(\theta)^{-1/2}g_t(\theta)\right]\right|\times\varpi.
\end{aligned}\tag{93}$$

Using the Minkowski inequality, it therefore suffices to prove

$$\lim_{\varpi\to 0}\sum\nolimits_{l=1}^{p_2k+1}\left\|\sup\nolimits_{\theta\in\Theta_\xi}\left|(\partial/\partial\theta_l)\left[V(\theta)^{-1/2}g_t(\theta)\right]\right|\right\|_4\times\varpi = 0, \tag{94}$$

which holds if $||\sup_{\theta\in\Theta_\xi}|(\partial/\partial\theta_l)[V(\theta)^{-1/2}g_t(\theta)]|||_4$ is bounded for each l, which is proved in Lemmas A.5 and A.8, cf. (54)-(55). ■

References

[1] Andrews, D.W.K., 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point, Econometrica 55, 1465-1471.

[2] Andrews, D.W.K. and W. Ploberger, 1994, Optimal Tests when a Nuisance Parameter is Present Only under the Alternative, Econometrica 82, 1383-1414.

[3] Andrews, D.W.K. and W. Ploberger, 1995, Admissibility of the Likelihood Ratio Test when a Nuisance Parameter is Present Only under the Alternative, The Annals of Statistics 23, 1609-1629.

[4] Bierens, H. J., 1981, Robust Methods and Asymptotic Theory in Nonlinear Econometrics (Springer Verlag: New York).

[5] Bierens, H. J., 1982, Consistent Model Specification Tests, Journal of Econometrics 20, 105-134.

[6] Bierens, H. J., 1984, Model Specification Testing of Time Series Regressions, Journal of Econometrics 26, 323-353.

[7] Bierens, H. J., 1987, ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands, Journal of Econometrics 35, 161-190.

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

[9] Bierens, H. J., 1991, Least Squares Estimation of Linear and Nonlinear ARMAX Models under Data Dependence, Annales d'Economie et de Statistique 20/21, 143-169.

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

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

[12] Billingsley, P., 1995 Probability and Measure (John Wiley & Sons: New York).

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

[14] Chan, K.S. and H. Tong, 1986a, On Estimating Thresholds in Autoregressive Models, Journal of Time Series Analysis 7, 179-190.

[15] Chan, K.S. and H. Tong, 1986b, On Tests for Nonlinearity in Time Series Analysis, Journal of Time Series Analysis 5, 217-228.

[16] Davidson, J., 1994, Stochastic Limit Theory (Oxford Univ. Press: Oxford).

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

[18] Davies, R.B., 1987, Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative, Biometrika 74, 33-43.

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

[20] van Dijk, D. and Franses, P.H., 1999, Modeling Multiple Regimes in the Business Cycle, Macroeconomic Dynamics 3, 311-340.

[21] van Dijk, D., T. Teräsvirta and P.H. Franses 2000, Smooth Transition Autoregressive Models-A Survey of Recent Developments, Econometric Institute Research Report EI2000-23/A, Erasmus University.

[22] Escribano, A. and O. Jordá, 1999, Improved Testing and Specification of Smooth Transition Autoregressive Models, in P. Rothman (ed.), Nonlinear Time Series Analysis of Economic and Financial Data, pp. 289-319 (Boston: Kluver).

[23] Evgrafov, M.A., 1978, Analytic Functions (Dover: New York).

[24] Franses, P.H. and D. van Dijk, 2000, Non-linear Time Series Models in Empirical Finance (Cambridge University Press).

[25] Gallant, A.R, and H. White, 1988, A Unified of Estimation and Inference for Nonlinear Dynamic Models (New York: Basil Blackwell).

[26] Giné, E., and J. Zinn, 1984, Annals of Probability 12, 929-989.

[27] Giné, E., and J. Zinn, 1990, Bootstrapping General Empirical Measures, The Annals of Probability 18, 851-869.

[28] Gonzalez-Rivera, G., 1998, Smooth-transition GARCH models, Studies in Nonlinear Dynamics and Econometrics 3, 61-78.

[29] Hagerud, G.E., 1997, A New Nonlinear GARCH Model, Unpublished Ph.D. thesis, IFE, Stockholm School of Economics.

[30] Hansen, B., 1991, Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis, Rochester Center for Economic Research Working Paper No. 296.

[31] Hansen, B., 1996, Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis, Econometrica 64, 413-430.

[32] Hansen, B., 1997, Inference in TAR Models, Studies in Nonlinear Dynamics 2, 1-14.

[33] Harville, D., 1997, Matrix Algebra from a Statistician's Perspective (Springer: New York).

[34] Holley, A., 1982, A Remark on Hausman's Specification Test, Econometrica 50, 749-760.

[35] Hong, Y. and H. White, 1995, Consistent Specification Testing via Nonparametric Series Regression, Econometrica 63, 1133-1159.

[36] Horn, R. and C.R. Johnson, 1985, Matrix Analysis (Cambridge Univ. Press: Cambridge).

