Consistent LM-Tests for Linearity Against Compound Smooth Transition Alternatives*

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

April 1, 2004

Abstract

We develop tests of linearity that are consistent against a class of Compound Smooth Transition Autoregressive (CoSTAR) models of the conditional mean. Our method is an extension of the sup-test developed by Bierens (1990) and Bierens and Plobeger (1997), provides maximal power against popular STAR alternatives and is consistent against any deviation from the null hypothesis. Moreover, the test method can be extended to consistent tests of number of threshold regimes, flexible parametric forms, conditional homoscedasticity against linear or smooth transition GARCH, and causality tests of out-of-sample predictive accuracy.

Of particular note, we improve on Bierens's (1990) test theory by considering a vector conditional moment that leads to a sup-test statistic that is never degenerate under the alternative of functional mis-specification. Moreover, our test is a true test against smooth transition alternatives, whereas the universally employed polynomial regression test of Luukkonen *et al* (1988) and Teräsvirta (1994) requires the assumption that the true data generating mechanism is STAR.

A simulation study demonstrates that the suggested STAR sup-statistic renders a test with superlative empirical size and power attributes, in particular in comparison to the Bierens (1990) test, the neural test by Lee, White and Granger (1993), and specifically the polynomial regression test employed throughout the STAR literature. Finally, we apply the new tests to various macroeconomic processes.

*I would like to thank Dick van Dijk, Alvaro Escribano and Phil Rothman for comments on an earlier version, as well as participants at the 12th annual meeting of the Society for Nonlinear Dynamics in Economics. All errors, of course, are mine alone.

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

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

1. Introduction Smooth Transition Threshold Autoregressive (STAR) models have gained significant popularity in the economics and finance literatures as a means to transcend well known explanatory and forecasting limitations of linear models and binary regime/switching models. Suggested by Chan and Tong (1986a,b) to account for sluggish regime dynamics in many time series, Teräsvirta (1994) develops a composite theory of estimation, diagnostic checking and inference for smooth transition processes with exponential and logistic transition functions. See, also, Luukkonen *et al* (1988) and van Dijk *et al* (2000).

Tests for linearity against STAR alternatives, however, have received almost no attention in the theory literature, although a standard practice dominates the applied literature. Detailed below, under the null hypothesis of linearity several parameters, say γ, unique to the STAR process are unidentified, and therefore standard Lagrange Multiplier (LM) statistics cannot be directly computed. The approach in the theory literature has been to apply functionals to the unidentified LM statistics and analyze the non-standard limiting distribution by numerical integration or bootstrap techniques. For example, Davies (1977, 1987) suggests a sup-statistic over feasible γ, $\sup_{\gamma} LM(\gamma)$; Andrews and Ploberger (1994) and de Jong (1996) consider methods of averaging over the nuisance parameter space; and Hansen (1996) develops a bootstrap technique for estimating the asymptotic p-value for LM-type statistics when parameters are present that are undefined under the null hypothesis.

In the applied smooth transition literatures, by comparison, Luukkonen *et al* (1988), Saikkonen and Luukkonen (1988), Teräsvirta (1994), Hagerud (1997), Gonzalez-Rivera (1998), Escribano and Jorda (2000), Madieros and Veiga (2000), Rothman *et al* (2001) and others proscribe a truncated Taylor expansion approximation of the STAR model as a means to transcend the nuisance parameter and non-standard distribution dilemma. The technique leads to a simple polynomial auxiliary regression in the spirit of the RESET tests by Ramsey and Schmidt (1976) and Keenan (1985), and standard F-tests of parametric zero-restrictions in order to determine whether the process is linear, exponential or logistic STAR. The simplicity of the auxiliary regression makes this method employable in any standard econometrics software and therefore has appeal for quick applications.

Several fundamental problems associated with polynomial regressions exist, however, and are well known in the econometrics literature. First, by construction rejection of the test does not necessarily lead to a STAR model when the null of linearity is rejected. The polynomial regression technique provides maximal power against local polynomial alternatives, and therefore can at most be weakly associated with STAR nonlinearity. This issue is particularly relevant if we admit *any* functional alternative to explain the data provided linearity is found inadequate: polynomial nonlinearity is known *not* to be "generically comprehensive" (Stinchcombe and White, 1998) in the sense that if a linear model is mis-specified, additive polynomial terms may not improve the model fit. This shortcoming of classic weight-based moment condition specification tests is well known in the inference theory and artificial neural network literatures: see, e.g., Davies (1977), Holley (1982), Bierens (1990), Kuan and White (1994) and Stinchcombe and White (1998).

Second, for STAR tests a nuisance "delay" parameter d (see Section 2) still exists in the polynomial regression. Teräsvirta (1994) and many others suggest performing the polynomial regression tests for various delay values, and selecting that d which generates the lowest test p-value. This is mathematically equivalent to maximizing an LM sup-statistic over possible d-values, a statistic known to have a non-standard limit distribution under the null. Nevertheless, in the applied literature the standard practice is simply to employ p-values derived from the chi-squared distribution.

Third, a fundamental assumption of the Taylor expansion method is that the true data generating mechanism is a smooth transition autoregression (see Teräsvirta, 1994). Thus, the resulting polynomial test is not a true test against STAR alternatives, per se, because a STAR functional form is assumed a priori. Rather, under the assumption the process is a STAR, the test is employed to detect whether the process is exponential, logistic, or simply a linear STAR process (i.e. identical regimes). Therefore, to date there does not exist a true test against smooth transition alternatives, without the necessity of a prior. Provided such a test can be formulated, and once STAR nonlinearity is detected, we can then exploit the extant procedure (e.g. Escribano and Jorda, 1999) for determining which type of STAR mechanism prevails.

In this paper, we expand the well established theory of conditional moment tests, cf. Newey (1985), Bierens (1982, 1990) and Bierens and Ploberger (1997), to the problem of testing linearity against finite-order STAR alternatives[1]. We develop a test in the spirit of Bierens' (1990) conditional moment sup-test that leads to a consistent (asymptotic power of one) test of linearity with maximal power against smooth transition alternatives. In order to make Bierens' (1990) theory operable, we need to augment traditionally scalar transition function variables to allow for multivariate transition processes with multiple thresholds: the result is a compound STAR model (CoSTAR) with a multiplicative transition mechanism. We do not require the assumption that the true process is STAR: our test is a test of whether STAR nonlinearity provides a better functional approximation to the underlying data degenerating mechanism. We do not consider an Integrated Conditional Moment test, cf. Bierens (1982) and Bierens and Ploberger (1997), because of the computational burden associated with numerical integration and the degree of arbitrariness associated with the choice of weight function. See, also, Corradi and Swanson (2002). Our intention is the development of a consistent, non-degenerate LM sup-test that can be relatively easily computed.

Consistent *non-parametric* moment based tests, however, exist. Zheng (1996) develops an analogue to the Bierens (1990) test, based on measuring the distance between the null conditional mean and a kernel estimator of the conditional mean. Hong and White (1995) similarly measure the distance between a para-

[1]For compactness, we only consider finite-order scalar autoregressive models in order to reduce notation and simplify asymptotic theory. However, extensions are straightforward for smooth transition VARMAX processes and GARCH models of volatiity.

metric null model and non-parametric general alternative functional form, where the non-parametric estimators are based on Fourier series and regression splines. See, also, Lee (1988), Yatchew (1992) and Wooldridge (1992). Non-parametric methods are suitable for testing whether a particular functional specification (e.g. AR or STAR) is correct with probability one, but cannot provide a better parametric specification if the null specification is false. Our test method is parametric in the sense that we test a null model with a specific class of alternatives (STAR) in mind, although traditional models of smooth transition nonlinearity asymptotically have a probability one of improving the fit of *any* mis-specified model.

It turns out that adjusting the conditional moment test weight to account for STAR nonlinearity essentially eradicates the set of nuisance parameters γ for which Bierens' (1990) test statistic is degenerate, an important improvement over original results established in Bierens (1990) and de Jong (1996). In their work, the LM statistic covariance matrix must be assumed to be nonsingular, and therefore the statistic is assumed to be nondegenerate: in the present work, no such assumption is required. Moreover, by utilizing LM-test theory, we do not require estimation of an alternative model. Likelihood Ratio tests, by comparison, require estimation of both the linear and STAR models, and it is well known that sharp estimates of STAR transition function parameters are difficult to obtain: see Teräsvirta (1994)[2]. Finally, our methods can be straightforwardly extended to encompass consistent tests of number of threshold regimes; conditional homoscedasticity against linear or nonlinear GARCH, or linear GARCH against nonlinear GARCH; and out-of-sample predictor accuracy, a la Corradi and Swanson (2002).

