

# Consistent and Non-Degenerate Model Specification Tests Against Smooth Transition and Neural Network Alternatives\*

Jonathan B. Hill<sup>†</sup>

**ABSTRACT.** – We develop a regression model specification test that directs maximal power toward smooth transition functional forms, and is consistent against *any* deviation from the null specification. We provide new details regarding whether consistent parametric tests of functional form are asymptotically degenerate: a test of linear autoregression against STAR alternatives is never degenerate. Moreover, a test of Exponential STAR has power attributes entirely associated with the choice of threshold. In a simulation experiment in which all parameters are randomly selected the proposed test has power nearly identical to a most-powerful test for true STAR, neural network and SETAR processes, and dominates popular tests. We apply the test to U.S. output, money, prices and interest rates.

---

## Tests de spécification de modèle convergents et non-dégénérés contre des alternatives avec transition douce et réseaux de neurones

**RÉSUMÉ.** – Nous développons un test de spécification pour un modèle de régression qui dirige une puissance maximale vers des formes fonctionnelles à transition douce et qui est convergent contre toute déviation par rapport à l'hypothèse nulle. Nous précisons également la dégénération asymptotique des tests paramétriques convergents de formes fonctionnelles : un test d'un processus autorégressif linéaire contre une alternative STAR n'est jamais dégénéré. De plus, un test d'une fonction exponentielle STAR possède une puissance entièrement associée au choix du seuil. Dans une simulation où tous les paramètres sont choisis aléatoirement, le test proposé possède une puissance presque identique à celle d'un test le plus puissant pour de vrais processus de type STAR, réseaux de neurones et SETAR, et domine tous les tests usuels. Nous appliquons ce test à des données américaines de niveau de production, de monnaie, de prix et de taux d'intérêts.

---

1 The comments and suggestions of two anonymous referees and Chief Editor Sébastien Roux are gratefully acknowledged. For comments on an earlier 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, and the 2004 North American Econometric Society meeting.

<sup>†</sup> Jonathan B. Hill : Dept. of Economics, University of North Carolina, Chapel Hill; jbhill@email.unc.edu; <http://www.unc.edu/~jbhill>. This paper was written while the author was affiliated with the Dept. of Economics at Florida International University. Their hospitality is gratefully acknowledged.

# 1 INTRODUCTION

---

## 1.1 STAR Methodologies

Smooth Transition Autoregression (STAR) models have gained significant popularity as a means to transcend well known explanatory and forecasting limitations of linear and binary regime switching models. See CHAN and TONG [1986a,b], TERASVIRTA [1994], LUUKKONEN *et al.* [1988], LIN and TERASVIRTA [1994], van DIJK *et al.* [2002], LUNDBERG *et al.* [2003] and LUNDBERG and TERASVIRTA [2005].

The standard setup models a time series  $\{y_t\}$  as a two-regime autoregression

$$(1) \quad y_t = \alpha + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^p \beta_i y_{t-i} \times F_{(y_{t-d}, \gamma, c)} + \varepsilon_t$$

where  $F$  is a smooth function taking values on  $[0, 1]$ , typically

$$\text{exponential: } F(y_{t-d}, \gamma, c) = \exp \{-\gamma(y_{t-d} - c)^2\}$$

$$\text{logistic: } F(y_{t-d}, \gamma, c) = \frac{1}{1 + \exp \{-\gamma(y_{t-d} - c)\}}.$$

The transition function  $F(y_{t-d}, \gamma, c) \in [0, 1]$  moves the data generating process between two linear regimes  $\alpha + \sum_{i=1}^p \phi_i y_{t-i}$  and  $\alpha + \sum_{i=1}^p (\phi_i + \beta_i) y_{t-i}$  based on past information  $y_{t-d}$ . The exponential  $F(\cdot)$  captures “inner” and “outer” regimes: if  $y_{t-d}$  is far from the threshold  $c$  then  $F(y_{t-d}, \gamma, c) \approx 0$  and  $y_t \approx \alpha + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t$ , and if  $y_{t-d}$  is close to  $c$  then  $F(y_{t-d}, \gamma, c) \approx 1$  and  $y_t \approx \alpha + \sum_{i=1}^p (\phi_i + \beta_i) y_{t-i} + \varepsilon_t$ . The logistic  $F(\cdot)$  captures “upper” and “lower” regimes: if  $c < y_{t-d} \rightarrow \infty$  then  $F(y_{t-d}, \gamma, c) \rightarrow 1$ , and  $F(y_{t-d}, \gamma, c) \rightarrow 0$  as  $c > y_{t-d} \rightarrow -\infty$ . The scale  $\gamma \geq 0$  gauges the speed of transition:  $\gamma = 0$  implies no transition in which case  $y_t$  is a linear AR; and small (large)  $\gamma > 0$  implies slow (fast) transition.

Tests of linearity against STAR alternatives, however, have received almost no attention in the theory literature, although a standard practice dominates the applied literature. Since a test of  $\beta = 0$  is not nuisance parameter-free the standard practice is to exploit  $\gamma = 0$ . The hypothesis is indirectly tested by performing a truncated Taylor approximation of  $F(y_{t-d}, \gamma, c)$  around  $\gamma = 0$ . This leads to a simple second or third order polynomial auxiliary regression in the spirit of RAMSEY [1970] and standard F-tests of parametric zero-restrictions are used to determine whether the process is linear AR, or exponential or logistic STAR. See LUUKKONEN *et al.* [1988], SAIKKONEN and LUUKKONEN [1988], LIN and TERASVIRTA [1994], TERASVIRTA [1994], GONZALEZ-RIVERA [1998], ESCRIBANO and JORDA [2000], ROTHMAN *et al.* [2001], LUNDBERG and TERASVIRTA [2002, 2005], and LUNDBERG *et al.* [2003], to name a few.

In order for the polynomial regression to have meaning in a STAR framework, however, the true data generating process is simply assumed to be a STAR. If no assumptions are made the test merely directs power toward low order polynomials. The test is therefore not a true test against smooth transition alternatives, per se.

The nuisance “delay” parameter  $d$  remains in the polynomial regression. If  $d$  is not simply assumed it is selected by minimizing the F-statistic  $p$ -value. The statistic has a non-standard limiting null distribution in the latter case (DAVIES [1977], STINCHCOMBE and WHITE [1992]), yet chi-squared or F- distributions are universally used. Similarly, in many instances the threshold  $c$  is simply fixed (e.g. GONZALEZ-RIVERA [1998]).

Finally, most smooth transition models in the applied literature incorporate only one threshold variable  $y_{t-d}$ , and in some cases only time  $t$  (LIN and TERASVIRTA [1994], VAN DIJK *et al.* [2000], LUNDBERG *et al.* [2003]). Test consistency will require *each stochastic variable* that enters into the null specification (e.g.  $y_{t-1}, \dots, y_{t-p}$ ) to enter into the weight function  $F(\cdot)$ , cf. BIERENS [1982, 1990] and STINCHCOMBE and WHITE [1998].

## 1.2 New STAR Test

In this paper we develop a consistent<sup>1</sup> parametric test of STAR functional form. Consistent parametric tests have been proposed by BIERENS [1990], BIERENS and PLOBERGER [1997], STINCHCOMBE and WHITE [1998], DETTE [1999] and HILL [2007]. See YATCHEW [1992], HARDLE and HALL [1993], HONG and WHITE [1996], STUTE [1997] and LI, HSIAO and ZINN [2003] for (semi) nonparametric methods.

Inconsistency arises because only a finite number of moment conditions are actually tested. A failure to reject the null may simply be due to the fact that some alternative *not* covered by the test statistic is true. In a STAR framework, even if we agree that finite-order polynomials adequately represents exponential and logistic functional forms, a failure to reject the test may be due to some *other* smooth transition mechanism (e.g. the Normal STAR: see CHAN and TONG [1986b]).

Our main contribution is a score test that directs power toward a general Smooth Transition Non-Linear Autoregression with Auxiliary variables (STARX). Single equation ARX models have a myriad applications in macroeconomics and finance (e.g. BAILLIE [1980]; BIERENS [1987, 1991]; PENA and SANCHEZ [2005]). The test is consistent against any deviation from the null, and nests specifications popularly employed in the STAR and Artificial Neural Network [ANN] literatures. Consult HORNIK, STINCHCOMBE and WHITE [1989], BIERENS [1990], HORNIK [1991] and LEE, WHITE and GRANGER [1996] for details on ANN models and their usage in economics. Whereas smooth transition models have simple behavioral interpretations<sup>2</sup>, neural nets are typically employed to absorb evident and otherwise unexplained nonlinearity (e.g. DONALDSON and KAMSTRA [1996]). A score test provides an intuitive sample check that smooth transition or neural net terms have not been omitted from a nonlinear ARX null specification.

- 
1. The power of the test statistic converges to one, as the sample size grows, under any deviation from the null.
  2. For example, as an exchange rate deviates from a target band, currency traders may expect open market transactions by a central bank to stabilize the rate. The planned transaction and its expectation by traders suggest traders may behave differently as the exchange rate increasingly deviates from the band.

Our test is consistent because we enforce  $\gamma > 0$ , permitting uncountably infinitely many moment conditions based on flexible test weights. We simply test whether the second regime term belongs

$$H_0 : \beta_i = 0, i = 1 \dots p \text{ vs. } H_1 : \text{at least one } \beta_i \neq 0,$$

and deliver a supremum test over  $\gamma$  in order to elevate small sample power.

In a second contribution we prove consistency of a test against an Exponential-STAR alternative is based on the threshold  $c$ . This suggests the practice of fixing  $c$  may curtail small sample test power.

Of separate interest, as a third contribution we prove a score test of linear autoregression against standard ANN or STAR alternatives is never degenerate except in a trivial case. This provides far more information concerning test degeneracy than previously characterized in BIERENS [1990] and DE JONG [1996], and provides a natural setting for the optimal tests of ANDREWS and PLOBERGER [1994, 1995] who simply assume non-degeneracy.

There are, however, some notable limitations. Although we permit non-stationary time series our test evidently cannot distinguish between non-stationarity (*e.g.* a unit root or stochastic break) and nonlinearity. See KAPETANIOS, SHIN and SNELL [2000] and KAPETANIOS and SHIN [2003] for tests in this genre. It also cannot handle some unbounded forms of global non-stationarity including linear trend in variance.

A simulation study demonstrates our test dominates standard tests, and vastly dominates the STAR polynomial regression test. In fact, the power of the proposed test against STAR, ANN and SETAR alternatives nearly matches that of uniformly most-powerful tests. Finally, we apply the test to a basket of U.S. macroeconomic variables.

In Section 2 we detail the STARX framework. Sections 3 and 4 contain the score statistic and construct smooth transition moment conditions. Asymptotic theory is developed in Section 5, and Section 6 characterizes test degeneracy. Sections 7 and 8 contain the simulation and empirical studies. Assumptions and proofs are in the appendices, and all tables are placed at the end.

Write  $\|x\|_p := (\sum_{i,j} |x_i, j|^p)^{1/p}$  and  $\|x\|_p := (\sum_{i,j} E|x_i, j|^p)^{1/p}$ . For arbitrary  $k$ -vectors  $a$  and  $x$ , vector powers  $x^a$  represent  $(x_1^{a_1}, \dots, x_k^{a_k})'$ .  $I_k$  denotes a  $k$ -dimensional identity matrix.  $\xrightarrow{p}$  denotes convergence in probability,  $\xrightarrow{d}$  convergence in finite dimensional distributions; and  $\Rightarrow$  weak convergence on a metric space;  $[x]$  is the integer part of  $x$ .  $\mathbb{C}[A]$  denotes the space of continuous functions endowed with the uniform metric on some compact space  $A$ .

## 2 STARX FRAMEWORK

---

Let  $\{W_t\} := \{y_t, x_t\}$  be a  $k$ -vector stochastic process, where  $x_t \in \mathbb{R}^{k-1}$ ,  $k \geq 1$  are regressors that do not contain lags of  $y_t$ . Assume  $\{W_t\}$  lies in  $L_2(\Upsilon, \mathfrak{F}, P)$  with probability measure  $P$  and  $\sigma$ -field  $\mathfrak{F}_t = \sigma(\{W_t\} : \tau \leq t)$ ,  $\mathfrak{F}_\infty = \sigma(\cup_{t \in \mathbb{Z}} \mathfrak{F}_t)$ . In the

case of a purely autoregressive framework  $k = 1$  and  $\mathfrak{I}_t = \sigma(y_\tau : \tau \leq t)$ . We assume  $x_t$  does not contain a constant, and the complete regressor set  $z_{t-1}$  contains lags of  $y_t$  and  $x_t$ :

$$z_{t-1} \equiv [(y_{t-1}, x'_{t-1})', \dots, (y_{t-p}, x'_{t-p})']' \text{ and } \tilde{z}_t \equiv (1, z'_t)' \in \mathbb{R}^{pk+1}.$$

## 2.1 STARX Model

Let  $B, D, \Gamma$  and  $\Phi$  be compact parameter spaces:

$$B \subset \mathbb{R}^l \quad (l \geq 1), \quad D \subset \mathbb{R}^q \quad (q \geq 0), \quad \Phi \subset \mathbb{R}^{pk+1}, \quad \Gamma \subset \mathbb{R}^{pk+1},$$

and consider known Borel-measurable response functions  $f$  and  $w$ :

$$f: \Phi \times \mathbb{R}^{pk+1} \rightarrow \mathbb{R} \text{ and } w: D \times \mathbb{R}^{pk+1} \rightarrow \mathbb{R}^l, \quad l \geq 1.$$

We are interested in whether the model

$$(2) \quad y_t = f(\phi, \tilde{z}_{t-1}) + \varepsilon_t$$

is correct for some  $\phi \in \Phi$  in the martingale difference sense  $E[\varepsilon_t | \mathfrak{I}_{t-1}] = 0$ , or whether a 2-regime smooth transition nonlinear ARX form

$$(3) \quad y_t = f(\phi, \tilde{z}_{t-1}) + \beta' w(\delta, \tilde{z}_{t-1}) \times F(\tau' \tilde{z}_{t-1}) + \varepsilon_t$$

improves the model fit, where

$$F: \mathbb{R} \rightarrow \mathbb{R}, \quad P(\inf_{\delta \in D} |w(\delta, \tilde{z}_{t-1})| > 0) = 1, \quad \beta \in B, \quad \delta \in D, \quad \tau \in \Gamma.$$

Traditionally  $F$  is the exponential or logistic restricted to  $[0, 1]$ , but we only require  $F$  to be non-polynomial and infinitely differentiable: see Section 4. The error term  $\varepsilon_t$  may be heteroscedastic. All regularity conditions are listed under Assumption A in Appendix B.