[37] Keenan, D.M., 1985, A Tukey Nonadditivity-Type Test for Time Series Nonlinearity, Biometrika 72, 39-44.

[38] 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.

[39] Kuan, C. and H. White, 1994, Artificial Neural Networks: An Economic Perspective, Econometric Reviews 13, 1-91.

[40] Lee, B.J., 1988, A Nonparametric Model Specification Test Using a Kernel Regression Method, Ph.D. Dissertation, University of Wisconsin, Madison.

[41] Lee, T., H. White and C.W.J. Granger, 1993, Testing for Neglected Nonlinearity in Time-Series Models: A Comparison of Neural Network Methods and Alternative Tests, Journal of Econometrics 56, 269-290.

[42] Li, Q., C. Hsiao, and J. Zinn, 2003, Consistent Specification Tests for Semiparametric/Nonparametric Models Based on Series Estimation Methods, Journal of Econometrics 112, 295-325.

[43] Lin, J., and T. Teräsvirta, 1994, Testing the Constancy of Regression Parameters against Continuous Structural Change, Journal of Econometrics 62, 211-228.

[44] Lundberg, S., and T. Teräsvirta, 2002, Evaluating GARCH Models, Journal of Econometrics 110, 417-435.

[45] Lundberg, S., and T. Teräsvirta, 2005, A Time Series Model for an Exchange Rate in a Target Zone with Applications, Journal of Econometrics: in press.

[46] Lundberg, S., T. Teräsvirta, and D. van Dijik, 2003, Time-Varying Smooth Transition Autoregressive Models, Journal of Business and Economics Statistics 21, 104-121.

[47] Luukkonen, R., P. Saikkonen and T. Teräsvirta, 1988, Testing Linearity against Smooth Transition Autoregressive Models, Biometrika 75, 491-9.

[48] Medeiros, M.C. and A. Veiga, 2000, Diagnostic Checking in a Flexible Nonlinear Time Series Model, manuscript, Dept. of Electrical Engineering Catholic University of Rio de Janeiro.

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

[50] McLeish, D.L., 1975, A Maximal Inequality and Dependent Strong Laws, Annals of Probability 3, 829-839.

[51] Newey, W. K., 1985, Maximum Likelihood Specification Testing and Conditional Moment Tests, Econometrica 53, 1047-1070.

[52] Pollard, D., 1984, Convergence of Stochastic Processes (Springer-Verlag: New York).

[53] Pötscher, B. M., and I. R. Prucha, 1991, Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts, Econometric Reviews 10, 125-216.

[54] Ramsey, J.B., 1970, Models, Specification Error, and Inference: A Discussion of Some Problems in Econometric Methodology, Bulletin of the Oxford Institute of Economics and Statistics 32, 301-318.

[55] Ramsey, J.B., and P. Schmidt, 1976, Some Further Results on the Use of OLS and BLUE Residuals in Specification Error Tests, Journal of the American Statistical Association 71, 389-390.

[56] Rothman, P., D. van Dijk and P.H. Franses, 2001, A Multivariate STAR Analysis of the Relationship between Money and Output, Macroeconomic Dynamics 5, 506-532.

[57] Royden, H., 1988, Real Analysis (Prentice Hall: New York).

[58] Saikkonen, P. and R. Luukkonen, 1988, Lagrange Multiplies Tests for Testing Nonlinearities in Time Series Models, Scandinavian Journal of Statistics 15, 55-68.

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

[60] Stock, J.H., and M.W. Watson, 1989, Interpreting the Evidence on Money-Income Causality, Journal of Econometrics 40, 161-181.

[61] Teräsvirta, T., 1994, Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models, Journal of the American Statistical Association 89, 208-218.

[62] Teräsvirta, T., D. van Dijk, and M. Medeiros, 2004, Linear Models, Smooth Transition Autoregressions, and Neural Networks for Forecasting Macroeconomic Time Series: A Re-examination, mimeo, Department of Economics PUC-Rio, Brazil.

[63] Thursby, J., and P. Schmidt, 1977, Some Properties of Tests for Specification Error in a Linear Regression Model, Journal of the American Statistical Association 72, 635-641.

[64] Tong, H., 1990, Nonlinear Time Series: A Dynamical System Approach (Oxford, U.K.: Oxford University Press).

[65] White, H., 1989, An Additional Hidden Unit Test of Neglected Nonlinearity in Multilayer Feedforward Networks, in Proceedings of the International Joint Conference on Neural Networks, Washington D.C., Vol. 2 (IEEE Press: New York).

[66] Wooldridge, J., 1992, A Test for Functional Form against Nonparametric Alternatives, Econometric Theory 8, 452-475.

[67] Yatchew, A.J., 1992, Nonparametric Regression Tests Based on Least Squares, Econometric Theory 8, 435-451.

[68] Zheng, J., 1996, A Consistent Test of Functional Form via Nonparametric Estimation Techniques, Journal of Econometrics 75, 263-289.