In a broad simulation study we provide concrete evidence that a STAR sup-test based on a scalar transition variable, and a CoSTAR sup-test with compound transition function dominate parametric tests by Bierens (1990), Lee *et al* (1993) and Teräsvirta (1994). Moreover, by combining polynomial and smooth transition weights multiplicatively and randomly selecting all nuisance parameters, the resulting statistic often dominates all tests, including the STAR and CoSTAR sup-tests. This demonstrates a multiplicative hybrid of polynomial and smooth transition terms may be a powerful tool for the detection of arbitrary nonlinearity and specifically for detecting STAR nonlinearity. Finally, our simulation is less restrictive than previous such studies (e.g. Luukonnen *et al*, 1988; Teräsvirta, 1994; Skalin, 1998): we do not fix slope parameters, and therefore control for the fact that conveniently chosen parameters may bias test results.

The rest of this paper contains the following topics. In Section 2, we detail STAR model specifications, and develop the test in Section 3. We compare the consistent STAR test with existing tests in a simulation study in Section 4. Finally, in Section 5 we demonstrate the use of the methods developed here

[2]Skalin (1998), by comparison, studies the comparative performances of the polynomial regression test and a Likelihood Ratio sup-test, where Hansen's (1996) bootstrap method for approximating the p-value is employed. From a limited simulation study with fixed STAR parameters, the auther concludes the polynomial test dominates.

on money supply, output, prices, the Treasury bill and commercial paper rates, and conclude with parting comments in Section 6.

Throughout we maintain the following notation conventions. $\rightarrow$ denotes convergence in probability; $\Longrightarrow$ denotes weak convergence with respect to finite dimensional distributions. $|\cdot|$ denotes the Euclidean norm for real-valued vectors, and the matrix norm for real-valued square matrices: $|x| = [Tr\,(x'x)]^{1/2}$. In all cases, m_i denote arbitrary weakly positive integers for arbitrary integer indices i, whose value is understood in context. For arbitrary k-vectors a and x, vector powers x^a are understood to represent $(x_1^{a_1}, ..., x_k^{a_k})'$.

2. Smooth Transition Threshold Autoregressive Models Consider a time series process $\{y_t\}$ defined in $L_2(\Omega, \mathfrak{F}_t, P)$, with $\mathfrak{F}_t$ an increasing σ-field, regressors $x_{it} = (1, y_{t-1}, ..., y_{t-p_i})'$, $i = 1, 2$, an innovations process $\{u_t\}$, and denote by v_t a stochastic scalar to be defined below. The class of L-regime STAR processes is represented as

$$y_t = \phi_1' x_{1t} + \sum\nolimits_{i=2}^{L} \phi_i' x_{it} F(v_{t-d_i}, \gamma_i, c_i) + u_t, \tag{1}$$

for some transition function $F_t(d_i, \gamma_i, c_i) = F(v_{t-d_i}, \gamma_i, c_i) : \mathbb{R}^3 \rightarrow [0, 1]$, transition scale $\gamma_i > 0$, threshold variable v_t, threshold c_i, and delay parameter d_i. In this setting, regressors are not restricted to be identical across regimes, hence the orders p_i need not equate. The transition function is assumed to be twice continuously differentiable in γ_i and c_i, and bounded $0 \leq F_t(d_i, \gamma_i, c_i) \leq 1$.

Following the Self Exciting Threshold Autoregression (SETAR) literatures (see, e.g., Tong, 1990), typically the threshold variable v_{t-d_i} is restricted to be some lag of y_t, say y_{t-d_i} In order to account for "lower" versus "upper" and "inner" versus "outer" regimes, Luukkonen *et al* (1988) and Teräsvirta (1994) suggests candidate transition functions include the logistic and exponential with self-exciting threshold variable $v_{t-d_i} = y_{t-d_i}$

$$F_t(d_i, \gamma_i, c_i) = \frac{1}{1 + e^{-\gamma_i (y_{t-d_i} - c_i)}}, \quad F_t(d_i, \gamma_i, c_i) = e^{-\gamma_i (y_{t-d_i} - c_i)^2}. \tag{2}$$

Other transition functions F_t discussed in the literature include the Gaussian probability distribution function: see Chan and Tong (1986b) for the so-called normal STAR model. See Teräsvirta (1994) and van Dijk *et al* (2000) for details on the various properties of the above transition functions and details on estimation.

2.1 STAR Limitations Several deficiencies noticeably persist in standard STAR representations, both from a forecasting perspective and the perspective of test statistic asymptotics. First, the threshold variable v_t is typically assumed to be scalar-valued, even in smooth transition VAR applications (e.g. Rothman *et al*, 2001). However, a consistent test of linearity will require all variables contained in x_{2t} to be included in the transition function: see Section

3, below. The restriction that only one process be allowed to embody the threshold mechanism essentially amounts to (untested) zero restrictions on embedded transition function vector-parameters for vector threshold processes.

Second, even if we allow for multiple threshold processes to enter the transition function, say, $v_t = x_{2t} = (1, y_{t-1}, ..., y_{t-p})'$ for some p, we are not guaranteed a set of unique variable-specific thresholds, say c_j, for each y_{t-j}. This problem is particularly acute for non-exponential transition functions. For example, the LSTAR transition function with multivariate threshold variable would be

$$F_t(\gamma) = \frac{1}{1 + \exp(-\gamma' v_t)} = \frac{1}{1 + \exp(-\gamma_0 - \sum_{i=1}^{k-1} \gamma_i y_{t-i})} \tag{3}$$

where $v_t = (1, y_{t-1}, ..., y_{t-p})'$, and $\gamma \in \Gamma$, a compact subset of $\mathbb{R}^k$, $k = p + 1$, where $\gamma = (\gamma_0, ..., \gamma_{k-1})'$. In order to parameterize (3) in a manner similar to (2), notice that

$$F_t(\gamma) = \frac{1}{1 + \exp(-\gamma_0 - \sum_{i=1}^{k-1} \gamma_i y_{t-i})} = \frac{1}{1 + \exp(-\sum_{i=1}^{k-1} \gamma_i \left[y_{t-i} - c_i\right])} \tag{4}$$

where c_i is defined by the identity

$$-\gamma_0 = \sum_{i=1}^{k-1} \gamma_i c_i. \tag{5}$$

It is clear from (5) that individual threshold values c_i cannot be identified in general for each threshold variable y_{t-i}. However, restricting $c_i = c$ for each i, a "universal" threshold c can be uniquely identified from the k-transition parameters, $c = -\gamma_0 / \sum_{i=1}^{k-1} \gamma_i$ provided $\sum_{i=1}^{k-1} \gamma_i \neq 0$. In practice, however, it may be difficult to obtain an intuitive environment in which the threshold variable is the linear combination $\sum_{i=1}^{k-1} \gamma_i y_{t-i}$ rather than the individual delayed processes y_{t-i}.

2.2 Compound STAR Model We solve the problems associated with scalar threshold processes and non-unique thresholds c_j by allowing for multivariate threshold variables and compound (multiplicative) transition functions. The Compound Smooth Transition AR model (CoSTAR) has the general form

$$y_t = \phi_1' x_{1t} + \sum_{i=2}^{L} \phi_i' x_{it} F^{(i)}(v_t, \gamma^{(i)}) + u_t, \tag{6}$$

where $F^{(i)}(\cdot)$ denotes the i^{th}-regime's compound transition function

$$F^{(i)}(v_t, \gamma^{(i)}) = \prod_{j=1}^{p} F_j^{(i)}(v_{t,j}, \gamma_j^{(i)}). \tag{6'}$$

The model allows for any sequence $F_j^{(i)}(v_{t,j}, \gamma_j^{(i)})$ of appropriate transition functions to transfer information concerning regime dynamics: because each transition function satisfies $0 \leq F_j^{(i)}(\cdot) \leq 1$, the compound transition function likewise satisfies $0 \leq F^{(i)}(\cdot) \leq 1$.

For logistic CoSTAR processes, define $v_{t,j} = (1, y_{t-j})'$, $\gamma_j^{(i)} = (\gamma_{j,0}^{(i)}, \gamma_{j,1}^{(i)})'$, $v_t = (1, y_{t-1}, ..., y_{t-p})'$ and $\gamma^{(i)} = (\gamma_1^{(i)\prime}, ..., \gamma_p^{(i)\prime})'$, a $p \times 2$ matrix. For exponential CoSTAR processes, define $v_{t,j} = (1, y_{t-j}, y_{t-j}^2)'$, and $\gamma_j^{(i)} = (\gamma_{j,0}^{(i)}, \gamma_{j,1}^{(i)}, \gamma_{j,2}^{(i)})'$, thus $\gamma^{(i)} = (\gamma_1^{(i)\prime}, ..., \gamma_p^{(i)\prime})'$, a $p \times 3$ matrix.