We use  $w(\delta, \tilde{z}_{t-1})$  with the imbedded parameter  $\delta$  to capture the ESTAR case, and bound  $|w(\delta, \tilde{z}_{t-1})| > 0$  to escape trivial or redundant cases (e.g.  $w(\delta, \tilde{z}_{t-1}) = \delta' \tilde{z}_{t-1}$ ). See Section 2.3, below, for examples. Model (3) nests (1) since  $f(\phi, \tilde{z}_{t-1}) = \phi' \tilde{z}_{t-1}$  and  $w(\delta, \tilde{z}_{t-1}) = \tilde{z}_{t-1}$  are special cases with  $\tilde{z}_{t-1} = (1, y_{t-1}, \dots, y_{t-p})$ .

It would be straightforward to permit different lags  $p_1$  and  $p_2$  in the two regimes, and to allow  $y_t$  and  $x_t$  to have different lags. Similarly, we could easily generalize  $\varepsilon_t$  to a finite-order moving average process producing a smooth transition ARMAX model (cf. DE JONG [1996]). Either generalization would only further complicate notation<sup>3</sup>.

3. Since none of the following theory requires  $p_1 = p_2 = p$ , we investigate  $p_1 \gtrless p_2$  in the simulation study of Section 7.

## 2.2 Persistence: $\nu$ -Stability

In order to have an accessible asymptotic theory applicable to heterogeneous nonlinear ARX data  $\{z_t\}$ , we utilize BIERENS' [1983, 1987, 1991, 1994] concept of  $\nu$ -stability on a strong mixing base. Consult Appendix A for a formal definition and properties, and see especially BIERENS [1991: *Annales d'Economie et de Statistique* 20/21].

Briefly,  $\nu$ -stability is essentially a version of Near-Epoch-Dependence and mixingale properties, and is equivalent to POTSCHER and PRUCHA'S [1991]  $L_p$ -approximability<sup>4</sup>. Under  $\nu$ -stability  $W_t$  can be an infinite order distributed lag in mean and/or variance with long or short memory, including ARFIMA( $p, d, q$ ) and/or FIGARCH( $p, d, q$ ),  $d \in [0, 1)$ , nonlinear difference equations  $W_t = h_t(\varepsilon_t, W_{t-1})$  with iid shocks  $\varepsilon_t$ , and bilinear, to name a few. Moreover, it covers mixing processes, in particular any strictly stationary geometrically ergodic process, including therefore Threshold Autoregressions, neural nets, Vector ARCH, STAR, nonlinear AR-GARCH processes, etc. (see, e.g. AN and HUANG [1996], NAJARIAN [2003], and MEITZ and SAIKKONEN [2008] *inter alia*).

The property does not characterize processes with a non-negligible infinite past (e.g. a unit root process), it encompasses seasonality, bounded trend in mean and variance, and stochastic breaks. In practice the analyst will need to pre-test for unbounded trend and unit roots and filter the series appropriately.

## 2.3 Examples

**LSTARX:** The Logistic-STARX model is

$$y_t = \phi' \tilde{z}_{t-1} + \beta' \tilde{z}_{t-1} \left[ 1 + \exp \left\{ - \sum_{i=1}^{pk} \gamma_i (\tilde{z}_{t-1,i} - c_i) \right\} \right]^{-1} + \varepsilon_t$$

$$= f(\phi, \tilde{z}_{t-1}) + \beta' w(\delta, \tilde{z}_{t-1}) \times F(\tau', \tilde{z}_{t-1}) + \varepsilon_t$$

where  $f(\phi, \tilde{z}_{t-1}) = \phi' \tilde{z}_{t-1}$ ,  $w(\delta, \tilde{z}_{t-1}) = \tilde{z}_{t-1}$ ,  $\gamma_i = -\tau_i \geq 0$  for  $i = 1 \dots pk$ ,  $c_i \in \mathbb{R}$ ,

and  $\tau_0 = \sum_{i=1}^{pk} \gamma_i c_i$ .

**ESTARX:** The Exponential-STARX model is

$$y_t = \phi' \tilde{z}_{t-1} + \beta' \tilde{z}_{t-1} \exp \left\{ - \sum_{i=1}^{pk} \gamma_i (\tilde{z}_{t-1,i} - c_i)^2 \right\} + \varepsilon_t.$$

This model is complicated by the quadratic transition mechanism. Since we ultimately require the test weight argument to be a *one-to-one function* of the transition variables  $\tilde{z}_{t-1}$  for test consistency, write

4. See GALLANT and WHITE [1988] and Davidson [1994]. Like NED, BIERENS'  $\nu$ -stability was originally inspired by the mixingale property (McLEISH [1975]). See BIERENS [1991].

$$\begin{aligned} \tilde{z}_{t-1} \exp \left\{ -\sum_{i=1}^{pk} \gamma_i (\tilde{z}_{t-1, i} - c_i)^2 \right\} &= \left( \tilde{z}_{t-1} \exp \left\{ \sum_{i=1}^{pk} \delta_i \tilde{z}_{t-1, i}^2 \right\} \right) \times \exp \{ \tau' \tilde{z}_{t-1} \} \\ &= w(\delta, \tilde{z}_{t-1}) \times F(\tau' \tilde{z}_{t-1}) \end{aligned}$$

say, where  $\delta_i = -\gamma_i < 0$  for  $i = 1 \dots pk$ ,  $c_i \in \mathbb{R}$ ,  $\{\tau_i = 2c_i \gamma_i\}_{i=1}^{pk}$  and  $\tau_0 = -\sum_{i=1}^{pk} \gamma_i c_i^2$ .

As long as  $\delta$  and  $\tau$  are treated as unrelated parameters (*i.e.* as long as the threshold  $c$  is not simply fixed) a consistent test is available based entirely on  $c$ , for any scale  $\gamma > 0$ . If  $c$  is fixed our proposed test cannot be proven to be consistent.

**ANN:** Since  $w(\delta, \tilde{z}_{t-1}) = 1$  is allowed under (3), a special case is a standard single-layer feed forward Artificial Neural Network. In the logistic case, for example,

$$y_t = \phi' \tilde{z}_{t-1} + \beta \times \left[ 1 + \exp \left\{ -\sum_{i=1}^{pk} \gamma_i (\tilde{z}_{t-1, i} - c_i) \right\} \right]^{-1} + \varepsilon_t.$$

Neural nets were popularized in the psychology and engineering literatures as purely non-theoretical means to efficiently approximate connections between data points. The most popular forms, the exponential and logistic, are universal approximators due to their infinite differentiability (HORNIK [1991]) making them highly useful objects for consistent test formation (BIERENS [1990], LEE *et al.* [1996], STINCHCOMBE and WHITE [1998]).

### 3 SCORE TEST OF STARX

---

Represent all nuisance parameters as

$$\theta \equiv [\delta', \tau']' \in \Theta = D \times \Gamma.$$

If the second regime  $w(\delta, \tilde{z}_{t-1})$  does not depend on  $\delta$  then  $\theta = \tau \in \Gamma$  (*e.g.*  $w(\delta, \tilde{z}_{t-1}) = \tilde{z}_{t-1}$  or  $w(\delta, \tilde{z}_{t-1}) = 1$ ).

Let  $\hat{s}_n(\phi, \beta, \theta)$  be the sample score associated with (3). If  $\hat{\phi}$  denotes the nonlinear least squares<sup>5</sup> estimator under  $H_0: \beta = 0$ , then

$$\hat{s}_n(\hat{\phi}, 0, \theta) = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t w(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1}) \in \mathbb{R}^l, \text{ where } \hat{\varepsilon}_t \equiv y_t - f(\hat{\phi}, \tilde{z}_{t-1}).$$

---

5. It is straightforward to extend all results to Generalized Method of Moments estimation.

By standard mean-value-theorem arguments an estimator of the asymptotic variance of  $\hat{s}_n(\hat{\phi}, 0, \theta)$  is

$$\hat{V}(\theta) = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2 \hat{g}_t(\theta) \hat{g}_t(\theta)' \in \mathbb{R}^{l \times l}$$

where

$$\hat{g}_t(\theta) = w(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1}) - \hat{b}(\hat{\phi}, \theta) \hat{A}(\hat{\phi})^{-1} \frac{\partial}{\partial \phi} f(\hat{\phi}, \tilde{z}_{t-1})$$

$$\hat{b}(\phi, \theta) = \frac{1}{n} \sum_{t=1}^n w(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1}) \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{t-1})$$

$$\hat{A}(\phi) = \frac{1}{n} \sum_{t=1}^n \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{t-1}) \frac{\partial}{\partial \phi'} f(\phi', \tilde{z}_{t-1}).$$

The score statistic under  $H_0$  is simply

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

We will show  $T_n(\theta) \xrightarrow{d} \chi^2(l)$  when model (2) is correct, for each point  $\theta$ , and  $T_n(\theta) \rightarrow \infty$  with probability one when (2) is not correct for uncountably infinitely many  $\theta$ . This is accomplished by considering (i) the ability of  $\hat{s}_n(\hat{\phi}, 0, \theta)$  to detect any deviation from the null (Section 4); (ii) whether  $T_n(\theta)$  converges on a space of continuous real functions (Section 5); and (iii) whether  $\hat{V}(\theta)$  converges to a singular matrix for certain points  $\theta \in \Theta$ , in which case  $T_n(\theta)$  is asymptotically degenerate (Section 6).

## 4 STARX CONDITIONAL MOMENTS

---

We need to show if  $\{\varepsilon_t, \mathfrak{F}_t\}$  in (2) is not a martingale difference sequence  $E[\varepsilon_t | \mathfrak{F}_{t-1}] \neq 0$ , then for any  $\delta \in D$

$$E[\varepsilon_t w(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1})] \neq 0 \text{ for "nearly every" } \tau \in \mathbb{R}^{pk+1}$$

We will make "nearly every" clear below. Lemma 1 is an easy, but required extension of Lemma 1 of BIERENS [1991], Theorem 1 of BIERENS and PLOBERGER [1997] and Theorem 2.3 of STINCHCOMBE and WHITE [1998].

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



Examples of analytic functions that are non-polynomial are  $\exp\{u\}$ ,  $[1 + \exp\{u\}]^{-1}$  and trigonometric functions<sup>6</sup>.

**LEMMA 1** *Let Assumption A apply, and  $P(E[\varepsilon_t | \mathfrak{F}_{t-1}] = 0) < 1$  where  $\mathfrak{F}_t = \sigma(\tilde{z}_\tau : \tau \leq t)$ . For each  $\delta \in D$  independent of  $\tau$ , any  $F$  under Assumption B, and any compact subset  $\Gamma \subset \mathbb{R}^{pk+1}$ , the set*

$$S = \bigcap_{i=1}^l \{ \tau \in \Gamma : E[\varepsilon_t w_i(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1})] = 0 \} \text{ and } P(\tau' \tilde{z}_{t-1} \in R_0) = 1$$

has Lebesgue measure zero and is nowhere dense in  $\mathbb{R}^{pk+1}$ .

*Remark:*  $S$  contains those  $\tau$  that render an asymptotically faulty score test  $T_n(\theta)$  since  $E[\varepsilon_t w_i(\delta, \tilde{z}_{t-1}) F(\tau', \tilde{z}_{t-1})] = 0$  even when  $E[\varepsilon_t | \mathfrak{F}_{t-1}] \neq 0$  with positive probability. Although there may be infinitely many such “bad” nuisance parameters  $\tau$ , Lebesgue measure zero means there can be at most countably<sup>7</sup> many of them. This means a consistent STARX test  $T_n(\theta)$  can be constructed simply by randomly selecting all nuisance parameters  $\theta = [\delta', \tau']'$  from any subset  $\Theta$ , or by computing the supremum of  $T_n(\theta)$  over compact  $\Theta$ .

Recall the ESTARX model from Section 2.3. If  $f(\phi, \tilde{z}_{t-1})$  is mis-specified then

$$\begin{aligned} & E \left[ \varepsilon_t \tilde{z}_{t-1} \exp \left\{ - \sum_{i=1}^{pk} \gamma_i (\tilde{z}_{t-1,i} - c_i)^2 \right\} \right] \\ &= E \left[ \varepsilon_t \left( \tilde{z}_{t-1} \exp \left\{ \sum_{i=1}^{pk} \delta_i \tilde{z}_{t-1,i}^2 \right\} \right) \times \exp \{ \tau' \tilde{z}_{t-1} \} \right] \neq 0 \end{aligned}$$

for any scale  $\gamma_i = -\delta_i > 0$  and uncountably infinitely many  $\tau \in \mathbb{R}^{pk+1}$ , hence uncountably infinitely many  $\{c_i = \tau_i / 2\gamma_i\}_{i=1}^{pk} \in \mathbb{R}^{pk}$ . Since  $\delta_i = -\gamma_i$  and  $\tau_i = 2c_i\gamma_i$ , and  $\delta_i$  and  $\tau_i$  must be treated as separate, the ability of the ESTARX moment condition to reveal model mis-specification is therefore *solely associated with the threshold  $c$* .

**COROLLARY 2 (ESTARX)** *Under the conditions of Lemma 1, if  $P(E[\varepsilon_t | \mathfrak{F}_{t-1}] = 0) < 1$  then for each  $\gamma > 0$  the set*

$$\left\{ c \in \mathbb{R}^{pk} : E \left[ \varepsilon_t \tilde{z}_{t-1} \exp \left\{ - \sum_{i=1}^{pk} \gamma_i (\tilde{z}_{t-1,i} - c_i)^2 Y_i \right\} \right] = 0, P(\tau' \tilde{z}_{t-1} \in R_0) = 1 \right\}$$

6. Lemma 1 is grounded on Theorem 2.3 of STINCHCOMBE and WHITE [1998]. However, they show (see their Corollary 3.9) that the analytic property can be relaxed, allowing  $F$  to be a normal cumulative distribution function. This supports the Normal STAR model of CHAN and TONG [1986b].

7. Any two  $\{\tau_1, \tau_2\} \in S$  are not “neighbors”:  $\inf\{\|\tau_1 - \tau_2\| : \tau_1, \tau_2 \in S\} > 0$ . Cf. BIERENS [1990: Lemma 1].

has Lebesgue measure zero and is nowhere dense in  $\mathbb{R}^{pk}$ , where  $\{\tau_i = 2\gamma_i c_i\}_{i=1}^{pk}$  and  $\tau_0 = -\sum_{i=1}^{pk} \gamma_i c_i^2$ .