Below we establish several representations of the compound transition function which will be useful in proving consistency of test statistic.

2.2.1 LoCoSTAR

Consider the logistic CoSTAR transition function where each $F_j^{(i)}(v_{t,j}, \gamma_j^{(i)})$ is logistic. For simplicity, set $L = 2$. Then $F^{(1)}(v_t, \gamma^{(1)}) = F(v_t, \gamma)$ satisfies

$$\begin{aligned}
F(v_t, \gamma) &= \prod\nolimits_{j=1}^{p} F_j(v_{t,j}, \gamma_j) = \prod\nolimits_{j=1}^{p} \left(\frac{1}{1 + \exp(-\gamma_j' v_{t,j})} \right) \qquad (7) \\
&= \prod\nolimits_{j=1}^{p} \left(\frac{\exp(\gamma_j' v_{t,j})}{1 + \exp(\gamma_j' v_{t,j})} \right) \\
&= \frac{\exp\left(\sum_{j=1}^{p} \gamma_j' v_{t,j} \right)}{\prod_{j=1}^{p} \left[1 + \exp(\gamma_j' v_{t,j}) \right]} \\
&= h(v_t, \gamma) \exp\left(\sum\nolimits_{j=1}^{p} \gamma_j' v_{t,j} \right) \\
&= h(v_t, \gamma) \exp\left(\sum\nolimits_{j=1}^{p} \gamma_{j,1} [y_{t-j} - c_j] \right),
\end{aligned}$$

say, where we define $c_j \equiv -\gamma_{j,0}/\gamma_{j,1}$ provided $\gamma_{j,1} \neq 0$, and $h(v_t, \gamma) \equiv \prod_{j=1}^{p}[1 + \exp(\gamma_{j,1}[y_{t-j} - c_j])]^{-1}$ is [0,1]-bounded. In order for each $F_j(v_{t,j}, \gamma_j)$ to satisfy the assumed interpretive and boundedness assumptions common in the STAR literature, in general we assume at least one $\gamma_{j,1} > 0$. Thus, $c_j = 0$ *if and only if* $\gamma_{j,0} = 0$, and if $\gamma_{j,1} = 0$ then c_j is simply not defined. The $F(v_t, \gamma)$ representation $h(v_t, \gamma) \exp(\sum_{j=1}^{p} \gamma_{j,1}[y_{t-j} - c_j])$ will be useful for deriving the first result, Lemma 1, below.

2.2.2 ECoSTAR

Consider the exponential CoSTAR transition function where each $F_j^{(i)}(v_{t,j}, \gamma_j^{(i)})$ is exponential. Again, set $L = 2$. Then $F^{(1)}(v_t, \gamma^{(1)}) = F(v_t, \gamma)$ satisfies

$$\begin{aligned}
F(v_t, \gamma) &= \prod\nolimits_{j=1}^{p} F_j(v_{t,j}, \gamma_j) \qquad (8) \\
&= \prod\nolimits_{j=1}^{p} \exp(-\gamma_j' v_{t,j}) \\
&= \exp\left(-\sum\nolimits_{j=1}^{p} \left[\gamma_{j,0} + \gamma_{j,1} y_{t-j} + \gamma_{j,2} y_{t-j}^2 \right] \right) \\
&= \exp\left(-\sum\nolimits_{j=1}^{p} \gamma_{j,2} \left[y_{t-j} - c_j \right]^2 \right) \\
&= \exp\left(-\sum\nolimits_{j=1}^{p} \gamma_{j,2} y_{t-j}^2 \right) \exp\left(-\sum\nolimits_{j=1}^{p} \left[\gamma_{j,0} + \gamma_{j,1} y_{t-j} \right] \right) \\
&= h(v_t, \gamma) \exp\left(-\sum\nolimits_{j=1}^{p} \left[\gamma_{j,0} + \gamma_{j,1} y_{t-j} \right] \right)
\end{aligned}$$

where each γ_j and c_j is restricted to solve $c_j \equiv -\gamma_{j,1}/2\gamma_{j,2}$, $\gamma_{j,0} = \gamma_{j,2}c_j^2$, provided $\gamma_{j,2} \neq 0$. Hence, we are effectively restricting each $\gamma_j = (\gamma_{j,0}, \gamma_{j,1}, \gamma_{j,2})'$ to satisfy $\gamma_{j,0} = \gamma_{j,1}^2/4\gamma_{j,2}$. We define $h(v_t, \gamma) \equiv \exp(-\sum_{j=1}^{p} \gamma_{j,2} y_{t-j}^2)$: this latter parameterization of the ECoSTAR model will be useful for Lemma 1, below. By convention, we assume at least one $\gamma_{j,2} > 0$, hence $h(v_t, \gamma)$ is [0,1]-bounded. The identities $c_j = -\gamma_{j,1}/2\gamma_{j,2}$ and $\gamma_{j,0} = \gamma_{j,2}c_j^2$, with $\gamma_{j,2} > 0$, implies $c_j = 0$ *if and only if* $\gamma_{j,0} = \gamma_{j,1} = 0$. If $\gamma_{j,2} = 0$, then c_j is not defined, however as a convention we enforce $\gamma_{j,0} = \gamma_{j,1} = 0$.

Notice that from line 5 of equation (7) and line 4 of equation (8), the compound transition function can be written identically as $F(v_t, \gamma) = F(v_t, \tilde{\gamma}, c)$ where $\tilde{\gamma} = \gamma_1$ in the logistic model and $\tilde{\gamma} = \gamma_2$ in the exponential model. Thus, in practice the restriction that $\gamma_j = (\gamma_{j,1}^2/4\gamma_{j,2}, \gamma_{j,1}, \gamma_{j,2})'$ can be straightforwardly satisfied by considering the more parsimonious parameterization $F(v_t, \tilde{\gamma}, c)$: we simply restrict each $\tilde{\gamma}_j > 0$ and each c_j to be within the observable range of y_{t-j}: see Section 4.

Moreover, we no longer need to define a "delay" parameter in multivariate transition functions. The above representations are general enough to include the scalar-case with the standard scalar threshold variate y_{t-d}, where it is typically assumed that $1 \leq d \leq p$ (see, e.g., Teräsvirta, 1994). For example, in the logistic case provided $\gamma_{1,j} = 0$, $j \neq d$, and $\gamma_{1,d} > 0$ for some $1 \leq d \leq p$, then the LoCoSTAR model reduces to the traditional LSTAR representation where d then denotes the traditional "delay" parameter.

3. Consistent Tests of Linearity against STAR Alternatives In this section, we develop the limit theory for consistent tests of linearity. For simplicity, we assume the CoSTAR model has two regimes ($L = 2$), and regressors are identical across regimes, $x_{1t} = x_{2t} = x_t$, a $k \times 1$ random vector:

$$y_t = \phi_1' x_t + \phi_2 x_t F_t(\gamma) + u_t, \tag{9}$$

where $F_t(\gamma) = F(v_t, \gamma)$ and v_t is defined below (6).

3.1 Null Hypothesis of Linearity The specific null hypothesis of linearity in a STAR framework, cf. (9), states

$$H_0 : \phi_2 = 0. \tag{10}$$

Under the null hypothesis, therefore,

$$y_t = \phi_1' x_t + \epsilon_t, \tag{11}$$

where $\epsilon_t = u_t$. In order to fix ideas, we assume the objective model is the conditional expectations, $E[y_t|\mathfrak{F}_{t-1}]$, where $\mathfrak{F}_t$ denotes an increasing σ-algebra induced by the information $\{y_{t-i}, x_{t-i}\}_{i=0}^{\infty}$. In autoregressive settings without exogenous information, it suffices for $\mathfrak{F}_{t-1}$ to be induced by $y_{t-1}, y_{t-2}, ..$: $\mathfrak{F}_{t-1}$ $= \sigma(y_{t-1}, y_{t-2}, ...)$. Under the null of linearity, the best L_2-predictor satisfies

$E\left[y_t|\mathfrak{F}_{t-1}\right] = \phi_1' x_t$, which is tantamount to $E\left[\epsilon_t|\mathfrak{F}_{t-1}\right] = 0$. Consult Appendix 1 for all maintained assumptions.

Denote by Φ a closed compact subset of $\mathbb{R}^k$. We state the fundamental hypotheses in a useful general format:

$$
\begin{aligned}
H_0 &: P\left(E\left[y_t - \phi' x_t|\mathfrak{F}_{t-1}\right] = 0\right) = 1, \text{ for some } \phi \\
H_1 &: \sup_{\phi \in \Phi} P\left(E\left[y_t - \phi' x_t|\mathfrak{F}_{t-1}\right] = 0\right) < 1.
\end{aligned} \tag{12}
$$

Under H_0 the linear model is *almost surely* correctly specified such that ϵ_t forms a martingale difference sequence. The alternative H_1 is simply that the null model is mis-specified, hence the alternative embraces any deviation from the null.

3.2 Lagrange Multiplier Framework: Null Score with STAR Alternatives Employing the *mean-squared-error* criterion, the score $s_n(\phi_1, \phi_2, \gamma)$ associated with (9) is exactly

$$
s_n(\phi_1, \phi_2, \gamma) = \left[\begin{array}{c} n^{-1}\sum_{t=1}^n u_t x_t \\ n^{-1}\sum_{t=1}^n u_t x_t F_t(\gamma) \\ n^{-1}\sum_{t=1}^n u_t \phi_2' x_{2t} \partial F_t(\gamma)/\partial\gamma \end{array}\right]. \tag{13}
$$

Denoting by $s_n(0, \gamma)$ the score evaluated under the null, $\phi_2 = 0$, we obtain

$$
s_n(0, \gamma) = \left[\begin{array}{c} n^{-1}\sum_{t=1}^n \epsilon_t x_t \\ n^{-1}\sum_{t=1}^n \epsilon_t x_t F_t(\gamma) \end{array}\right], \quad 2k \times 1, \tag{14}
$$

where $\epsilon_t = y_t - \phi_1' x_t$, the null innovations. Using the least squares estimator of ϕ_1 under the null $H_0 : \phi_2 = 0$, and denoting by $\hat{\epsilon}_t$ the resulting residuals $y_t - \hat{\phi}_1' x_{1t}$, the estimated score under the null compactly reduces to

$$
\begin{aligned}
\hat{s}_n(0, \gamma) &= n^{-1}\sum_{t=1}^n \hat{\epsilon}_t x_t F_t(\gamma) \\
&= n^{-1}\sum_{t=1}^n \hat{\epsilon}_t z_t(\gamma), \quad k \times 1,
\end{aligned} \tag{15}
$$

say, due to least squares orthogonality, $n^{-1}\sum_{t=1}^n \hat{\epsilon}_t x_t = 0$.

3.3 STAR Conditional Moments Denote by $F(x, \gamma)$ the exponential $e^{\gamma' x}$ or logistic $(1 + e^{\gamma' x})^{-1}$ function. The following lemma is a direct extension of Lemma 1 of Bierens (1990) to account for moment condition vector weights applicable for STAR models. While it is straightforward to extend Bierens' (1990) theory to account for STAR moment conditions, the optimal choice of weight, as it turns out, is non-trivial: a test statistic which is non-degenerate under all deviations from the null can be generated from specific weights that precisely coincides with STAR nonlinearity: see Lemma 3.

In the following, define the closed, bounded compact parameter subspaces $\Gamma \subseteq \mathbb{R}^k$ and $\Delta \subseteq \mathbb{R}^k$. Moreover, assume $h(x_t, \delta)$ is any bounded, continuous mapping from $\mathbb{R}^k \times \Delta$ to $\mathbb{R}^k$, measurable with respect to $\mathfrak{F}_{t-1}$, such that

$P(\inf_{\delta\in\Delta}|h(x_t,\delta)|>0)=1$ and $\sup_{\delta\in\Delta}|h(x_t,\delta)|<\infty$ with probability one. For example, $h(x_t,\delta)$ cannot be $\delta' x_t$ because $\delta' x_t=0$ with probability one when $\delta=0$; however $h(x_t,\delta)$ can be $\exp(-\delta' x_t^2)$ for bounded real vectors $\delta\geq 0$.

Lemma 1 *Let ϵ be a random variable satisfying $E|\epsilon|<\infty$, and let x be an $\mathfrak{F}$-measurable bounded vector in $\mathbb{R}^k$, $0<k<\infty$, $P(|x|>0)=1$, such that $P[E(\epsilon|x)=0]<1$. The sets*

$$S_i=\{\gamma\in\Gamma:\sup_{\delta\in\Delta}|E[\epsilon h_i(x,\delta)F(x,\gamma)]|=0\},\ i=1...k, \tag{16}$$

have Lebesgue measure zero, and are not dense in $\mathbb{R}^k$.

Proof. All proofs are contained in Appendix 3. ■

Remark 1: Conditioning on x in $E(\epsilon|x)$ is equivalent to conditioning on any bounded, measurable, one-to-one function of x, say $\Psi(x):\mathbb{R}^k\to\mathbb{R}^k$, since any such functional induces the same σ-field as x: see Billingsley (1995: Theorem 5.1).

Remark 2: Although we restrict attention to STAR models proper in which the transition functions are bounded $0\leq F_t(\gamma)\leq 1$ and twice continuously differentiable in γ, all results in this paper hold for essentially any real analytic function defined on a compact subset on which $\gamma'\Psi(x)$ takes it values: see Stinchcombe and White (1998).

Remark 3: Using arguments similar to Theorem 1 of de Jong (1996), we may extend Lemma 1 to include any strictly stationary time series process $\{y_t,x_t\}$ with innovations ϵ_t that are martingale difference sequences under the null.

Remark 4: If x in (16) is infinite dimensional, as in the case of covariance stationary ARMA processes, further regulatory conditions on the serial dependence in x must be met: see de Jong (1996).

Remark 5: By setting $h(x,\delta)=x$, Lemma 1 demonstrates a vector moment condition implied by a STAR process leads to sets $S_i=\{\gamma\in\Gamma:E[\epsilon x F(x,\gamma)]=0\}$, $i=1...k$, with Lebesgue measure zero under H_1. Thus, while the moment condition $E[\epsilon h(x,\delta)F(x,\gamma)]$ is sensitive to any deviation from the null hypothesis, we can arbitrarily direct maximal power toward an implicit alternative of smooth transition nonlinearity.

Because we only require $h(x_t,\delta)$ to be an $\mathfrak{F}_{t-1}$-measurable, continuous, bounded mapping into $\mathbb{R}^k$, Lemma 1 holds for compound transition functions (7)-(8) with $\Delta=\Gamma$, $\delta=\gamma$. For example, in the 2-regime LoCoSTAR case $v_t=x_t=(1,y_{t-1},\ldots,y_{t-p})'$ and we define

$$\begin{aligned}\tilde{h}(x_t,\gamma) &\equiv x_t\left(\prod_{j=1}^{p}\left[1+\exp(\gamma_{j,1}[y_{t-j}-c_j])\right]\right)^{-1} \\ \tilde{F}(x_t,\gamma) &\equiv \exp\left(\sum_{j=1}^{p}\gamma_{j,0}+\gamma_{j,1}y_{t-j}\right)\end{aligned} \tag{17}$$

where $c_j \equiv -\gamma_{j,0}/\gamma_{j,1}$, $\gamma_{j,1} > 0$, thus

$$\begin{aligned}\tilde{h}(x_t,\gamma)\tilde{F}(x_t,\gamma) &= x_t \frac{\exp\left(\sum_{j=1}^{p}\gamma_{j,0}+\gamma_{j,1}y_{t-j}\right)}{\prod_{j=1}^{p}\left[1+\exp(\gamma_{j,1}[y_{t-j}-c_j])\right]} \\ &= x_t F(v_t,\gamma),\end{aligned} \tag{18}$$

which is identically the LoCoSTAR nonlinear term with logistic compound transition function $F(v_t,\gamma)$, cf. (6) and (7). In the ECoSTAR case, define $\tilde{h}(v_t,\gamma) \equiv x_t \exp(-\sum_{j=1}^{p}\gamma_{j,2}y_{t-j}^2)$ and $\tilde{F}(x_t,\gamma) \equiv \exp(-\sum_{j=1}^{p}[\gamma_{j,0}+\gamma_{j,1}y_{t-j}])$, hence $\tilde{h}(x_t,\gamma)\tilde{F}(x_t,\gamma) = x_t F(v_t,\gamma)$, the ECoSTAR nonlinear term, cf. (6) and (8).