## 5 STARX TEST THEORY

---

In this section we derive the weak limit distribution of the STARX score test statistic  $T_n(\theta)$ . A weak limit is required since  $\theta$  is unknown (see BILLINGSLEY [1999]). Define

$$(4) \quad s_n(\phi, 0, \theta) \equiv \frac{1}{n} \sum_{t=1}^n \varepsilon_t g_t(\theta), \text{ where } \varepsilon_t \equiv y_t - f(\phi, \tilde{z}_{t-1})$$

and

$$(5) \quad g_t(\theta) \equiv w(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1}) - b(\theta, \phi) A(\phi)^{-1} \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{t-1})$$

$$b(\phi, \theta) \equiv E \left[ F(\tau' \tilde{z}_{t-1}) w(\delta, \tilde{z}_{t-1}) \frac{\partial}{\partial \phi'} f(\phi, \tilde{z}_{t-1}) \right]$$

$$A(\phi) \equiv E \left[ \frac{\partial}{\partial \phi} (\phi, \tilde{z}_{t-1}) \frac{\partial}{\partial \phi'} f(\phi, \tilde{z}_{t-1}) \right] \text{ and } V(\theta) \equiv E[\varepsilon_t^2 g_t(\theta) g_t(\theta)'].$$

In Lemma A.1 of Appendix D it is shown

$$\sup_{\theta \in \Theta} |\hat{V}(\theta)^{-1/2} \sqrt{n} \hat{s}_n(\hat{\phi}, 0, \theta) - V(\theta)^{-1/2} \sqrt{n} s_n(\phi, 0, \theta)| \xrightarrow{p} 0.$$

We therefore need only consider  $V(\theta)^{-1/2} \sqrt{n} s_n(\phi, 0, \theta)$ .

In order to optimize small sample power we propose a supremum or average score statistics over  $\Theta$ . This requires treating  $V(\theta)^{-1/2} \sqrt{n} s_n(\phi, 0, \theta)$  as a random function of  $\theta$ . We will show the vector  $V(\theta)^{-1/2} \sqrt{n} s_n(\phi, 0, \theta)$  converges weakly to a Gaussian element of the space of continuous functions  $C[\Theta]$ . Gaussian elements of  $C[\Theta]$  are completely characterized by their mean and covariance functions  $V(\cdot, \cdot)$ , the latter defined for our purposes as

$$V(\theta_1, \theta_2) \equiv E[\varepsilon_t^2 g_t(\theta_1) g_t(\theta_2)'].$$

Notice  $V(\theta) = V(\theta, \theta)$  where  $V(\theta)$  is in (5). See ROYDEN [1968] and BILLINGSLEY [1999].

We invoke for now the following assumption to ensure  $V(\theta)^{-1/2}$  exists. For arbitrary  $\xi > 0$  define the compact subspace

$$\Theta_\xi = \{\theta = [\delta', \tau'] \in \Theta : |\tau| \geq \xi\}.$$

By convention  $\Theta_0 = \Theta$ . Bounding  $|\tau| \geq \xi > 0$  is required in order to demonstrate tightness of the distributions governing  $\{V(\theta)^{-1/2} \sqrt{ns_n}(\phi, 0, \theta)\}_{n \in \mathbb{N}}$  by ensuring  $V(\theta)^{-1/2}$  is not “nearly” singular. If the reader wants to randomize  $\theta$  then tightness arguments can be safely ignored. Denote by  $\lambda_{\min}(V(\theta))$  the minimum eigenvalue of  $V(\theta)$ .

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

## 5.1 Weak Convergence

The null hypothesis in its most general form is simply  $f(\phi, \tilde{z}_{t-1})$  is a version of  $E[y_t | \mathfrak{F}_{t-1}]$ . In the framework of (3) this translates to  $\beta = 0$ . In general,

$$H_0 : P(E[y_t - f(\phi, \tilde{z}_{t-1}) | \mathfrak{F}_{t-1}] = 0) = 1, \text{ for some } \phi \in \Phi$$

$$H_1 : \sup_{\phi \in \Phi} P(E[y_t - f(\phi, \tilde{z}_{t-1}) | \mathfrak{F}_{t-1}] = 0) < 1.$$

The general alternative  $H_1$  embraces *any* deviation from the null, and not just (3) with  $\beta \neq 0$ . Those interested only in STARX models will not distinguish the two: the hypotheses are  $H_0 : \beta = 0$  and  $H_1 : \beta \neq 0$ . The point here is that irrespective of whether the analyst is interested only in smooth transition models (3), the alternative is any deviation from  $E[y_t | \mathfrak{F}_{t-1}] = f(\phi, \tilde{z}_{t-1})$ , where  $f(\phi, \tilde{z}_{t-1}) + \beta' w(\delta, \tilde{z}_{t-1}) F(\tau \tilde{z}_{t-1})$  is guaranteed to provide a better fit.

### THEOREM 3

*i. Under  $H_0$  and Assumptions A and C there exists an  $l$ -vector Gaussian element  $z(\theta)$  of  $\mathbb{C}[\Theta_\xi]$  with covariance function  $E[z(\theta_1)z(\theta_2)'] = V(\theta_1)^{-1/2} V(\theta_1, \theta_2) V(\theta_2)^{-1/2}$  satisfying*

$$V(\theta)^{-1/2} \sqrt{ns_n}(\phi, 0, \theta) \Rightarrow z(\theta)$$

*ii. Under Assumptions A-C and  $H_1$ , there exists a non-stochastic vector function  $\eta : \Theta_\xi \rightarrow \mathbb{R}^{pk+1}$  with the property  $V(\theta)^{-1/2} \eta(\theta) \neq 0$  for all  $\theta = [\delta', \tau'] \in \Theta_\xi$  except possibly for  $\tau$  in a set  $S$  with Lebesgue measure zero, such that*

$$\sup_{\theta \in \Theta_\xi} \left| V(\theta)^{-1/2} s_n(\phi, 0, \theta) - V(\theta)^{-1/2} \eta(\theta) \right| \xrightarrow{p} 0.$$

## 5.2 Test Statistic

Theorem 3.i and the continuous mapping theorem suffice to show the STARX score statistic  $T_n(\theta)$  satisfies under  $H_0$

$$T_n(\theta) \Rightarrow z(\theta)'z(\theta) = T(\theta)$$

a chi-squared process on  $\mathbb{C}[\Theta_\xi]$ . For any fixed or randomized  $\theta$  the distribution  $T(\theta)$  is  $\chi^2(I)$ .

Under  $H_1$  Theorem 3.ii implies

$$T_n(\theta) / n \xrightarrow{p} \eta(\theta)'V(\theta)^{-1}\eta(\theta) > 0$$

for every  $\theta \in \Theta_\xi$  except possibly for countably many  $\tau \in S$ . But this means

$$T_n(\theta) \rightarrow \infty \text{ with probability one}$$

for uncountably infinitely many  $\theta$ . The STARX score test  $T_n(\theta)$  is therefore consistent since  $H_1$  captures any deviation from  $H_0$ .

Popular methods for handling  $\theta$  include randomization (LEE *et al.* [1996]), or continuous functionals  $h(T_n(\theta))$  including  $\sup_{\theta \in \Theta_\xi} T_n(\theta)$  and  $\text{ave}_{\Theta_\xi} T_n(\theta) = \int_{\Theta \in \Theta_\xi} T_n(\theta) d\mu(\theta)$  for some probability measure  $\mu(\theta)$  absolutely continuous with respect to Lebesgue measure (DAVIES [1977]; BIERENS [1990]; ANDREWS and PLOBERGER [1994, 1995]). Theorem 3 and the mapping theorem guarantee under  $H_0$

$$h(T_n(\theta)) \Rightarrow h(T(\theta))$$

and under  $H_1$

$$h(T_n(\theta) / n) \xrightarrow{p} h(\eta(\theta)'V(\theta)^{-1}\eta(\theta))$$

In the average and supremum cases  $h(T_n(\theta)) \rightarrow \infty$  with probability one under  $H_1$ .

Test statistic functionals like  $\sup_{\theta \in \Theta_\xi} T_n(\theta)$  and  $\text{ave}_{\Theta_\xi} T_n(\theta)$  have non-standard limit distributions. See HILL [2008] for details on a monte-carlo technique for approximating the asymptotic  $p$ -value, and a proof of asymptotic validity. Cf. GINE and HALL [1990] and HANSEN [1996].

## 6 NON-DEGENERATE STAR TESTS

---

There exist trivial cases in which  $V(\theta)$  is singular. If  $w(\delta, \tilde{z}_{t-1}) = \tilde{z}_{t-1} w(\delta, \tilde{z}_{t-1})$  and  $\theta = \tau = 0$ , for example, then  $f(0' \tilde{z}_{t-1}) = F(0)$  is a constant, so

$$\hat{s}_n(\hat{\phi}, 0, \theta) = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t \hat{z}_{t-1} F(0' \hat{z}_{t-1}) = K \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t \hat{z}_{t-1} = 0$$

by the least squares first order condition, and  $V(0) = 0$  a zero-matrix.

In this section we analyze the set of all  $\tau$  for which  $V(\theta)$  is singular. Define the set of parameters  $\tau$  that render  $V(\theta)$  singular:

$$S_\xi^* = \{\tau \in \Gamma : |\tau| \geq \xi, \lambda_{\min}(V(\theta)) = 0 \text{ and } P(\tau' \hat{z}_{t-1} \in R_0) = 1\}.$$

A proof that  $S_0^*$  has Lebesgue measure zero, similar in spirit to Lemma 2 of BIERENS [1990] and Lemma 2 of de JONG [1996], is easy to deliver in the present environment and is therefore omitted for the sake of brevity. Our aim is to provide fresh insight into the contents of  $S_\xi^*$ .

### 6.1 Neural Network Tests

Consider the case  $w(\delta, \tilde{z}_{t-1}) = 1$  (hence  $\theta = \tau$ ). Model (3) reduces to a single layer feedforward neural network form

$$y_t = f(\phi, \tilde{z}_{t-1}) + \beta \times F(\tau' \tilde{z}_{t-1}) + \varepsilon_t$$

In this case the set  $S$  from Lemma 1 and  $S_0^*$  are identically those considered in BIERENS [1991]. Assume the conditional variance is positive (a mild assumption).

**Assumption D**  $P(E[\varepsilon_t^2 | \mathfrak{F}_{t-1}] \geq \zeta) = 1$  for some constant  $\zeta > 0$ .

The following is a somewhat trivial argument, but important to note. If  $V(\tau) = 0$  then (5) implies

$$g_t(\tau)^2 E[\varepsilon_t^2 | \mathfrak{F}_{t-1}] = 0, \text{ a.s.}$$

Under Assumption D use (5) to deduce

$$F(\tau' \tilde{z}_{t-1}) = b(\tau, \phi) A(\phi)^{-1} \frac{\partial}{\partial \phi} f(\phi, \hat{z}_{t-1}), \text{ a.s.}$$

Since we assume  $E[\varepsilon_t(\partial/\partial\phi)f(\phi, \tilde{z}_{t-1})] = 0$  under Assumption A as a standard regulatory condition<sup>8</sup>, under either hypothesis  $V(\tau) = 0$  implies

$$E[\varepsilon_t F(\tau' \tilde{z}_{t-1})] = b(\tau, \phi) A(\phi)^{-1} E \left[ \varepsilon_t \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{t-1}) \right] = 0.$$

This is trivial under the null because  $E[\varepsilon_t | \mathfrak{F}_{t-1}] = 0$ , but it importantly implies  $S_0^* \subseteq S$  under the alternative: any  $\tau$  in a neural network setting that renders an asymptotically degenerate score test,  $V(\tau) = 0$ , also renders a score that is insensitive to model mis-specification,  $E[\varepsilon_t F(\tau' \tilde{z}_{t-1})] = 0$ .

**THEOREM 4** *If  $w(\delta, \tilde{z}_{t-1}) = 1$  then  $S_0^* \subseteq S$  under  $H_1$  and Assumption D.*

*Remark:* BIERENS [1990] and DE JONG [1996] show the set  $S_0^*$  has Lebesgue measure zero, but it is not known whether  $\tau \in S_0^*$  corresponds to  $\tau \in S$ . Under  $H_1$  and Assumption D we now know  $S_0^* \subseteq S$ . In other words, degeneracy is actually a secondary problem: the test  $T_n(\theta)$  fails to work in every sense possible.

## 6.2 Smooth Transition Tests

We can go further for linear specifications under the null when the chosen weight  $F(\tau' \tilde{z}_{t-1})$  has a non-zero derivative with positive probability. This covers standard ESTARX and LSTARX smooth transition functions  $F(\tau' \tilde{z}_{t-1})$ .

Denote by

$$\Theta_\xi^* : \text{any compact subset of } \{\theta = [\delta', \tau'] \in \Theta_\xi \mid P(\tau' \tilde{z}_{t-1} \in R_0) = 1\}$$

where  $R_0$  is the interval in Assumption B on which  $F$  is analytic and non-polynomial. Write  $F'(u) = (\partial/\partial u)F(u)$ .