The above re-parameterization $\tilde{h}(x_t,\gamma)\tilde{F}(x_t,\gamma)$ of $x_t F(v_t,\gamma)$ for CoSTAR models is unavoidably important with respect to Lemma 1, in particular for the ECoSTAR model. By remark 1 of Lemma 1, the sets S_i have Lebesgue measure zero for any $\mathfrak{F}_{t-1}$-measurable, bounded, one-to-one mapping $\Psi(x_t) : \mathbb{R}^k \rightarrow \mathbb{R}^k$ in $F(\Psi(x_t),\gamma)$. For ECoSTAR models, however, we require $\Psi(x_t) = (1, y_{t-1}, ..., y_{t-p}, y_{t-1}^2, ..., y_{t-p}^2)'$ in $F(\Psi(x),\gamma)^3$, a mapping from $\mathbb{R}^k$ to $\mathbb{R}^{k+p}$. Such a mapping does not generate the same Borel field as $(1, y_{t-1}, ..., y_{t-p})'$ due simply to the dimensionality problem. Likewise, if we simply use $\Psi(x) = (y_{t-1}^2, ..., y_{t-p}^2)'$ in $F(\Psi(x_t),\gamma)$ (i.e. $\gamma_{j,0} = \gamma_{j,1} = 0$), the mapping is not one-to-one.

However, by Lemma 1 the sets S_i have Lebesgue measure zero for any bounded function $\tilde{h}(x_t,\gamma)$, non-zero with probability one. For each CoSTAR re-parameterized weight $\tilde{h}(x_t,\gamma)\tilde{F}(x_t,\gamma)$, the function $\tilde{h}(x_t,\gamma)$ is bounded by the assumptions x_t and Γ are bounded, and each $\tilde{F}(x_t,\gamma)$ involves a simple one-to-one mapping with respect to $x_t = (1, y_{t-1}, ..., y_{t-p})'$: the LoCoSTAR $\tilde{F}(x_t,\gamma)$ effectively uses $\Psi(x) = -x$, and the ECoSTAR $\tilde{F}(x_t,\gamma)$ effectively uses $\Psi(x) = x$. Thus, by re-parameterizing the smooth transition weights $x_t F(v_t,\gamma)$ into $\tilde{h}(x_t,\gamma)\tilde{F}(x_t,\gamma)$, Lemma 1 applies for compound STAR weights provided $\tilde{F}(x_t,\gamma)$ incorporates a measurable, bounded one-one mapping with respect to the argument x_t. In the ECoSTAR model, therefore, in order for Lemma 1 to hold we require $\gamma_{j,0} \neq 0$ and $\gamma_{j,1} \neq 0$ for at least one $j = 1...p$: if $\gamma_{j,0} = \gamma_{j,1} = 0$ for all j, then $\tilde{F}(x_t,\gamma) = 1$ is degenerate and the weight $\tilde{h}(x_t,\gamma)\tilde{F}(x_t,\gamma)$ reduces to $\tilde{h}(v_t,\gamma) \times 1 \equiv x_t \exp(-\sum_{j=1}^{p}\gamma_{j,2}y_{t-j}^2) = x_t F(\Psi(x_t),\gamma)$, where $\Psi(x_t) = (y_{t-1}^2, ..., y_{t-p}^2)'$ is not one-to-one. By requiring $\gamma_{j,0} \neq 0$ and $\gamma_{j,1} \neq 0$ for at least one j, we effectively require at least one threshold c_j to be non-zero.

For matters of hypothesis testing, a non-zero restriction $c_j \neq 0$ in and of itself is not important: we only seek evidence that the null specification is false and whether *some* form of smooth transition alternative can be used to improve the model fit. By Lemma 1 consistency of the smooth transition moment condition $\sqrt{n}\hat{s}_n(0,\gamma)$ under the alternative fails to hold for countably many $\gamma = (\gamma_1, ..., \gamma_p)'$, and therefore for only countably many restricted $\gamma_j = (\gamma_{j,1}^2/4\gamma_{j,2}, \gamma_{j,1}, \gamma_{j,2})'$ with $\gamma_{j,1} \neq 0$ and $\gamma_{j,2} > 0$ (hence $\gamma_{j,0} \neq 0$ and $c_j \neq 0$), thus the implied restrictions will not diminish asymptotic power. In practice we

[3] See line 3 of equation (8).

simply enforce $\gamma_{j,2} > 0$, and confine feasible c_j to be within some range of y_{t-j}, excluding 0: see Section 4.

3.4 Consistent Test of Linearity Against STAR Alternatives

Recall the least squares score under the null with possibly compound transition weights $F_t(\gamma) = F(x_t, \gamma)$:

$$\hat{s}_n(0,\gamma) = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{\phi}_1' x_t)x_t F_t(\gamma). \tag{19}$$

All assumptions are detailed in Appendix 1. Under Assumptions 1-4, the null score converges in law to a multivariate normal random vector with covariance matrix $V(\gamma)$. Define the closed, bounded compact parameter subspaces $\Gamma \subseteq \mathbb{R}^{m_1 \times m_2}$, where the dimensions m_1 and m_2 depend on the compound case of logistic ($m_1 = p$, $m_2 = 2$) or exponential ($m_1 = p$, $m_2 = 3$).

Theorem 2 *Assume Assumptions* $1-4$ *hold, and denote by* $F_t(\gamma)$ *the logistic or exponential function with the nuisance vector* $\gamma \in \Gamma$ *defined accordingly. Then,* (i) *under* H_0

$$\sqrt{n}\hat{s}_n(0,\gamma) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(y_t - \hat{\phi}_1' x_t)x_t F_t(\gamma) \Longrightarrow N(0, V(\gamma)), \tag{20}$$

point-wise in $\gamma \in \Gamma$ *where*

$$\begin{aligned} V(\gamma) &= E\left[\epsilon_t^2 \left\{F_t(\gamma) I_k - b(\gamma)A^{-1}\right\} x_t x_t' \left\{F_t(\gamma) I_k - A^{-1} b(\gamma)\right\}\right] \\ b(\gamma) &= E\left[F_t(\gamma) x_t x_t'\right] \\ A &= E\left[x_t x_t'\right], \end{aligned} \tag{21}$$

and I_k *denotes a* k*-dimensional identity matrix. Moreover,* (ii) *under* H_1 *there exists a subset* S *of* $\mathbb{R}^{m_1 \times m_2}$ *with Lebesgue measure zero such that for each* $\gamma \in \Gamma/S$

$$\hat{s}_n(0,\gamma) = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{\phi}_1' x_t)x_t F_t(\gamma) \to \eta(\gamma) \neq 0 \tag{22}$$

with probability one for some vector-functional $\eta(\gamma)$*,* $\eta_i(\gamma) \neq 0$*,* $i = 1...k$*.*

For test purposes, by Assumptions 1-4 the covariance matrix can be consistently estimated for each γ as

$$\begin{aligned} \hat{V}(\gamma) &= \frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t^2 \left[F_t(\gamma) I_k - \hat{b}(\gamma)\hat{A}^{-1}\right] x_t x_t' \left[F_t(\gamma) I_k - \hat{A}^{-1}\hat{b}(\gamma)\right] \\ \hat{\epsilon}_t &= y_t - \hat{\phi}' x_t \\ \hat{b}(\gamma) &= \frac{1}{n}\sum_{t=1}^{n} F_t(\gamma) x_t x_t' \\ \hat{A} &= \frac{1}{n}\sum_{t=1}^{n} x_t x_t'. \end{aligned} \tag{23}$$

Notice that $\hat{V}(\gamma)$ is robust against an unknown form of heteroscedasticity in the innovations series ϵ_t.

There exist cases in which $V(\gamma)$ is singular: in particular, for $\gamma = 0$, $F_t(\gamma)$ is a constant (and the STAR model reduces to a linear AR) thus $\hat{s}(0,\gamma) = 0$ by the least-squares first-order conditions implying $V(\gamma) = 0$, a zero-matrix. It is interesting to point out that in Bierens (1990) the dilemma of a degenerate score variance occurs when $\gamma = 0$ only provided a constant term is included in x_t: in our case, $V(\gamma) = 0$ for *any* x_t when $\gamma = 0$.

For a consistent test statistic with non-degenerate limit distribution, we must therefore analyze the set of all γ for which $V(\gamma)$ is non-positive definite. Consider the following assumption and result. Define the set

$$S^* = \{\gamma \in \Gamma : r'V(\gamma)r \not> 0,\ \ r \in \mathbb{R}^k,\ r \neq 0\}. \tag{24}$$

Assumption 5 For each t, $E[\epsilon_t^2] > 0$, $P(E[\epsilon_t^2|x_t] > 0) = 1$ and $P(|x_t| > 0) = 1$.