**THEOREM 5** *Let  $f(\phi, \tilde{z}_{t-1}) = \phi' \tilde{z}_{t-1}$ ,  $w(\delta, \tilde{z}_{t-1}) = \tilde{z}_{t-1}$ , and*

$$P(F'(\tau' \tilde{z}_{t-1}) \neq 0 \cap \tau' \tilde{z}_{t-1} \in R_0) > 0.$$

*Under Assumptions A, B, and D the score statistic  $T_n(\theta)$  is not degenerate on  $\Theta_\xi^*$ :  $\inf_{\theta \in \Theta_\xi^*} \lambda_{\min}(V(\theta)) > 0$  for any  $\xi > 0$ . Hence  $S_\xi^*$  is empty.*

*Remark 1:* The result relies on a generalization of (5) when  $V(\theta)$  is singular. Any  $\tau \in S_\xi^*$  implies  $r'V(\theta)r = 0$  for some  $r \in \mathbb{R}^{pk+1}$ ,  $r'r = 1$ , which implies

$$F'(\tau' \tilde{z}_{t-1})r' \tilde{z}_{t-1} = \beta(r, \theta)' \tilde{z}_{t-1}, \text{ a.s.}$$

---

8. This implies  $f(\phi, \tilde{z}_{t-1})$  is best in some weak sense, even if  $E[\varepsilon_t | \tilde{z}_{t-1}] \neq 0$ . For example, when  $f(\phi, \tilde{z}_{t-1}) = \phi' \tilde{z}_{t-1}$  it is standard to assume  $E[\varepsilon_t \tilde{z}_{t-1}] = 0$

where

$$\beta(r, \theta) \equiv (E[\tilde{z}_{t-1}\tilde{z}'_{t-1}])^{-1} \times E[\tau'\tilde{z}_{t-1}F(\tau'\tilde{z}_{t-1})r'\tau'\tilde{z}_{t-1}]$$

Notice  $\beta(r, \theta)$  is simply the slope of the best linear  $L_2$ -metric projection of  $F(\tau'\tilde{z}_{t-1})r'\tilde{z}_{t-1}$  on  $\tilde{z}_{t-1}$ . For any  $\tau \in S_\xi^*$  such that  $T_n(\theta)$  is degenerate, the weight  $F(\tau'\tilde{z}_{t-1})r'\tilde{z}_{t-1}$  can be *almost surely* approximated by a linear function  $\beta(r, \theta)'\tilde{z}_{t-1}$  which itself cannot reveal model mis-specification by assumption since  $E[\varepsilon_t \tilde{z}_{t-1}] = 0$  under Assumption A. But this contradicts the revealing nature of the test weight  $F(\tau'\tilde{z}_{t-1})$  a la Lemma 1. BIERENS [1990] and DE JONG [1996] exploit Lemma 1 to deduce that the set of such  $\tau$  has Lebesgue measure zero. We prove that *no such*  $\tau \in S_\xi^*$  exists for exponential  $F(u) = \exp\{u\}$ , logistic  $F(u) = [1 + \exp\{u\}]^{-1}$ , etc.

*Remark 2:* The result can be extended to other specifications for  $w(\delta, \tilde{z}_{t-1})$  and  $f(\phi, \tilde{z}_{t-1})$  under appropriate modifications to the line of proof.

*Remark 3:* In a test of linear ARX against a general nonlinear alternative, the non-singularity Assumption C is superfluous, and may simply be replaced with the mild heteroscedasticity Assumption D.

*Remark 4:* In maximum likelihood settings the functional  $\text{ave}_{\Theta_\xi^*} T_n(\theta)$  can be interpreted as the limit of a (Gaussian) weighted average power optimal test, where power is directed toward alternatives near the null (ANDREWS and PLOBERGER [1994]). Similarly,  $\sup_{\theta \in \Theta_\xi^*} T_n(\theta)$  directs power toward distant alternatives but is only known to be asymptotically admissible (ANDREWS and PLOBERGER [1995]). In both cases the covariance matrix is required to be *uniformly positive definite* in the nuisance parameter space:  $\inf_{\theta \in \Theta_\xi^*} \lambda_{\min}(V(\theta)) > 0$ . Consistent CM tests of linear autoregression against a smooth transition alternative therefore provide a natural setting for ANDREWS and PLOBERGER'S [1994, 1995] optimal tests.

The last result covers the ESTARX model as a special case. For any finite  $\delta_0 \in \mathbb{R}$  define  $\tilde{\delta} \equiv [\delta_0, \delta']' \in \mathbb{R}^{pk+1}$ .

COROLLARY 6 *Let  $f(\phi, \tilde{z}_{t-1}) = \phi'\tilde{z}_{t-1}$ , and for  $\delta \in \mathbb{R}^{pk}$  and  $\tau \in \Gamma$*

$$w(\delta, \tilde{z}_{t-1})F(\tau'\tilde{z}_{t-1}) = \tilde{z}_{t-1} \exp \left\{ \sum_{i=1}^{pk} \delta_i \tilde{z}_{t-1,i}^2 \right\} \exp(\tau'\tilde{z}_{t-1}).$$

*Assume  $2\tilde{\delta}'\tilde{z}_{t-1}^2 + \tau'\tilde{z}_{t-1} \neq 0$  a.s. where  $\tilde{\delta} \neq 0$  and/or  $\tau \neq 0$ . Under Assumptions A, B, and D the statistic  $T_n(\theta)$  is never degenerate on  $\Theta^*$ :  $\inf_{\theta \in \Theta^*} \lambda_{\min}(V(\theta)) > 0$ .*

*Remark:* Notice  $\inf_{\theta \in \Theta^*} \lambda_{\min}(V(\theta)) > 0$  is based on the *complete* space  $\Theta^*$  and not the truncated subspace  $\Theta_\xi^*$ . A test directed toward an ESTARX alternative is *never* degenerate for any  $\theta = [\tau', \delta']' \in \Theta^*$ , hence for any scale  $\gamma = -\delta > 0$  and any threshold  $c \in \mathbb{R}^{pk}$  as long as  $2\tilde{\delta}'\tilde{z}_{t-1}^2 + \tau'\tilde{z}_{t-1} \neq 0$  a.s. The latter condition is trivial for purely autoregressive processes ( $\tilde{z}_{t-1} = [1, y_{t-1}, \dots, y_{t-p}]'$ ) under standard regulatory conditions (e.g.  $\mathfrak{I}_{t-1} \subset \mathfrak{I}_t$  and  $\mathfrak{I}_t \not\subset \mathfrak{I}_{t-1}$ ).

Together, Theorem 5 and Corollary 6 imply tests of linear ARX against popular smooth transition alternatives are never asymptotically degenerate.

## 7 SIMULATION STUDY

---

We now investigate the empirical size and power properties of  $\sup_{\theta \in \Theta_\xi} T_n(\theta)$  under a null of linear autoregression, and nonlinear alternatives. The statistic  $\text{ave}_{\theta_\xi} T_n(\theta)$  is non-negligibly dominated by  $\sup_{\theta \in \Theta_\xi} T_n(\theta)$ , in simulation evidence not presented here because the following alternatives are “distant” from the null. See ANDREWS and PLOBERGER [1994, 1995].

Write  $\tilde{z}_{it-1} = (1, y_{t-1}, \dots, y_{t-p_i})'$  for some orders  $p_i$ ,  $i = 1, 2$ . The iid innovations  $\{\varepsilon_t\}_{t=1}^n$  are drawn from a standard normal distribution for sample sizes  $n \in \{100, 500\}$ . The simulated models are

$$H_0: y_t = \phi' \tilde{z}_{1t-1} + \varepsilon_t$$

$$H_1^L: y_t = \phi' \tilde{z}_{1t-1} + \beta' \tilde{z}_{2t-1} [1 + \exp\{-\gamma' \tilde{z}_{1t-1}\}]^{-1} + \varepsilon_t$$

$$H_1^E: y_t = \phi' \tilde{z}_{1t-1} + \beta' \tilde{z}_{2t-1} \exp\left\{-\sum_{i=1}^{p_1} \gamma_i (\tilde{z}_{1t-1, i} - c_i)^2\right\} + \varepsilon_t$$

$$H_1^{AN}: y_t = \phi' \tilde{z}_{1t-1} + \beta \times [1 + \exp\{-\gamma' \tilde{z}_{1t-1}\}]^{-1} + \varepsilon_t$$

$$H_1^{SE}: y_t = \phi' \tilde{z}_{1t-1} + \beta' \tilde{z}_{2t-1} I(y_{t-1} > c_1) + \varepsilon_t$$

$$H_1^{BL}: y_t = \phi' \tilde{z}_{1t-1} + \beta \times y_{t-1} \varepsilon_{t-1} + \varepsilon_t, \quad |\beta| < 1$$

Notice  $\tilde{z}_{1t-1}$  and  $\tilde{z}_{2t-1}$  may have a different number of lags. Under  $H_0$  the true data generating process is a linear autoregression; under  $H_1^L$  and  $H_1^E$  a 2-regime LSTAR and ESTAR, respectively; under  $H_1^{AN}$  a logistic AR-ANN; under  $H_1^{SE}$  a Self Exciting Threshold Autoregression (SETAR), equivalent to the LSTAR  $y_t = \phi' \tilde{z}_{1t-1} + \beta' \tilde{z}_{2t-1} [1 + \exp\{-\gamma(y_{t-1} - c)\}]^{-1} + \varepsilon_t$  with  $\gamma \rightarrow \infty$ ; under  $H_1^{BL}$  the process is bilinear.

A total of  $3n$  observations are simulated and the last  $n$  are retained. For each simulated series  $p_i \in \{1, \dots, 10\}$ ,  $\phi \in [-.45, .45]^{p_1}$ ,  $\gamma \in [.5, 5]^{p_1} = \Gamma$ , and  $\gamma \in [.5, 5]^{p_1} = \Gamma$  are randomly drawn from uniform distributions. For  $H_1^{AN}$  and  $H_1^{BL}$ ,  $\beta$  is drawn uniformly from  $[-.45, .45]$ . For all other cases denote by  $\varphi \in \mathbb{R}^{\max\{p_1, p_2\}}$  the sum of  $\phi$  and  $\beta$  over elements  $1, \dots, \min\{p_1, p_2\}$  with zeros in the remaining elements  $\min\{p_1, p_2\}, \dots, \max\{p_1, p_2\}$ . If  $p_1 = p_2$  then  $\varphi = \phi + \beta$ ; if



$p_2 < p_1$  then if  $\varphi = \phi + [\beta', 0']$ , where  $\mathbf{0}$  denotes a  $p_1 - p_2$  vector of zeros; and so on. We use only those  $\{\phi, \beta\}$  such that  $\varphi$  has all polynomial roots outside the unit circle.

Since  $\varepsilon_t$  has a strictly positive, continuous density function on  $\mathbb{R}$ , given the parameter restrictions each  $\{y_t\}$  is strictly stationary and strong mixing. This follows since each is geometrically ergodic (DOUKHAN [1994], AN and HUANG [1996], NAJARIAN [2003], LEIBSCHER [2005]), hence strong mixing with geometrically decaying coefficients  $\{\alpha_i\}$  (DOUKHAN [1994], DAVIDSON [1994]). Therefore,  $\{y_t\}$  is properly heterogeneous in the sense of BIERENS [1987, 1994] and  $\nu$ -stable on a strong mixing process with coefficients  $\sum_{i=0}^{\infty} \alpha_i < \infty$ . Since the errors are iid normal random variables, the root condition ensures each  $y_t$  has infinitely many bounded moments and is properly heterogeneous (BIERENS [1987]). Together this implies Assumption A holds.

For each series an AR( $p^*$ ) model is estimated where  $p^* \in \{1, \dots, 10\}$  minimizes the AIC. The consistent STAR test, the standard STAR polynomial test, and a selection of extant tests of nonlinearity are applied to each time series.

## 7.1 Star Tests

Define the lag set  $\tilde{z}_{t-1}^* = [y_{t-1}, \dots, y_{t-p^*}]$ . The sup-STAR test is computed with exponential and logistic test weights  $\omega(\delta, \tilde{z}_{t-1}^*)F(\tau'\tilde{z}_{t-1}^*) = \tilde{z}_{t-1}^*[1 + \exp\{\tau'\tilde{z}_{t-1}^*\}]^{-1}$  and  $\omega(\delta, \tilde{z}_{t-1}^*)F(\tau'\tilde{z}_{t-1}^*) = \tilde{z}_{t-1}^* \exp\{\tau'\tilde{z}_{t-1}^*\}$ . The weights  $F(\tau'\tilde{z}_{t-1}^*)$  are constructed from standardized regressors  $\tilde{z}_{t-1}^*$  in order to stabilize the test statistic:  $F(\tau'\tilde{z}_{t-1}^*) = F(\sum_{i=1}^{p_1} \tau_i(\tilde{z}_{t-1,i} - \bar{\tilde{z}}_i)/s_i)$ , where  $\bar{\tilde{z}}_i$  and  $s_i$  denote the sample mean and standard deviation of  $\tilde{z}_{t-1,i}$ . As a default rule-of-thumb we simply use the same regressor set  $\tilde{z}_{t-1}^*$  in all components of the test weight. The supremum is computed over an increasing set of uniformly randomly selected nuisance parameters  $\{\tau_i\}_{i=1}^{\lfloor n/2 \rfloor} \in \Gamma$ . These are the LSTAR and ESTAR tests. Asymptotic  $p$ -values are computed according to the monte carlo method detailed in HILL [2008]. Covariance matrix estimators robust to unknown forms of conditional heteroscedasticity are used in all applicable cases here and below.

For the STAR polynomial test the following model is estimated:

$$y_t = \phi'\tilde{z}_{t-1}^* + \sum_{i=1}^L \vartheta_i'\tilde{z}_{t-1}^* y_{t-d}^i + u_t \text{ for } d = 1 \dots p$$

Under a null of linearity against a logistic STAR (or exponential STAR) alternative,  $L = 3$  (or 4) and  $\vartheta_i = 0$ ,  $i = 1 \dots 3$  (or 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 TERASVIRTA [1994]. These are the LPOLY and EPOLY tests, respectively.

## 7.2 Tests of Nonlinearity

We perform the neural test of neglected nonlinearity (LEE *et al.*, 1996), the BIERENS [1990] test, and the McLEOD-LI and RESET tests.

The BIERENS test is simply a sup-STAR test with  $\omega(\delta, \tilde{z}_{t-1}^*) = 1$  (denoted LBIER and EBIER).

The neural test of neglected nonlinearity is equivalent to a randomized STAR test with  $\omega(\delta, \tilde{z}_{t-1}^*) = 1$ , where  $\tau$  is uniformly randomly selected from  $\Gamma$  (denoted LNEUR and ENEUR).

The McLEOD and LI [1983] test is a standard portmanteau test on the squared null residuals  $\hat{\varepsilon}_t^2$  et for lags  $L = 1 \dots 3$ . The statistic is

$$ML_h = \frac{\sum_{t=h+1}^n (\hat{\varepsilon}_t^2 - \hat{\sigma}_2)(\hat{\varepsilon}_{t-h}^2 - \hat{\sigma}_2)}{\sum_{t=h+1}^n (\hat{\varepsilon}_t^2 - \hat{\sigma}_2)^2}$$

Recall the test's construction is based on the property that independent Gaussian innovations  $\varepsilon_t$  have white noise squares  $\varepsilon_t^2 - \sigma_2$ , and the minimum-mean-squared error predictor of a Gaussian time series  $y_t$  is linear.

For the Regression Specification Error Test (RESET) test we follow THURSBY and SCHMIDT [1977] by estimating an auxiliary regression based on the null residuals  $\hat{\varepsilon}_t$ ,

$$\hat{\varepsilon}_t = \beta'_0 x_t + \sum_{i=2}^L \sum_{j=2}^k \beta_{i,j} x_{t,j}^i + u_t, \text{ where } L = 3$$

A standard LM test of  $H_0: \beta_{i,j} = 0$  is performed.

## 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 test. Because  $\phi$  and  $\sigma = 1$  are known, for any  $\phi$  and  $\theta$  each STAR and ANN model can be represented as  $y_t(\phi) = \beta' z_{t-1}(\theta) + \varepsilon_t$ , where  $y_t(\phi) = y_t - \phi' \tilde{z}_{t-1}^*$  and  $z_{t-1}(\theta) = \omega(\delta, \tilde{z}_{t-1}^*) F(\tau' \tilde{z}_{t-1}^*)$ . For an arbitrary point  $(\phi, \theta)$  the least squares estimator of  $\beta$  is  $\hat{\beta}(\theta, \phi) = (z(\theta)' z(\theta))^{-1} z(\theta)' y(\phi)$ , where  $y(\phi) = \{y_t(\phi) : p^* + 1 \leq t \leq n\}$ , etc. The best test is simply the likelihood ratio which in the present known standard normal setting reduces to

$$\exp \{.5 \times y(\phi)' z(\theta) [z(\theta)' z(\theta)]^{-1} z(\theta)' y(\phi)\} = \exp \{.5 \times T_n(\phi, \theta)\}$$

say. We compute  $\sup_{\theta \in \Theta_\xi} T_n(\phi, \theta)$ , the MP-LSTAR and MP-ESTAR tests<sup>9</sup>.

9. The SETAR process is simply an LSTAR with  $\tau = \infty$ . Thus, the logistic sup-MP-STAR 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 are reported in Table 1. For each test statistic empirical size is comparable to the nominal size, although the polynomial test tends to under-reject the null.

In the present environment in which all parameters are randomly selected the popular STAR polynomial regression test is dominated by every test against each alternative (except the McLeod-Li test in some cases). Indeed, the consistent sup-STAR tests massively out-perform the conventional STAR tests. The conventional test is useless against a neural network alternative: recall ANN forms are nested within the smooth transition alternative.

Impressively, with just  $n = 500$  the sup-STAR tests obtain empirical power *nearly identical to the most-powerful* sup-MP-STAR tests against an LSTAR alternative (within .006), with a rejection rate above 90%. Similarly, the sup-STAR tests are comparable to sup-MP-STAR tests against AR-ANN and SE-TAR alternatives (in particular, the sup-ESTAR test).

The sup-STAR tests dominate each test performed against every alternative, except the McLeod-Li test against the bilinear alternative<sup>10</sup>. Finally, because smooth transition vector weights nest neural network alternatives it is not surprising that the sup-STAR test out-performs the Bierens test.

TABLE 1  
*Simulation Results*

	$n = 100$					
	$H_0$	$H_1^L$	$H_1^E$	$H_1^{AN}$	$H_1^{SE}$	$H_1^{BL}$
LSTARb	.057a	.478	.303	.281	.387	.301
ESTAR	.061	.426	.336	.302	.325	.296
MP-LSTAR	.042	.635	.801	.466	.570	.165
MP-ESTAR	.024	.536	.698	.359	.421	.129
LBIER	.011	.325	.325	.450	.213	.208
EBIER	.040	.297	.376	.406	.211	.226
LNEUR	.039	.357	.356	.365	.228	.211
ENEUR	.039	.342	.392	.376	.261	.206
LPOLY	.001	.043	.004	.023	.002	.018
EPOLY	.001	.043	.004	.023	.002	.018
RESET	.045	.261	.021	.027	.112	.092
ML1	.052	.113	.031	.182	.078	.516
ML2	.057	.124	.050	.177	.094	.517
ML3	.064	.151	.078	.163	.113	.524

10. This is not surprising since the test is particularly sensitive to multiplicative forms of omitted non-linearity.

$n = 500$						
	$H_0$	$H_1^L$	$H_1^E$	$H_1^{AN}$	$H_1^{SE}$	$H_1^{BL}$
LSTAR	.029	.912	.822	.803	.825	.766
ESTAR	.029	.915	.824	.872	.831	.729
MP-LSTAR	.016	.918	.901	.822	.980	.565
MP-ESTAR	.043	.921	.926	.881	.899	.572
LBIER	.024	.722	.617	.784	.588	.624
EBIER	.058	.786	.653	.759	.792	.624
LNEUR	.046	.635	.606	.736	.589	.469
ENEUR	.059	.622	.622	.777	.593	.479
LPOLY	.003	.433	.180	.019	.316	.029
EPOLY	.003	.433	.180	.019	.316	.029
RESET	.038	.411	.109	.039	.571	.006
ML1	.070	.306	.065	.058	.254	.975
ML2	.088	.366	.127	.058	.355	.987
ML3	.105	.409	.151	.058	.371	.996

Notes: a. Values denote rejection frequencies at the 5% level. b. LSTAR, etc., are the consistent sup-STAR tests; MP-LSTAR, etc., are the Most Powerful sup-STAR tests; LBIER, etc., are the Bierens tests (a sup-STAR test with scalar weight); LNEUR, etc., are the neural tests of neglected nonlinearity (a randomized STAR test with scalar weight); LPOLY, etc., are the popular STAR-polynomial tests; ML1, ML2 and ML3 are the McLeod-Li portmanteau tests of squared residuals over 1, 2 and 3 lags.

## 8 EMPIRICAL APPLICATION

We apply all tests in the simulation study (except the MP tests) to macroeconomic processes modeled in ROTHMAN *et al.* [2001] as a Logistic Smooth Transition VECM process. The variables studied are the logarithm of nominal, seasonally adjusted  $M1$  ( $m$ ), the logarithm of unadjusted output measured by the industrial production index ( $y$ ), the logarithm of the producer price index ( $p$ ), the commercial paper rate ( $r_p$ ), the 90-day Treasury bill rate ( $r_b$ ), and the rate spread  $r_b - r_p$ . All data were taken from the Saint Louis Federal Reserve data base, are monthly for the period 1959:01 - 2003:08, and seasonally adjusted at the source when applicable. Based on augmented Dickey-Fuller tests all variables, except for the rate spread, are differenced  $\{\Delta m, \Delta y, \Delta p, \Delta r_p\}$ . Evidence suggests the Treasury bill and commercial paper rates are cointegrated of order one such that the spread is  $I(0)$ . The sample size is 536 months, before lag and differencing adjustments.

Test results are reported in Table 2. The sup-STAR tests produce highly significant evidence in favor of smooth transition nonlinearity in money growth, inflation, and fluctuations in the commercial paper rate and the rate spread. The strongest evidence points to Logistic-STAR nonlinearity in each univariate series  $\{\Delta m, \Delta, \Delta r_p, r_b - r_p\}$ . By comparison, the polynomial regression tests provide weaker evidence of STAR

nonlinearity, and do not detect a smooth transition structure in the rate spread series. At the 5% level the neural test of neglected nonlinearity only finds logistic nonlinearity in the commercial paper rate and exponential nonlinearity in the rate spread. The RESET test fails to detect nonlinearity in any series.

TABLE 2  
*Macroeconomic Fluctuations*

	$\Delta_m$	$\Delta_y$	$\Delta_p$	$\Delta r_b$	$\Delta r_p$	$r_b - r_p$
LSTAR	.0000	1.000	.0000	.6700	.0200	.0000
ESTAR	.0000	1.000	.0000	.9500	.0100	.0000
LNEUR	.0921	.5621	.4432	.3444	.0076	.0327
ENEUR	.2154	.2715	.1108	.5181	.0695	.5619
LBIER	.3800	.2800	.4400	.1400	.0500	.1700
EBIER	.4800	.1900	.3900	.1100	.6600	.1000
LPOLY	.0793	.4401	.0525	.1143	.0144	.5239
EPOLY	.0019	.3310	.1059	.2018	.0090	.1277
RESET	.3933	.3103	.2875	.2953	.9339	.2066

Notes: Numbers are *p-values* of the respective tests.

## 9 CONCLUSION

We present a new test of regression model specification against a general class of Smooth Transition Autoregressions. The test obtains an asymptotic power of one against any form of model mis-specification, and delivers a nonlinear STARX alternative that is guaranteed to improve the model fit. The test solves major shortcomings of the seemingly universal practice of linearizing the transition function and performing F-tests on polynomial regression coefficients. The conventional test is not consistent against a general alternative, it is ineffective against non-exponential or non-logistic smooth transition forms, and uses the wrong limit distribution in cases when the *p*-value is optimized in order to select a delay parameter. Our test performs impressively well against conventional tests of functional form, non-negligibly dominates the conventional STAR test, and nearly matches the empirical power of a Most Powerful test. ■

## References

- AN H.Z., HUANG F.C. (1996). - « The Geometrical Ergodicity of Nonlinear Autoregressive Models », *Statistica Sinica*, 6, p. 943-956.
- ANDREWS D.W.K., PLOBERGER W. (1994). - « Optimal Tests when a Nuisance Parameter is Present Only under the Alternative », *Econometrica*, 82, 1383-1414.

- ANDREWS D.W.K., PLOBERGER W. (1995). - « Admissibility of the Likelihood Ratio Test when a Nuisance Parameter is Present Only under the Alternative », *Annals of Statistics*, 23, p. 1609-1629.
- BAILLIE R. T. (1980). - « Predictions From ARMAX Models », *Journal of Econometrics*, 12, p. 365-374.
- BIERENS H. J. (1982). - « Consistent Model Specification Tests », *Journal of Econometrics*, 20, p. 105-134.
- BIERENS H. J. (1983). - « Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions », *Journal of the American Statistical Association*, 78, p. 699-707.
- BIERENS H. J. (1984). - « Model Specification Testing of Time Series Regressions », *Journal of Econometrics*, 26, p. 323-353.
- BIERENS H. J. (1987). - « ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands », *Journal of Econometrics*, 35, p. 161-190.
- BIERENS H. J. (1990). - « A Consistent Conditional Moment Test of Functional Form », *Econometrica*, 58, p. 1443-1458.
- BIERENS H. J. (1991). - « Least Squares Estimation of Linear and Nonlinear ARMAX Models under Data Dependence », *Annales d'Economie et de Statistique*, 20/21, p. 143-169.
- BIERENS H. J. (1994). - *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models*, Cambridge University Press, Cambridge.
- BIERENS H. J., PLOBERGER W. (1997). - « Asymptotic Theory of Integrated Conditional Moment Tests », *Econometrica*, 65, p. 1129-1151.
- BILLINGSLEY P. (1999). - *Convergence of Probability Measures*, John Wiley and Sons, New York.
- CHAN K.S., TONG H. (1986a). - « On Tests for Nonlinearity in Time Series Analysis », *Journal of Time Series Analysis*, 5, p. 217-228.
- CHAN K.S., TONG H. (1986b). - « On Estimating Thresholds in Autoregressive Models », *Journal of Time Series Analysis*, 7, p. 179-190.
- DAVIDSON J. (1994). - *Stochastic Limit Theory*, Oxford University Press, Oxford.
- DAVIDSON J. (2004). - « Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model », *Journal of Business and Economic Statistics*, 22, p. 16-29.
- DAVIES R.B. (1977). - « Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative », *Biometrika*, 64, p. 247-254.
- DONALDSON R.G., KAMSTRA M. (1996). - « Using Dividend Forecasting Models to Reject Bubbles in Asset Prices: The Case of the Crash of 1929 », *Review of Financial Studies*, 9, p. 333-383.
- DE JONG R. (1996). - « The Bierens Test under Data Dependence », *Journal of Econometrics*, 72, p. 1-32.
- DETTE, H. (1999). - « A Consistent Test for the Functional Form of a Regression Based on a Difference of Variance Estimators », *Annals of Statistics*, 27, p. 1012-40.
- VAN DIJK D., TERASVIRTA T., FRANSES P.H. (2002). - « Smooth Transition Autoregressive Models - A Survey of Recent Developments », *Econometric Reviews*, 21, p. 1-47.
- DOUKHAN P. (1994). - « Mixing: Properties and Examples », *Lecture Notes in Statistics* 85, Berlin, Springer.
- ESCRIBANO A., JORDA O. (1999). - « Improved Testing and Specification of Smooth Transition Autoregressive Models », in P. Rothman (ed.), *Nonlinear Time Series Analysis of Economic and Financial Data*, p. 289-319, Boston, Kluwer.
- GALLANT A.R., WHITE H. (1988). - *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, New York, Basil Blackwell.
- GINE E., ZINN J. (1990). - « Bootstrapping General Empirical Measures », *Annals of Probability* 18, 851-869.
- GONZALEZ-RIVERA G. (1998). - « Smooth-transition GARCH Models », *Studies in Nonlinear Dynamics and Econometrics*, 3, p. 61-78.