For compound transition functions, recall $\gamma = (\gamma_1', ..., \gamma_p')'$ where each γ_j is 2×1 for the logistic (3×1 for the exponential) model. In either case, the first column of γ is identically $(\gamma_{1,0}, ..., \gamma_{p,0})'$, the lag-specific intercepts.

For the following result, initially assume x_t does not contain a constant term such that $\gamma = \gamma_1 = (\gamma_{1,1}, ..., \gamma_{p,1})'$ for the logistic and $\gamma = \gamma_2 = (\gamma_{1,2}, ..., \gamma_{p,2})'$ for the exponential, each p-vectors[4].

Lemma 3 *Under Assumption* 5 *the set S^* has Lebesgue measure zero. In particular, $S^* = \{0\}$.*

Remark 1: If x_t (and therefore v_t) contains a constant term, then $\gamma = 0$ is only one element of S^*: the "intercepts" need not be zero. Consider the LoCoSTAR model, assume $x_{t,1} = 1$ by convention, denote by $\tilde{\gamma}$ the second column of γ, and denote by $\tilde{S}^*$ and $\tilde{\Gamma}$ the relevant sets associated with $\tilde{\gamma}$:

$$\tilde{S}^* = \{\tilde{\gamma} \in \tilde{\Gamma} \subseteq \mathbb{R}^{m_1} : r'V(\gamma)r \not> 0,\ \ r \in \mathbb{R}^k,\ \ r \neq 0\}. \tag{25}$$

Then $\tilde{S}^* = \{0\}$ follows from the line of proof of Lemma 3. In particular, we deduce $S^* = \{\gamma = (\gamma_0, \tilde{\gamma}) = (w, 0)$: $w \in \mathbb{R}^{m_1}, 0 \in \mathbb{R}^{m_1}\}$. Thus, a score test based on $\sqrt{n}\hat{s}(0,\gamma)$ will be non-degenerate for any γ-vectors such that the "slopes" $\tilde{\gamma}$ are non-zero. This can be easily enforced, as detailed in the subsequent sections: test statistics may be derived by selecting $\tilde{\gamma}$ from, e.g., any positive, bounded, subset of $\mathbb{R}^{m_1}$, since all such subsets will result in a non-degenerate test statistic with non-singular asymptotic covariance matrix $V(\gamma)$.

In the ECoSTAR case, denote by $\tilde{\gamma}$ the third column of the matrix γ, hence $S^* = \{\gamma = ([\gamma_0, \gamma_1], \tilde{\gamma}) = (w, 0)$: $w \in \mathbb{R}^{m_1 \times 2}, 0 \in \mathbb{R}^{m_1}\}$. In this case, because we enforce $\gamma_{j,0} = \gamma_{j,1} = 0$ when $\tilde{\gamma}_j \equiv \gamma_{j,2} = 0$, we therefore only need to restrict $\gamma_{j,2} > 0$ for at least one $j = 1...p$ in order to ensure a non-degenerate test. While non-degeneracy requires at least one $\gamma_{j,2} > 0$, recall that consistency, cf. Lemma 1, requires $\gamma_{j,0} \neq 0$, $\gamma_{j,1} \neq 0$ and $\gamma_{j,2} > 0$ for at least one j.

[4] Recall in the ECoSTAR case the intercepts satisfy $\gamma_{j,0} = \gamma_{j,1}/4\gamma_{j,2}^2$, $\gamma_{j,2} > 0$, hence $\gamma_{j,0} = 0$ *if and only if* $\gamma_{j,1} = 0$. Whenever $\gamma_{j,2} = 0$, then $\gamma_{j,0} = \gamma_{j,1} = 0$ by convention.

Lemma 3 differs in important ways from results derived by Bierens (1990: Lemma 2) and de Jong (1996: Lemma 2) for the variance of their associated scalar-valued null scores. Under the auxiliary assumptions that $E[\epsilon_t^2|x_t] > 0$ with probability one, and some measurable function $\mu : \mathbb{R}^k \rightarrow \mathbb{R}^k$ exists such that $(\mu(x_t), x_t)$ has a nonsingular covariance matrix, they prove the set of γ for which scalar variance functionals $V(\gamma) = 0$ is countable and therefore has Lebesgue measure zero. Of course, $\mu(x_t) = x_t$ will work in general. For computational purposes, however, because γ will have to be arbitrarily selected in practice, Bierens (1990) and de Jong (1996) each simply assume $V(\gamma) > 0$. Using the vector weight $x_t F_t(\gamma)$ rather than a scalar weight $F_t(\gamma)$, however, under the minimal Assumption 5 a non-degenerate test statistic is available for any parameter vector $\tilde{\gamma}$ chosen from any bounded compact subset of $\mathbb{R}^{m_1}$ that does not include the zero sub-vector $\tilde{\gamma} = 0$.

Consider general weighted moment conditions of the form

$$\sqrt{n}\hat{s}_n(0,\gamma) = \sqrt{n}\frac{1}{n}\sum\nolimits_{t=1}^{n}(y_t - \hat{\phi}_1' x_t)h(x_t,\delta)F_t(\gamma). \tag{26}$$

Bierens (1990: p. 1449) incorrectly claims the asymptotic covariance matrix $V(\gamma)$ will have to be assumed to be non-singular because the subset Γ will have to be chosen somewhat arbitrarily in practice, and $\gamma \in \Gamma$ could be chosen such that $V(\gamma)$ is singular. Lemma 3, however, proves that for *some* weights, in particular $x_t F_t(\gamma)$, $V(\gamma)$ is non-singular by construction for every $\tilde{\gamma} \neq 0$. In this regard, we may well argue that a moment condition test of functional form with weights implied by smooth transition models dominates standard parametric tests with classic (e.g. exponential or logistic) neural network interpretations. Indeed, LoCoSTAR and ECoSTAR structures are simply generalized versions of artificial neural networks, and as such conventional smooth transition nonlinearity is "totally revealing" in the sense of Kuan and White (1994) and Stinchcombe and White (1998). Moreover, a nondegenerate, consistent pointwise LM test is available where smooth transition nonlinearity, rather than feedforward neurons (exponential or logistic), are used to improve model fit.

Consequently, by Lemmas 1 and 3, and Theorem 2, and by standard asymptotic theory, the test statistic

$$T_n(\gamma) = n\hat{s}_n(0,\gamma)'\hat{V}(\gamma)^{-1}\hat{s}_n(0,\gamma) \tag{27}$$

converges in law under H_0 to a random variable which is $\chi^2(k)$ distributed point-wise in γ, except for $\tilde{\gamma} = 0$.

Next, define the sup-statistic,

$$g_n = \sup\nolimits_{\gamma \in \Gamma} T_n(\gamma). \tag{28}$$

The subsequent corollary follows immediately. Recall $\tilde{\gamma}$ denotes the last ("far right") column of γ.

Corollary 4 *Under Assumptions* $1 - 5$ *and under* H_0*,*

$$T_n(\gamma) \Longrightarrow \chi^2(k), \tag{29}$$

pointwise in γ except for $\tilde{\gamma} = 0$. Moreover, under H_1, there exist a subset S of $\mathbb{R}^{m_1 \times m_2}$ with Lebesgue measure zero such that for every $\gamma \in \Gamma / S$

$$\frac{T_n(\gamma)}{n} \to \tilde{\eta}(\gamma) \ a.s. \tag{30}$$

for some real-scalar $\tilde{\eta}(\gamma) > 0$. In particular, for arbitrarily large $N > 0$, and for every $\gamma \in \Gamma / S$,

$$P\left(\lim_{n \to \infty} T_n(\gamma) > N\right) = 1. \tag{31}$$

Remark 1: It follows immediately that the sup-statistic g_n has the consistency property $P(\lim_{n\to\infty} g_n > N) = 1$ for any deviation from the null. Under H_0, the limit distribution of g_n depends, in general, on Γ and S, and therefore on the distribution of $\{y_t, x_t\}$. Hence, p-values will have to be derived by simulation and/or bootstrap.

4. Simulation Study We now investigate the empirical size and power properties of the sup-statistic g_n and an associated randomized statistic $T_n(\gamma)$ under a null of linearity, and under LSTAR, ESTAR and bilinear alternatives.

Our simulations are based on the following models:

$$\begin{aligned}
&H_0 : y_t = \phi_1' x_t + \epsilon_t \\
&H_1^L : y_t = \phi_1' x_t + \phi_2' x_t \left(1 + \exp(-\gamma x_{t,2})\right)^{-1} + \epsilon_t \\
&H_1^E : y_t = \phi_1' x_t + \phi_2' x_t \exp(-\gamma x_{t,2}^2) + \epsilon_t \\
&H_1^{BL_1} : y_t = \phi_1' x_t + \phi_2 y_{t-1}\epsilon_{t-1} + \epsilon_t, \ \ |\phi_2| < 1 \\
&H_1^{BL_2} : y_t = \phi_1' x_t + y_{t-1}\epsilon_{t-1} + \epsilon_t \\
&H_1^{BL_3} : y_t = \phi_1 y_{t-1}\epsilon_{t-1} + \epsilon_t, \ \ |\phi_1| < 1
\end{aligned}$$

where ϵ_t are *iid* standard normal, and $x_t = (1, y_{t-1}, ..., y_{t-p})'$ for some $p \geq 1$. Notice that the transition parameters γ are scalar-valued, only $x_{t,2} = y_{t-1}$ is employed in the transition functions and $c = 0$ for all simulations. Under H_0 the true data generating process is linear; under H_1^L and H_1^E the true process is a 2-regime LSTAR and ESTAR, respectively; and under each $H_1^{BL_i}$, the process is bilinear.

4.1 Set-up We consider sample sizes $n = 100$, 500, and 1000: in each case, we generate $3n$ observations, and retain the last n in order to reduce dependence on starting values. For each simulated series, the order p is randomly chosen from the set $\{1, ..., 10\}$, and the vectors ϕ_i, $i = 1, 2$, are randomly chosen from the hypercube $[-.95, .95]^{p+1}$. For all simulations we fix $\gamma = 3$. Because we require the null model to be covariance stationary, only vectors ϕ_1 with characteristic polynomial roots outside the unit circle are considered.

We generate 1000 replications of each series above. For each series a linear model is estimated and the resulting residuals are tested. In order to specify the null model, we employ both a minimum AIC model selection criterion for the order p, as well as the true order for benchmark comparisons.

4.2 CoSTAR Tests In order to test for linearity, consider a general CoSTAR model

$$\begin{aligned} y_t &= \phi_1' x_t + \phi_2' x_t F(v_t, \gamma) + u_t \\ F(v_t, \gamma_i) &= \prod_{i=1}^{p} F(v_{t,i}, \gamma_i), \end{aligned} \tag{32}$$

Using a traditional STAR parameterization, for logistic processes define $c_i = -\gamma_{i,0}/\gamma_{i,1}$ and $\tilde{\gamma} \equiv \gamma_{i,1} > 0$; and for exponential processes $c_i = \gamma_{i,1}/2\gamma_{i,2}$ and $\tilde{\gamma} \equiv \gamma_{i,2} > 0$:

$$\begin{aligned} y_t &= \phi_1' x_t + \phi_2' x_t F(v_t, \tilde{\gamma}, c) + u_t \\ F(v_t, \tilde{\gamma}, c) &= \prod_{i=1}^{p} F(v_{t,i}, \tilde{\gamma}_i, c_{,i}). \end{aligned} \tag{33}$$

Define the sets $\Gamma = \{.10, .11, ..., 10\}$ and $C_i = \{y^{(i)}_{[.15n]}, ..., y^{(i)}_{[.85n]}\}/0$ [5], where $y^{(i)}_{[j]}$ denotes the j^{th} order statistic of the i^{th} lagged series y_{t-i}. We maximize the test statistic $T_n(\tilde{\gamma}, c)$ over $\tilde{\gamma}_i \in \Gamma$ and $c_i \in C_i$, $i = 1...p$ [6].

Clearly, computation time will be burdensome for large p and n. We reduce the complexity of the maximization problem by searching over a subset of $\tilde{\gamma}_i \in \Gamma$. In particular, we maximize $T_n(\tilde{\gamma}, c)$ under two separate sets of restrictions: (i) for $\tilde{\gamma}_i \in \Gamma$ and $\tilde{\gamma}_j = 0$, $i = 1...p$, $j \neq i$; and (ii) for $\tilde{\gamma}_1 = \tilde{\gamma}_2 = ... = \tilde{\gamma}_i \in \Gamma$ and $\tilde{\gamma}_j = 0$, $i = 1...p$, $j > i$. In the former case, we compute a simple, non-compound STAR test statistic based on using only one lag y_{t-i} at a time as the transition variable: this is precisely how the standard polynomial regression method is performed. In the latter case, we compute CoSTAR test statistics by incrementally adding threshold information. Denoting by $\tilde{\gamma}^*$ and c^* the nuisance vectors that maximize $T_n(\tilde{\gamma}, c)$ over the grid search in either case, the sup-statistics satisfy $g_n = T_n(\tilde{\gamma}^*, c^*)$.

Once the statistic g_n is generated, we employ Hansen's (1996) parametric bootstrap method for approximating the asymptotic p-value. For Hansen's method we simulate J *iid* standard normal random n-vectors $(u_{t,j})_{t=1}^{n}$, $j = 1...J$, generate J scores $\hat{s}_{n,j}(0, \gamma) = n^{-1} \sum_{t=1}^{n} \hat{\epsilon}_t u_{t,j} x_t F_t(\gamma)$, J test statistics, $T_{n,j}(\tilde{\gamma}, c)$, and J statistic functionals $g_{n,j} = T_{n,j}(\tilde{\gamma}^*_j, c^*_j)$. The p-value is the percent frequency of the event $g_{n,j} > g_n$[7]. Under our Assumptions 1-4, Hansen's (1996) Assumptions 1-3 and Theorems 1-2 hold[8], and therefore the approximate

[5] Thus, the set C_i contains a middle range of lag values y_{t-i}, except for the value 0.

[6] The restriction that c_i be fixed to order statistics between the 15^{th} lower and upper quantiles was suggested by Luukonnen *et al* (1988) and Teräsvirta (1994), a standard used throughout the STAR literature.

[7] The bootstrap p-value $\hat{p}$ satisfies

$$\hat{p} = \frac{1}{J} \sum_{j=1}^{J} I\left(g_{n,j} > g_n\right)$$

where $I(A) = 1$ if the event A is true, and 0 otherwise.

[8] Specifically, Hansen's (1996) Assumption 1 defines the process $\{y_t, x_t\}$ as strictly stationary and absolutely regular, which holds by our Assumption 1. Hansen's Assumption 2 bounds the nonlinear component, which holds for STAR models given the boundedness condition, $0 \leq F_t(\gamma) \leq 1$. Finally, the author's Assumption 3 details asymptotic bounds of various sample moments essentially identical to our Assumption 4.

p-value converges in probability to the true p-value. For all simulations, we set $J = 500$. These are the $STAR$ and $CoSTAR$ tests.

Randomized STAR Tests

As an experimental means to improve small sample power, we also perform a point-wise randomized test of linearity against a STAR alternative with hybrid polynomial-smooth transition terms. In particular, we include one "neuron" $\phi_2 F_{t,2}$, where ϕ_2 is a scalar, one STAR term $\phi_3' x_t F_{t3}$, and one randomized polynomial term $\phi_4' x_t^a F_{t,4}$:

$$y_t = \phi_1' x_t + \phi_2 F_{t,2} + \phi_3' x_t F_{t,3} + \phi_4' x_t^a F_{t,4} + u_t, \tag{34}$$

where the k-vector components a_i are randomly selected from the integer set $\{0, 1, 2, 3\}$. Each $F_{t,i}$ denotes a non-compound transition function $F_t(y_{t-d_i}, \tilde{\gamma}_i, c_i)$. The nuisance parameters c_i, d_i and $\tilde{\gamma}_i$ are randomly selected from the respective sets $\{y_{[.15n]}, y_{[.85n]}\}$, $\{1, 2, 3\}$ and $\Gamma = [.1, 10]$. The test is a standard LM test of the zero restrictions $\phi_2 = \phi_3 = \phi_4 = 0$. This is the $PSTAR$ test.

Polynomial Regression Tests, Neural Tests, etc.

For comparisons, we also perform the standard Bierens test both by implementing Bierens (1990) criterion (*BIER*) and by using Hansen's (1996) method for evaluating the asymptotic distribution of the Bierens sup-statistic (*BIER_ han*). In this manner, we control for the possibility that differences between the STAR sup-test and the Bierens test is merely due to the use of Hansen's (1996) method, rather than due to use of the vector weight $x_t F_t$. The model implied by the Bierens test is a neural network model with one feedforward layer,

$$y_t = \phi_1' x_t + \phi_2 F(x_t, \gamma) + u_t, \tag{35}$$

where γ denotes a k-vector, and $F(x_t, \gamma)$ denotes the logistic $(1 + \exp(\gamma' x_t))^{-1}$ or exponential $\exp(\gamma' x_t)$. The test is an LM test of the hypothesis $\phi_2 = 0$, and test statistics are maximized over $\gamma \in \Gamma^k$.