- HANSEN B. (1996). - « Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis », *Econometrica*, 64, p. 413-430.
- HARDLE W., MAMMEN E. (1993). - « Comparing Nonparametric Versus Parametric Regression Fits », *Annals of Statistics*, 21, p. 1926-1947.
- HILL J.B (2007). - « Stochastically Weighted Average Conditional Moment Tests of Functional Form », Dept. of Economics, University of North Carolina-Chapel Hill, [www.unc.edu/~jbhill/swacm\\_test.pdf](http://www.unc.edu/~jbhill/swacm_test.pdf).
- HILL J.B. (2008). - « Technical Appendix for “Consistent and Non-Degenerate Model Specification Tests Against Smooth Transition and Neural Network Alternatives” », Dept. of Economics, University of North Carolina-Chapel Hill, [www.unc.edu/~jbhill/tech\\_append\\_star\\_tests.pdf](http://www.unc.edu/~jbhill/tech_append_star_tests.pdf).
- HONG Y., WHITE H. (1995). - « Consistent Specification Testing via Non-parametric Series Regression », *Econometrica*, 63, p. 1133-1159.
- HORNIK K. (1991). - « Approximating Capabilities of Multilayer Feedforward Networks », *Neural Networks*, 4, p. 251-257.
- HORNIK, K., STINCHCOMBE M., WHITE H. (1989). - « Multilayer Feedforward Networks are Universal Approximators », *Neural Networks*, 2, p. 359-366.
- KAPETANIOS G, SHIN Y., SNELL A. (2000). - « Testing for a Unit Root Against Nonlinear STAR Models », *Journal of Econometrics*: forthcoming.
- KAPETANIOS G., SHIN Y. (2003). - « Unit Root Tests in Three-Regime SE-TAR Models », *Econometrics Journal*, 9, p. 252-278.
- KUAN, C., WHITE H. (1994). - « Artificial Neural Networks: An Economic Perspective », *Econometric Reviews*, 13, p. 1-91.
- LEE T., WHITE H., GRANGER C.W.J. (1993). - « Testing for Neglected Nonlinearity in Time-Series Models: A Comparison of Neural Network Methods and Alternative Tests », *Journal of Econometrics*, 56, p. 269-290.
- LEIBSCHER, E. (2005). - « Towards a Unified Approach for Proving Geometric Ergodicity and Mixing Properties of Nonlinear Autoregressive Processes », *Journal of Time Series Analysis*, 26, p. 669-689.
- LI Q., HSIAO C., ZINN J. (2003). - « Consistent Specification Tests for Semiparametric/ Nonparametric Models Based on Series Estimation Methods », *Journal of Econometrics*, 112, p. 295-325.
- LIN J., TERASVIRTA T. (1994). - « Testing the Constancy of Regression Parameters against Continuous Structural Change », *Journal of Econometrics*, 62, p. 211-228.
- LUNDBERG S., TERASVIRTA T. (2002). - « Evaluating GARCH Models », *Journal of Econometrics*, 110, p. 417-435.
- LUNDBERG S., TERASVIRTA T. (2005). - « A Time Series Model for an Exchange Rate in a Target Zone with Applications », *Journal of Econometrics*: in press.
- LUNDBERG S., TERASVIRTA T., VAN DIJK D. (2003). - « Time-Varying Smooth Transition Autoregressive Models », *Journal of Business and Economic Statistics*, 21, p. 104-121.
- LUUKKONEN R., SAIKKONEN P., TERASVIRTA T. (1988). - « Testing Linearity against Smooth Transition Autoregressive Models », *Biometrika*, 75, p. 491-9.
- MCLEISH D.L. (1974). - « Dependent Central Limit Theorems and Invariance Principles », *Annals of Probability*, 2, p. 620-628.
- MEITZ M., SAIKKONEN P. (2008). - « Stability of Nonlinear AR-GARCH Models », *Journal of Time Series Analysis*, 29, p. 453-475.
- NAJARIAN, K. (2003). - « On Stochastic Stability of Dynamical Neural Models in Presence of Noise », *Proceedings of the Fourth International Conference on Dynamical Systems and Differential Equations*, p. 656-663.
- PENA D., SANCHEZ I. (2005). - « Multifold Predictive Validation in ARMAX Time Series Models », *Journal of the American Statistical Association*, 100, p. 135-146.
- POTSCHER B.M., PRUCHA I.R. (1991). - « Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometrics Models, Part I: Consistency and Approximation Concepts », *Econometric Reviews*, 10, p. 125-216.
- 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, p. 301-318.

- ROTHMAN P., VAN DIJK D. , FRANSES P.H. (2001). - « A Multivariate STAR Analysis of the Relationship between Money and Output », *Macroeconomic Dynamics*, 5, p. 506-532.
- ROYDEN H.L. (1968). - *Real Analysis*, Macmillan, London.
- SAIKKONEN P., LUUKKONEN R. (1988). - « Lagrange Multiplier Tests for Testing Nonlinearities in Time Series Models », *Scandinavian Journal of Statistics*, 15, p. 55-68.
- STINCHCOMBE M.B., WHITE H. (1992). - « Some Measurability Results for Extrema of Random Functions over Random Sets », *Review of Economic Studies*, 59, p. 495-514.
- STINCHCOMBE M.B., WHITE H. (1998). - « Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative », *Econometric Theory*, 14, p. 295-325.
- STUTE W. (1997). - « Nonparametric Model Checks for Regression », *Annals of Statistics*, 35, p. 613-641.
- TERASVIRTA T. (1994). - « Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models », *Journal of the American Statistical Association*, 89, p. 208-218.
- THURSBY J., SCHMIDT P. (1977). - « Some Properties of Tests for Specification Error in a Linear Regression Model », *Journal of the American Statistical Association*, 72, p. 635-641.
- 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.
- YATCHEW A.J. (1992). - « Nonparametric Regression Tests Based on Least Squares », *Econometric Theory*, 8, p. 435-451.



## Appendix A: Nu-Stability

We exploit in various proofs a uniform law of large numbers and central limit theorem for  $\nu$ -stable random variables  $\{W_t\} = \{y_t, x_t\}$  that are properly heterogeneous in the sense of BIERENS [1987: p. 151]. Let  $F_{t,m}$  denote the joint distribution of  $\{W_t, \dots, W_{t-m}\}$ . Proper heterogeneity implies  $1/n \sum_{t=1}^n F_{t,m} \rightarrow H_m$  where  $H_m$  is a proper distribution function. Strict stationarity trivially implies proper heterogeneity.

The  $\nu$ -stable property, due to BIERENS [1983, 1987, 1991, 1994], is defined as follows.

**Definition** A stochastic process  $\{W_t\} \in \mathbb{R}^k$  is  $L_r$ - $\nu$ -stable on a base process  $\{\varepsilon_t\}$  if there exists a bounded non-stochastic mapping  $\nu: \mathbb{N} \rightarrow \mathbb{R}_+$  satisfying

$$\sup_{t \in \mathbb{Z}} \left\| W_t - E[W_t | \{\varepsilon_{t-i}\}_{i=0}^m] \right\|_r = O(\nu(m)) \text{ where } \nu(m) \rightarrow 0 \text{ as } m \rightarrow \infty.$$

*Remark 1:* Any  $L_2$ - $\nu$ -stable process is  $L_1$ - $\nu$ -stable by Liapouov's inequality (DAVIDSON 1994: Theorem 9.23).

*Remark 2:* Uniform boundedness in  $t$  rules out forms of global non-stationarity where the  $r^{\text{th}}$ -moment of  $W_t$  is not uniformly bounded.

The  $\nu$ -stability condition has since been categorized under Near Epoch Dependence (DAVIDSON [1994]). The only differences between NED and  $\nu$ -stability are (i)  $\nu$ -stability imposes uniform boundedness in  $t$ ; and (ii) NED uses a two-sided future-past lag of the shock, and divides the right-hand-side into time-dependent constants  $d_t$  (not necessarily bounded) and lag-dependent coefficients  $\nu(m)$  (DAVIDSON [1994]):

$$\left\| z_t - E[z_t | \{\varepsilon_{t-i}\}_{i=-m}^m] \right\|_r \leq d_t \times \nu(m)$$

Neither property characterizes processes with a non-negligible infinite past (e.g. a random walk). For all practical purposes the two concepts are identical for globally stationary, one-sided processes. Indeed, although there are few results establishing which processes are  $\nu$ -stable, all results establishing NED for one-sided time series with  $\sup_{t \in \mathbb{Z}} d_t < \infty$  apply to  $\nu$ -stability, including ARFIMA( $p, d, q$ ) and FIGARCH( $p, d, q$ ) with  $d \in [0, 1]$ , bilinear process, and a host of nonlinear distributed lags all with innovations  $\varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ ,  $\sigma^2 < \infty$ . Further, since any process is  $\nu$ -stable on itself, geometrically ergodic and therefore geometrically strong mixing data are  $\nu$ -stable. Under appropriate smoothness conditions on the probability density of  $\{y_t\}$ ,  $\nu$ -stable therefore covers linear and nonlinear GARCH, AR-GARCH, neural nets, contraction mappings, Threshold Autoregressions, Vector

ARCH, and more. See GALLANT and WHITE [1988], DOUKHAN [1994], DAVIDSON [1994, 2004], AN and HUANG [1996], NAJARIAN [2003], LEIBSCHER [2005] and MEITZ AND SAIKKONEN [2008].

## Appendix B: Assumptions

---

### Assumption A.1

$W_t = \{y_t, x_t\} \in \mathbb{R} \times \mathbb{R}^{k-1}$  exists in  $L_2(Y, \mathfrak{F}, P)$ ,  $\mathfrak{F}_t = \sigma(W_i : \tau < t)$ ,  $\mathfrak{F}_{t-1} \subset \mathfrak{F}_t$ ,  $\mathfrak{F} = \sigma(\cup_t \mathfrak{F}_t)$   $\{W_t\}$  is properly heterogeneous, governed by a non-degenerate joint distribution function with non-degenerate marginal distributions, and for some  $\kappa > 0$ ,  $\sup_{t \in \mathbb{N}} \|W_t\|_{4+\kappa} < \infty$ .

$\{W_t\}$  is  $L_2$ - $\nu$ -stable on a strong-mixing base with mixing coefficients  $\{\alpha_i\}$ ,  $\sum_{i=1}^{\infty} \alpha_i < \infty$ <sup>11</sup>. The error  $\varepsilon_t$  has an almost everywhere positive, continuous density, and is weakly orthogonal in the sense  $E[\varepsilon_t \tilde{z}_{t-1}] = 0$  and  $E[\varepsilon_t (\partial / \partial \phi) f(\phi, \tilde{z}_{t-1})] = 0$  where  $\tilde{z}_{t-1} = [1, (y_{t-1}, x'_{t-1}), \dots, (y_{t-p}, x'_{t-p})]'$ . Moreover  $\|\varepsilon_t\|_{4+\kappa} < \infty$  for some  $\kappa > 0$ .

### Assumption A.2

The function  $f(\phi, \tilde{z}_{t-1}) : \Phi \times \mathbb{R}^{pk+1} \rightarrow \mathbb{R}$  is for each  $\tilde{z}_{t-1} \in \mathbb{R}^{pk+1}$  a continuous real function and twice continuously differentiable on compact  $\Phi \in \mathbb{R}^{pk+1}$ . Moreover,  $f(\phi, \tilde{z}_{t-1})$  is for each  $\phi \in \Phi$  a Borel-measurable function on  $\mathbb{R}^{pk+1}$ . Define  $\partial f(\phi, \tilde{z}_{t-1}) \equiv (\partial / \partial \phi) f(\phi, \tilde{z}_{t-1})$  and  $\partial \partial f(\phi, \tilde{z}_{t-1}) \equiv (\partial^2 / \partial \phi \partial \phi) f(\phi, \tilde{z}_{t-1})$ . For each  $t$ :  $\|\sup_{\phi \in \Phi} |f(\phi, \tilde{z}_{t-1})|_1\|_{4+\kappa} < \infty$ ,  $\|\sup_{\phi \in \Phi} |\partial f(\phi, \tilde{z}_{t-1})|_1\|_{4+\kappa} < \infty$ ,  $\|\sup_{\phi \in \Phi} |\partial f(\phi, \tilde{z}_{t-1}) \partial f(\phi, \tilde{z}_{t-1})|_1\|_{4+\kappa} < \infty$ , and  $\|\sup_{\phi \in \Phi} |\partial \partial f(\phi, \tilde{z}_{t-1})|_1\|_{4+\kappa} < \infty$  for some  $\kappa > 0$ .

The second regime  $\omega(\delta, \tilde{z}_{t-1})$  denotes a Borel-measurable mapping from  $D \times \mathbb{R}^{pk+1}$  to  $\mathbb{R}^l$ ,  $l \geq 1$ , where  $D$  is a compact subset of  $\mathbb{R}^q$ ,  $q \geq 0$ .  $P(a_0 < |\omega(\delta, \tilde{z}_{t-1})|_1 < a_1) = 1$  uniformly in  $\delta$  for constants  $0 < a_0 < a_1 < \infty$ .

**Assumption A.3** There exists a unique element  $\phi = \arg \inf_{\phi \in \Phi} E(y_t - f(\phi, \tilde{z}_{t-1}))^2$  in the interior of  $\Phi$ . Under either hypothesis  $E[\varepsilon_t \partial f(\phi, \tilde{z}_{t-1})] = 0$  and  $E[\varepsilon_t \tilde{z}_{t-1}] = 0$ , where  $\tilde{z}_{t-1} = (1, z'_{t-1}, \dots, z'_{t-p})'$ .

---

11. For example if  $\alpha_i^* = O(\rho^i)$ ,  $|\rho| < 1$ , such that memory is “short” and decay is geometric, then  $\left| \sum_{i=1}^{\infty} \alpha_i \right| \leq K\rho / (1-\rho) < \infty$  is trivial. Geometric strong mixing is implied by geometric ergodicity, which holds for a host of nonlinear time series. Long memory cases include mixing with size  $\lambda > 1$  since  $\alpha_i = O(i^{-\lambda})$  implies  $\sum_{i=1}^{\infty} |\alpha_i| < \infty$ .

**Assumption A.4** The non-stochastic matrix  $A(\phi) = E[\partial f(\phi, \tilde{z}_{t-1})\partial' f(\phi, \tilde{z}_{t-1})]$  is uniformly positive definite on  $\Phi$ . Let  $\theta = [\delta', \tau'] \in \Theta = D \times \Gamma$ , a compact subset of  $\mathbb{R}^{q+(pk+1)}$  with positive Lebesgue measure. Let  $T$  denote an arbitrary compact subset of  $\mathbb{R}^{pk+1}$ . For some  $\kappa > 0$  and  $K < \infty$

$$\begin{aligned} & \left\| \sup_{\theta \in \Theta} |F(\tau' \tilde{z}_{t-1}) \omega(\delta, \tilde{z}_{t-1}) \omega(\delta, \tilde{z}_{t-1})'| \right\|_{4+\kappa} \leq K \\ & \left\| \sup_{\theta \in \Theta} \sup_{\tau \in T} |F(\tau' \tilde{z}_{t-1}) \partial f(\phi, \tilde{z}_{t-1}) \partial' f(\phi, \tilde{z}_{t-1})| \right\|_{4+\kappa} \leq K \\ & \left\| \sup_{\phi \in \Theta} \sup_{\theta \in \Theta} |F(\tau' \tilde{z}_{t-1}) \omega(\delta, \tilde{z}_{t-1}) \partial' f(\phi, \tilde{z}_{t-1})| \right\|_{4+\kappa} \leq K \\ & \left\| \sup_{\theta \in \Theta} |F(\tau' \tilde{z}_{t-1}) \times \omega(\delta, \tilde{z}_{t-1})| \right\|_{4+\kappa} \leq K \\ & \left\| \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \{F(\tau' \tilde{z}_{t-1}) \times \omega(\delta, \tilde{z}_{t-1})\} \right| \right\|_{4+\kappa} \leq K \end{aligned}$$

## Appendix C: Proofs of Main Results

**Proof of Theorem 3.** By Lemma A.2 in Appendix D the finite dimensional distributions of  $V(\theta)^{-1/2} \sqrt{ns_n}(\phi, 0, \theta)$  converge to normal distributions under the null hypothesis. Lemma A.3 proves the sequence  $\{P_n\}$  of probability measures  $P_n$  associated with  $V(\theta)^{-1/2} \sqrt{ns_n}(\phi, 0, \theta)$  is tight in  $\mathcal{C}[\Theta_\xi]$ . The result under  $H_1$  follows from Lemma A.4.

**Proof of Theorem 5.** Write  $\lambda_{\min}(\theta) = \lambda_{\min}(V(\theta))$ . By assumption  $(\partial/\partial\phi)f(\phi, \tilde{z}_{t-1}) = \tilde{z}_{t-1} = \omega(\delta, \tilde{z}_{t-1})$  such that  $\theta = \tau$ . Step 1 proves  $\inf_{\theta \in \Theta_\xi^*} \lambda_{\min}(\theta) = \min_{\theta \in \Theta_\xi^*} \lambda_{\min}(\theta)$  and Step 2 proves  $\min_{\theta \in \Theta_\xi^*} \lambda_{\min}(\theta) > 0$ .

**Step 1:** By construction  $\lambda_{\min}(\theta) = \inf_{r,r=1} E[\varepsilon_t^2 (r'g_t(\theta))^2] \geq 0$ , hence  $\lambda_{\min} : \Theta_\xi^* \rightarrow \mathbb{R}_+$  is a proper function. Under Assumption A the function  $\lambda_{\min}(\cdot)$  is uniformly continuous on the compact subset<sup>12</sup>  $\Theta_\xi^*$ . The image of a compact

12. For any  $\delta > 0$ ,  $\sup_{\theta_1 \in \Theta_\xi^*} \sup_{\theta_2 \in \Theta_\xi^* : \|\theta_1 - \theta_2\|_1 < \delta} |\lambda_{\min}(\theta_1) - \lambda_{\min}(\theta_2)| \leq \sup_{\theta \in \Theta_\xi^*} |(\partial/\partial\theta)\lambda_{\min}(\theta)|_1 \times \delta$ , where  $\sup_{\theta \in \Theta_\xi^*} |(\partial/\partial\theta)\lambda_{\min}(\theta)|_1 = \sup_{\theta \in \Theta_\xi^*} |(\partial/\partial\theta) \inf_{r,r=1} E[\varepsilon_t^2 (r'g_t(\theta))^2]|_1$ . By Chebychev's and Liapouon's inequalities, for some finite  $B > 0$ ,

$$\sup_{\theta \in \Theta_\xi^*} \left| \frac{\partial}{\partial \theta} \inf_{r,r=1} E[\varepsilon_t^2 (r'g_t(\theta))^2] \right|_1 \leq B \|\varepsilon_t\|_4^2 \left\| \sup_{\theta \in \Theta_\xi^*} |g_t(\theta)|_1 \right\|_4 \left\| \sup_{\theta \in \Theta_\xi^*} \left| \frac{\partial}{\partial \theta} g_t(\theta) \right|_1 \right\|_4$$

Each component on the right hand side is bounded from above by Assumption A and Lemma A.6. Hence  $\lambda_{\min}(\cdot)$  is uniformly continuous.

set under a continuous mapping is compact, hence  $\lambda_{\min}(\cdot)$  is compact on  $\Theta_{\xi}^*$ . Therefore the image of  $\lambda_{\min}(\cdot)$  is closed and bounded, and  $\lambda_{\min}(\cdot)$  admits a unique minimum on  $\Theta_{\xi}^*$ . This implies  $\inf_{\theta \in \Theta_{\xi}^*} \lambda_{\min}(\theta) = \min_{\theta \in \Theta_{\xi}^*} \lambda_{\min}(\theta)$ .

**Step 2:** It suffices to prove  $S_{\xi}^*$  is empty for any  $\xi > 0$  since  $S_{\xi}^* = \emptyset$  implies  $\lambda_{\min}(\theta) > 0$  for all  $\theta$  in the closed and bounded  $\Theta_{\xi}^*$ , and therefore  $\inf_{\theta \in \Theta_{\xi}^*} \lambda_{\min}(\theta) = \min_{\theta \in \Theta_{\xi}^*} \lambda_{\min}(\theta)$ .

Any  $\tau \in S_{\xi}^*$  implies  $\lambda_{\min}(V(\theta)) = 0$ . This means  $r'V(\theta)r = 0$  for some  $r \in \mathbb{R}^{pk+1}$ ,  $r'r = 1$ . By the construction of  $V(\theta)$  in (5), this in turn implies

$$r'g_t(\theta)g_t(\theta)'rE[\varepsilon_t^2 | \mathfrak{F}_{t-1}] = 0, \text{ a.s.}$$

From Assumption D we deduce  $r'g_t(\theta) = 0$  a.s., hence

$$(6) \quad r'\tilde{z}_{t-1}F(\tau'\tilde{z}_{t-1}) = \beta(r, \theta)'\tilde{z}_{t-1}, \text{ a.s., where } \beta(r, \theta) \equiv A^{-1}b(\theta)'r$$

For fixed  $r$  and  $\tau \in S_{\xi}^*$ , (6) defines a functional identity with respect to  $\tilde{z}_{t-1}$  with probability one. Differentiating both sides of (6) with respect to  $\tilde{z}_{t-1}$ , multiplying by  $\tilde{z}_{t-1}$ , and using identity (6) we find

$$(7) \quad \begin{aligned} r'\tilde{z}_{t-1}F(\tau'\tilde{z}_{t-1}) + r'\tilde{z}_{t-1}F'(\tau'\tilde{z}_{t-1})\tau'\tilde{z}_{t-1} \\ = \beta(r, \theta)'\tilde{z}_{t-1} = r'\tilde{z}_{t-1} \exp(\tau'\tilde{z}_{t-1}), \text{ a.s.} \end{aligned}$$

Notice  $r'\tilde{z}_{t-1} \neq 0$  and  $\tau'\tilde{z}_{t-1} \neq 0$  each with probability one due to  $r \neq 0$ ,  $\tau \neq 0$ , and the non-singularity of  $E[\tilde{z}_{t-1}\tilde{z}_{t-1}']$  under Assumption A. By canceling like terms in (7), any  $\tau \in S_{\xi}^*$  implies

$$F'(\tau'\tilde{z}_{t-1}) = 0, \text{ a.s.}$$

But any  $\tau \in S_{\xi}^*$  satisfies  $\tau'\tilde{z}_{t-1} \in R_0$  a.s. Therefore  $P(F'(\tau'\tilde{z}_{t-1}) \neq 0 \cap \tau'\tilde{z}_{t-1} \in R_0) = 0$ , a contradiction of the assumption  $P(F'(\tau'\tilde{z}_{t-1}) \neq 0 \cap \tau'\tilde{z}_{t-1} \in R_0) > 0$ . Therefore  $S_{\xi}^*$  is empty.

**Proof of Corollary 6.** Without loss of generality we may substitute  $\tilde{z}_{t-1} \exp\left(\sum_{i=1}^{pk} \delta_i \tilde{z}_{t-1}^2\right) \exp(\tau'\tilde{z}_{t-1})$  with  $\tilde{z}_{t-1} \exp\{\delta_0\} \exp\left(\sum_{i=1}^{pk} \delta_i \tilde{z}_{t-1}^2\right) \exp(\tau'\tilde{z}_{t-1})$  for any finite  $\delta_0 \in \mathbb{R}$ . Using an argument identical to the line of proof of Theorem 5, any  $\tau \in S_{\xi}^*$  satisfies

$$(8) \quad r'\tilde{z}_{t-1} \exp(\delta^{-1}\tilde{z}_{t-1}^2 + \tau'\tilde{z}_{t-1}) = \beta(r, \theta)'\tilde{z}_{t-1}, \text{ a.s. where } \tilde{\theta} = (\delta_0, \delta)'$$

where  $\tilde{\theta} = 0$  is possible. Differentiate both sides of (8) with respect to  $\tilde{z}_{t-1}$  and multiply by  $\tilde{z}_{t-1}$ :

$$\begin{aligned} r'\tilde{z}_{t-1} \exp(\tilde{\delta}'\tilde{z}_{t-1} + \tau'\tilde{z}_{t-1}) + r'\tilde{z}_{t-1} \exp(\tilde{\delta}'\tilde{z}_{t-1} + \tau'\tilde{z}_{t-1})(2\tilde{\delta}'\tilde{z}_{t-1}^2 + \tau'\tilde{z}_{t-1}) \\ = \beta(r, \theta)' \tilde{z}_{t-1} \end{aligned}, a.s.$$

Canceling like terms by using (8), and noting  $r'\tilde{z}_{t-1} \neq 0$  *a.s.* and  $\exp(\tilde{\delta}'\tilde{z}_{t-1} + \tau'\tilde{z}_{t-1}) \neq 0$  *a.s.* under Assumption A and the boundedness of  $D$  and  $\Gamma$ , it must be the case that

$$2\tilde{\delta}'\tilde{z}_{t-1}^2 + \tau'\tilde{z}_{t-1} = 0, a.s.,$$

a contradiction of the assumption  $2\tilde{\delta}'\tilde{z}_{t-1}^2 + \tau'\tilde{z}_{t-1} = 0, a.s.$  given  $\tau \neq 0$ . But for any  $\tilde{\delta} \neq 0$  an identical argument implies  $\tau \neq 0 \notin S_0^*$ , hence  $S_0^*$  is empty.

## Appendix D: Supporting Lemmata

---

**Lemma A.1** *Under Assumptions A and C and  $H_\rho$*

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

**Lemma A.2** *Under Assumptions A and C and  $H_0$  the finite dimensional distributions of  $V(\theta)^{-1/2} \sqrt{n} \hat{s}_n(\phi, 0, \theta)$  converge to multivariate normal distributions, where  $\theta \in \Theta_\xi$ .*

**Lemma A.3** *Under Assumptions A and C and  $H_0$ ,  $V(\theta)^{-1/2} \sqrt{n} \hat{s}_n(\phi, 0, \theta)$  is tight on  $\Theta_\xi$ .*

**Lemma A.4** *Under Assumptions A-C and  $H_1$  there exists a non-stochastic function  $\eta: \Theta_\xi \rightarrow \mathbb{R}^{pk+1}$  and a subset  $S \subseteq \mathbb{R}^{pk+1}$  with Lebesgue measure zero such that*

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

where  $V(\theta)^{-1/2} \eta(\theta) \neq 0$  for every  $\theta \in \Theta_\xi$  except for  $\tau \in S$ .

**LEMMA A.5 (Hill 2008)** *Under Assumptions A and B and either  $H_0$  or  $H_1$ ,*

$$\begin{aligned} & \text{i. } \left| \hat{A}(\hat{\phi}) - A(\phi) \right|_1 = o_p(1), \text{ ii. } \sup_{\theta \in \Theta} \left| \hat{b}(\hat{\phi}, \theta) - b(\phi, \theta) \right|_1 = o_p(1) \\ & \text{iii. } \sup_{\theta \in \Theta} \sup_{\phi \in \Phi} \left| \hat{b}(\phi, \theta) - b(\phi, \theta) \right|_1 = o_p(1), \text{ iv. } \sup_{\theta \in \Theta} \left| \hat{V}(\theta) - V(\theta) \right|_1 = o_p(1) \end{aligned}$$

**LEMMA A.6 (Hill 2008)** *Under Assumption C for some positive constant  $K < \infty$*

$$\begin{aligned} & \text{i. } \left| A(\phi)^{-1} \right|_1 \leq K, \text{ ii. } \sup_{\theta \in \Theta_\xi} |b(\phi, \theta)|_1 \leq K, \\ & \text{iii. } \sup_{\theta \in \Theta_\xi} \left| \frac{\partial}{\partial \theta} b(\phi, \theta) \right|_1 \leq K, \text{ iv. } \left\| \sup_{\theta \in \Theta_\xi} |g_l(\theta)|_1 \right\|_{4+k} \leq K, \\ & \text{v. } \left\| \sup_{\theta \in \Theta_\xi} \left| \frac{\partial}{\partial \theta} g_l(\theta) \right|_1 \right\|_{4+k} \leq K, \\ & \text{vi. } \sup_{\theta \in \Theta_\xi} \left| V(\theta)^{-1/2} \right|_1^2 \leq K(pk+1) \left[ \inf_{\theta \in \Theta_\xi} \lambda_{\min}(V(\theta)) \right]^{-1} \leq K \\ & \text{vii. } \sup_{\theta \in \Theta_\xi} \left| \frac{\partial}{\partial \theta} \in V(\theta)^{-1/2} \right|_1 \leq K, \text{ for } l = 1 \dots pk+1. \end{aligned}$$

**Proof of Lemma A.1.**

**Step 1:** Properties of the  $l_1$ -norm and Minkowski's inequality imply

$$\begin{aligned} & \sup_{\theta \in \Theta_\xi} \left| \hat{V}(\theta)^{-1/2} \sqrt{n} \hat{s}_n(\hat{\phi}, 0, \theta) - V(\theta)^{-1/2} \sqrt{n} s_n(\phi, 0, \theta) \right|_1 \\ & \leq \sup_{\theta \in \Theta_\xi} \left| V(\theta)^{-1/2} \right|_1 \times \sup_{\theta \in \Theta_\xi} \left| \sqrt{n} \hat{s}_n(\hat{\phi}, 0, \theta) - \sqrt{n} s_n(\phi, 0, \theta) \right|_1 \\ & \quad + \sup_{\theta \in \Theta_\xi} \left| \hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2} \right|_1 \times \sup_{\theta \in \Theta_\xi} \left| \sqrt{n} s_n(\phi, 0, \theta) \right|_1 \\ & + \sup_{\theta \in \Theta_\xi} \left| \hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2} \right|_1 \times \sup_{\theta \in \Theta_\xi} \left| \sqrt{n} \hat{s}_n(\phi, 0, \theta) - \sqrt{n} s_n(\phi, 0, \theta) \right|_1. \end{aligned}$$

By Lemma A.6.vi  $\sup_{\theta \in \Theta_\xi} |V(\theta)^{-1/2}| < \infty$ , and  $\sup_{\theta \in \Theta_\xi} |\hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2}| = o_p(1)$  follows from Lemmas A.5.iv and A.6.vi, and Assumption C. Thus  $\sup_{\theta \in \Theta_\xi} |\hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2}| \times \sup_{\theta \in \Theta_\xi} |\sqrt{n}s_n(\phi, 0, \theta)| = o_p(1)$  by Lemmas A.2 and A.3, Cramér's Theorem and the mapping theorem.

**Step 2:** We need only show

$$\sup_{\theta \in \Theta_\xi} \sqrt{n} |\hat{s}_n(\hat{\phi}, 0, \theta) - s_n(\phi, 0, \theta)| = o_p(1).$$

By the mean-value-theorem there exists  $\phi_*(\theta) \in \mathbb{R}^{pk+1}$  such that  $|\phi_*(\theta) - \phi| \leq |\hat{\phi} - \phi|$  a.s. and

$$\sqrt{n}\hat{s}_n(\hat{\phi}, 0, \theta) = \sqrt{n}\hat{s}_n(\phi, 0, \theta) + \frac{\partial}{\partial \phi} \hat{s}_n(\phi_*(\theta), 0, \theta) \sqrt{n}(\hat{\phi} - \phi).$$

where for any  $\phi \in \Phi$

$$\begin{aligned} \frac{\partial}{\partial \phi} \hat{s}_n(\phi, 0, \theta) &= \frac{1}{n} \sum_{t=1}^n \frac{\partial}{\partial \phi} (y_t - f(\phi, \tilde{z}_{t-1})) \omega(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1}) \\ &= -\frac{1}{n} \sum_{t=1}^n \omega(\delta, \tilde{z}_{t-1}) \partial' f(\phi, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1}) = -\hat{b}(\phi, \theta). \end{aligned}$$

Standard nonlinear least squares algebra and Lemma A.5 show under  $H_0$

$$\left| \sqrt{n}(\hat{\phi} - \phi) - A(\phi)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{t-1}) \varepsilon_t \right| = o_p(1)$$

hence  $\hat{\phi} = \phi + O_p(1/\sqrt{n})$  by the martingale central limit theorem of BIERENS [1987: Theorem 29]. Lemma A.5.iii and  $|\phi_*(\theta) - \phi| \leq |\hat{\phi} - \phi| = O_p(1/\sqrt{n})$  imply  $\sup_{\theta \in \Theta_\xi} |\hat{b}(\phi_*(\theta), \theta) - b(\phi, \theta)| = o_p(1)$ , and

$$\begin{aligned} &\sup_{\theta \in \Theta_\xi} \sqrt{n} |\hat{s}_n(\hat{\phi}, 0, \theta) - \hat{s}_n(\phi, 0, \theta) + b(\phi, \theta)(\hat{\phi} - \phi)| \\ &\leq \sup_{\theta \in \Theta_\xi} \sqrt{n} |\hat{s}_n(\hat{\phi}, 0, \theta) - \hat{s}_n(\phi, 0, \theta) + \hat{b}(\phi_*(\theta), \theta)(\hat{\phi} - \phi)| \\ &\quad + \sup_{\theta \in \Theta_\xi} |\hat{b}(\phi_*(\theta), \theta) - b(\phi, \theta)| \sqrt{n} |\hat{\phi} - \phi| = o_p(1). \end{aligned}$$

Finally, use the identity

$$\sqrt{n}\hat{s}_n(\phi, 0, \theta) - b(\phi, \theta)A(\phi)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{t-1}) \varepsilon_t = \sqrt{n}s_n(\phi, 0, \theta)$$

and Cramér's Theorem to conclude

$$\begin{aligned} & \sup_{\theta \in \Theta_\xi} \sqrt{n} \left| \hat{s}_n(\hat{\phi}, 0, \theta) - s_n(\phi, 0, \theta) \right| \\ & \leq \sup_{\theta \in \Theta_\xi} \sqrt{n} \left| \hat{s}_n(\phi, 0, \theta) - s_n(\phi, 0, \theta) - b(\phi, \theta)(\hat{\phi} - \phi) \right| + o_p(1) \\ & = \sup_{\theta \in \Theta_\xi} \sqrt{n} \left| b(\phi, \theta) A(\phi)^{-1} \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{i-1}) \varepsilon_i - b(\phi, \theta)(\hat{\phi} - \phi) \right| + o_p(1) \\ & \leq \sup_{\theta \in \Theta_\xi} |b(\phi, \theta)| \times \sup_{\theta \in \Theta_\xi} \left| A(\phi)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \phi} f(\phi, \tilde{z}_{i-1}) \varepsilon_i - \sqrt{n}(\hat{\phi} - \phi) \right| + o_p(1) = o_p(1) \end{aligned}$$

**Proof of Lemma A.2.** Define for any  $h \in \mathbb{N}$ , any sequence  $\{\theta_1, \dots, \theta_h\}$ ,  $\theta_i \in \Theta$ ,  $r \in \mathbb{R}^{pk+1}$ ,  $r'r = 1$  and  $s \in \mathbb{R}^h$ ,  $s's = 1$ ,

$$\omega_t(r, s, \theta) = r' \sum_{i=1}^{k+1} s_i V(\theta_i)^{-1/2} g_t(\theta_i) \varepsilon_t$$

Thus  $\sqrt{nr'} \sum_{i=1}^h V(\theta_i)^{-1/2} s_n(\phi, 0, \theta_i) = 1/\sqrt{n} \sum_{i=1}^n w_t(r, s, \theta)$ . Clearly  $\{\omega_t(r, s, \theta), \mathfrak{F}_{t-1}\}$  forms a martingale difference sequence for any  $r'r = 1$  and  $s's = 1$  under the null by Assumption A and the  $\mathfrak{F}_{t-1}$ -measurability of  $g_t(\theta)$ . Under  $H_0$ ,  $1/\sqrt{n} \sum_{t=1}^n w_t(\theta) \rightarrow N(0, 1)$  in distribution pointwise in  $\Theta_\xi$  follows from BIERENS' [1991: Theorem 29] expanded version of McLEISH's [1974] martingale difference central limit theorem, cf. Lemma A.2.1. A Cramér-Wold device delivers the desired result.

**LEMMA A.2.1 (Hill 2008)** *Under the conditions of Lemma A.2, for each  $\theta \in \Theta_\xi$   $p \lim_{n \rightarrow \infty} 1/\sqrt{n} \sum_{t=1}^n w_t(\theta)^2 = \lim_{n \rightarrow \infty} 1/n \sum_{t=1}^n E[w_t(\theta)^2] = 1$ , and  $p \lim_{n \rightarrow \infty} \sum_{t=1}^n E|w_t(\theta)/\sqrt{n}|^{2+\kappa} = 0$  for some  $\kappa > 0$ .*

**Proof of Lemma A.3.** For any  $r \in \mathbb{R}^{pk+1}$ ,  $r'r = 1$ , write

$$r'V(\theta)^{-1/2} \sqrt{n} s_n(\phi, 0, \theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t \omega_t(r, \theta)$$



say, where  $\omega_t(r, \theta) = r'V(\theta)^{-1/2}g_t(\theta)$ . Using Lemma A.1 of BIERENS and PLOBERGER [1997] we need to show

$$(10) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E[\varepsilon_t^2 K_t^2] < \infty$$

$$(11) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E[\varepsilon_t^2 \omega_t(\theta_0)^2] < \infty$$

for at least one point  $\theta_0 \in \Theta$ , where  $K_t \equiv \sup_{\theta \in \Theta_\xi} |(\partial/\partial\theta)\omega_t(\theta)|_1$ .

Inequality (10) easily follows from Assumption A, the Cauchy-Schwartz and envelope inequalities, and Lemma A.6: for any  $\theta \in \Theta_\xi$

$$E[\varepsilon_t^2 \omega_t(r, \theta)^2] \leq \|\varepsilon_t\|_4^2 \times |r|_1^2 \times |V(\theta)^{-1/2}|_1^2 \times \left\| \sup_{\theta \in \Theta_\xi} |g_t(\theta)|_1 \right\|_4^2 \leq M < \infty$$

Now consider (9). By the Cauchy-Schwartz inequality

$$E[\varepsilon_t^2 K_t^2] \leq \|\varepsilon_t\|_4^2 \times |r|_1^2 \times \left\| \sup_{\theta \in \Theta_\xi} \left| \frac{\partial}{\partial\theta} \psi_t(\theta) \right|_1 \right\|_4^2$$

We need only show  $\left\| \sup_{\theta \in \Theta_\xi} |(\partial/\partial\theta)\psi_t(\theta)|_1 \right\|_4^2 \leq K < \infty$ . The  $(l, j)^{th}$ -component  $(\partial/\partial\theta)\psi_{t,j}(\theta)$  of the  $s \times pk+1$ -matrix  $(\partial/\partial\theta)\psi_t(\theta)$ , where  $s = q(pk+1) + pk+1$ , is exactly

$$\frac{\partial}{\partial\theta_l} \psi_{t,j}(\theta) = \sum_{i=1}^{pk+1} \frac{\partial}{\partial\theta_l} V(\theta)_{j,i}^{-1/2} \times g_{t,i}(\theta) + \sum_{i=1}^{pk+1} V(\theta)_{j,i}^{-1/2} \times \frac{\partial}{\partial\theta_l} g_{t,i}(\theta)$$

Use Minkowski's inequality repeatedly and Lemma A.6 to get

$$\left\| \sup_{\theta \in \Theta_\xi} \left| \frac{\partial}{\partial\theta} \psi_t(\theta) \right|_1 \right\|_4 \leq \left\| \sum_{l=1}^s \sum_{j=1}^{pk+1} \sup_{\theta \in \Theta_\xi} \left| \sum_{i=1}^{pk+1} \frac{\partial}{\partial\theta_l} V(\theta)_{j,i}^{-1/2} \times g_{t,i}(\theta) \right|_1 \right\|_4$$

$$\begin{aligned}
& + \left\| \sum_{l=1}^s \sum_{j=1}^{pk+1} \sup_{\theta \in \Theta_\xi} \left| \sum_{i=1}^{pk+1} \frac{\partial}{\partial \theta_l} V(\theta)_{j,i}^{-1/2} \times \frac{\partial}{\partial \theta_l} g_{t,i}(\theta) \right| \right\|_4 \\
& \leq (pk+1) \left\| \sum_{l=1}^s \sup_{\theta \in \Theta_\xi} \left| \sum_{i,j=1}^{pk+1} \frac{\partial}{\partial \theta_l} V(\theta)_{j,i}^{-1/2} \right| \sum_{i=1}^{pk+1} |g_{t,i}(\theta)| \right\|_4 \\
& + s(pk+1) \left\| \sum_{l=1}^s \sup_{\theta \in \Theta_\xi} \left| \sum_{i,j=1}^{pk+1} V(\theta)_{j,i}^{-1/2} \right| \sum_{l=1}^s \sum_{i=1}^{pk+1} \left| \frac{\partial}{\partial \theta_l} g_{t,i}(\theta) \right| \right\|_4 \\
& \leq (pk+1) \sum_{l=1}^s \sup_{\theta \in \Theta_\xi} \left| \frac{\partial}{\partial \theta} V(\theta)^{-1/2} \right|_1 \left\| \sup_{\theta \in \Theta_\xi} |g_t(\theta)|_1 \right\|_4 \\
& + s(pk+1) \sup_{\theta \in \Theta_\xi} \left| V(\theta)^{-1/2} \right|_1 \left\| \sup_{\theta \in \Theta_\xi} \left| \frac{\partial}{\partial \theta} g_t(\theta) \right|_1 \right\|_4 \leq K
\end{aligned}$$

**Proof of Lemma A.4.** Define  $\eta(\theta) \equiv E[\varepsilon_t \omega(\delta, \tilde{z}_{t-1}) F(\tau' \tilde{z}_{t-1})]$ . Minkowski's inequality and properties of the  $l_1$ -norm imply

$$\begin{aligned}
(12) \quad & \sup_{\theta \in \Theta_\xi} \left| \hat{V}(\theta)^{-1/2} \hat{s}_n(\hat{\phi}, 0, \theta) - V(\theta)^{-1/2} \eta(\theta) \right|_1 \\
& \leq \sup_{\theta \in \Theta_\xi} \left| V(\theta)^{-1/2} \right|_1 \times \sup_{\theta \in \Theta_\xi} \left| \hat{s}_n(\hat{\phi}, 0, \theta) - \eta(\theta) \right|_1 \\
& \quad + \sup_{\theta \in \Theta_\xi} \left| \hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2} \right|_1 \times \sup_{\theta \in \Theta_\xi} \left| \eta(\theta) \right|_1 \\
& \quad + \sup_{\theta \in \Theta_\xi} \left| \hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2} \right|_1 \times \sup_{\theta \in \Theta_\xi} \left| \hat{s}_n(\hat{\phi}, 0, \theta) - \eta(\theta) \right|_1
\end{aligned}$$

Each  $V(\theta)^{-1/2}$ ,  $\hat{V}(\theta)^{-1/2} - V(\theta)^{-1/2}$  and  $\eta(\theta)$  is uniformly  $l_1$ -bounded under Assumption A and Lemmas A.5-A.6.

Consider  $\hat{s}_n(\hat{\phi}, 0, \theta) - \eta(\theta)$ . By the mean-value-theorem

$$\begin{aligned} & \sup_{\theta \in \Theta_\xi} \left| \hat{s}_n(\hat{\phi}, 0, \theta) - \frac{1}{n} \sum_{t=1}^n \varepsilon_t \omega(\delta, \tilde{z}_{t-1}) F(\tau', \tilde{z}_{t-1}) \right| \\ & \leq \sup_{\theta \in \Theta_\xi} \left| \hat{b}(\hat{\phi}, \theta) - b(\hat{\phi}, \theta) \right| \times \left| \hat{\phi} - \phi \right| + \sup_{\theta \in \Theta_\xi} \left| b(\hat{\phi}, \theta) \right| \times \left| \hat{\phi} - \phi \right| = o_p(1) \end{aligned}$$

and under Assumption A Theorem 17 of BIERENS [1991] applies:

$$\sup_{\theta \in \Theta_\xi} \left| \frac{1}{n} \sum_{t=1}^n \varepsilon_t \omega(\delta, \tilde{z}_{t-1}) F(\tau', \tilde{z}_{t-1}) - \eta(\theta) \right| = o_p(1).$$

Now use the triangular inequality to conclude each term on right-hand-side of (11) is  $o_p(1)$ .

Finally, Lemma 1 guarantees for any  $\delta$  the set  $S = \{\tau \in \mathbb{R}^{p_{k+1}} : \eta(\theta) \neq 0 \text{ and } P(\tau' \tilde{z}_{t-1} \in \mathbb{R}_0) = 1\}$  has Lebesgue measure zero.

Because  $V(\theta)$  is uniformly positive definite in  $\Theta_\xi$  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  $\tau \in S$ .