Similarly, we perform the neural test of neglected nonlinearity, cf. Lee *et al* (1996), which is equivalent to a randomized Bierens test over the nuisance parameter space. For each test we include two scalar neurons, hence the implied ANN model is

$$y_t = \phi_1' x_t + \phi_2 F(x_t, \gamma_1) + \phi_3 F(x_t, \gamma_2) + u_t, \tag{36}$$

where each γ_i denotes a k-vector, and $F(x_t, \gamma_i)$ denotes the logistic $(1 + \exp(\gamma_i' x_t))^{-1}$ or exponential $\exp(\gamma_i' x_t)$. The vectors γ_i, $i = 1, 2$, are randomly selected from the set Γ^k.

We also employ the polynomial regression method of Luukonen *et al* (1988) and Teräsvirta (1994), the RESET test, and the McLeod-Li test. For the STAR polynomial test, we estimate models of the form

$$y_t = \beta_0' x_t + \sum\nolimits_{i=1}^{L} \beta_i' \tilde{x}_t y_{t-d}^i + u_t, \tag{37}$$

where $\tilde{x}_t = (y_{t-1}, ..., y_{t-p})'$. Under a null of linearity against an LSTAR alternative, $L = 3$ and (37) implies $\beta_i = 0$, $i = 1..3$. Under a null of linearity against an ESTAR alternative, $L = 4$ and (37) implies $\beta_i = 0$, $i = 1..4$. In order to decide between LSTAR and ESTAR alternatives based on the polynomial regression, Teräsvirta (1994) suggests a test of $\beta_i = 0$, $i = 1..4$ first in order to substantiate concern for STAR nonlinearity at all, then a sequence of F-tests on parameter sub-sets from (37). Because we are interested in whether the test procedure can find *any* deviation from the null of linearity, we do not pursue the test sequence approach and simply report null rejection frequencies based on tests of (37) with $L = 3$ or 4. Rejection in either case is consistent with evidence in favor of polynomial nonlinearity and STAR nonlinearity. The test is performed for $d \in \{1, ..., p\}$, where p is either assumed known or selected by minimizing the AIC of linear models, and the statistic with the smallest p-value is selected. These are the $POLY$ tests.

For the McLeod-Li test, we perform a standard portmanteau test on the squared null residuals for lags 1...5. For the RESET test, we follow the procedure detailed in Thursby and Schmidt (1977) by estimating the auxiliary regression based on the null residuals $\hat{u}_t$,

$$\hat{u}_t = \beta_0' x_t + \sum_{i=2}^{L} \sum_{j=2}^{k} \beta_{i,j} x_{t,j}^i + w_t, \tag{38}$$

where we set $L = 3$. A standard LM test for the linearity hypothesis $H_0 : \beta_{i,j} = 0$ is performed.

For all LM tests employed in this study, covariance estimators robust to unknown forms of conditional heteroscedasticity are used (e.g. (23) for STAR tests).

4.3 Results Results for H_0 are contained in Tables 1-2, and Tables 3-4 contain empirical powers for the various alternatives. See Appendix 2.

Linear AR

For linear processes, the STAR, CoSTAR and PSTAR tests compare well with the popularly used neural and polynomial regression tests. The polynomial test tends to under-reject the null for any process (i.e. low empirical size and power). The STAR and CoSTAR sup-tests over rejects the null for small n and substantially under-rejects the null for large n. This suggests the bootstrap p-value method may not capture the exact distributional dynamics of the proposed test statistics, although the complex nature of distortion favors the test's performance, in particular when the test is studied under the various alternatives, below.

We control for the possibility that it is merely Hansen's (1996) method for evaluating the asymptotic p-value that differentiates the simple STAR test from the Bierens test. In general, however, even when Hansen's p-value is used with the Bieren's sup-statistic, the null hypothesis is still over-rejected for all sample sizes when the exponential is used. Thus, it appears the vector-weights themselves augment the test statistic's performance, and not merely the method of analyzing the p-value.

LSTAR

Under the alternative of LSTAR, the STAR and CoSTAR sup-tests dominate all tests, particularly for small samples $n \leq 500$. The randomized PSTAR test performs well, particularly for large n, however in the bench-mark tests where the true AR-order is known the neural test generates larger empirical powers based on either logistic or exponential tests. In the realistic case where the AR-order is selected by minimizing the AIC, however, the PSTAR tests out-perform the neural tests with noticeable improvements in the mid-sample size range. In all cases, the polynomial regression test based on a Taylor expansion of the STAR model performs reasonably well, but never rejects more than 60% of the false null hypotheses, and is dominated (often substantially) by the neural, STAR and PSTAR tests.

ESTAR

Under the alternative of ESTAR, the STAR, CoSTAR, PSTAR and neural tests fared equally well. The STAR, CoSTAR and PSTAR tests dominate the polynomial regression tests.

The randomized PSTAR test performed well for large n, generating a rejection frequency in par with the two STAR sup-tests: this suggests that the randomized, non-compound polynomial term $x_t^a F_t(\gamma)$ included in the PSTAR test can absorb the nonlinear structure contained in the null residuals as well as the compound or non-compound terms $x_t F_t(\gamma)$ with an optimally selected nuisance parameter γ. While interesting in its own right, in practical terms this implies that the computational burden of generating a sup-statistic may be reasonably by-passed in favor of a randomized test.

Comparatively, however, the neural test always performs better than the three STAR tests in the bench-mark tests of known AR order, although the margin of improvement diminishes to about 2% for the STAR (under 1% for the CoSTAR) for large samples. When the AR order is selected by minimizing the AIC, however, the STAR, CoSTAR and PSTAR tests out-perform the neural tests for large n, and the PSTAR test can detect nonlinearity more often than the STAR and CoSTAR sup-tests for $n = 500$ or 1000.

The Bierens tests performed well, in particular when the exponential is used. However, given the size distortions encountered above, evidently these tests are dominated by the extremely low test sizes and the ample empirical power of the STAR, CoSTAR and PSTAR tests. Indeed, the Logistic STAR and CoSTAR sup-tests substantially dominate the Bierens test for small samples.

Bilinear

For all remaining hypotheses, we focus only on tests of residuals from minimum AIC models. The pecking order essentially continues for bilinear processes. The polynomial regression performs well for large n in 2 out of 3 bilinear models, although the randomized PSTAR test is better. For for $n = 500$ or 1000, the PSTAR test out-performs the STAR and COSTAR sup-tests, and out-performs the neural test for large n. The improvement in rejection accuracy is particularly noticeable for the standard bilinear process under $H_1^{BL_3}$. In this case, the randomized PSTAR test provides a large improvement over the STAR, CoSTAR, neural and polynomial tests, detecting nonlinearity in up to 11% more simulated series than the other tests.

Overall, for small samples the neural test seems to provide the best probability of detecting nonlinearity, however for medium-to-large samples the pointwise PSTAR and logistic Bierens tests typically dominate all other tests (recall, the exponential Bierens test over-rejects the null). Indeed, under $H_1^{BL_2}$ the PSTAR test dominates for large n, with rejection frequencies near 60% for $n = 1000$, while the polynomial test is only able to detect non-linearity in under 2% of all simulated series, a dismal performance.

Finally, the McLeod-Li test, created essentially for bilinear nonlinearity detection, works particularly well for each bilinear model simulated here, howerver is sub-optimal for STAR processes. Moreover, it is interesting to point out that the RESET test performed better than the polynomial regression test for LSTAR processes, and for the stationary bilineary processes.

In summary, the STAR and CoSTAR sup-tests and PSTAR randomized test dominate the polynomial regression test for all STAR and bilinear processes considered, and in general dominates the Bierens test under the null and under STAR and non-stationary bilinear alternatives ($H_1^{BL_2}$). The logistic Bierens test works particularly well for conventional bilinear processes, $H_1^{BL_1}$ and $H_1^{BL_3}$. Hansen's (1996) bootstrap technique does not appear to be the fundamental reason why the simple STAR test performs so well relative to the Bierens test: evidence suggests the implied CoSTAR or STAR moment condition weights help to smooth out null hypothesis rejection frequencies, and provide a significant power lift under the studied alternatives. Moreover, application of a CoSTAR test with compound transition function does not significantly improve the performance of the class of STAR tests developed here. Based on simulated simple STAR and bilinear processes, evidence suggests we can safely implement a randomized STAR test with simple transition functions, or a simple STAR sup-test, without affecting test performance.

5. Empirical Applications In this penultimate section, we briefly exemplify the information content of the STAR, CoSTAR and PSTAR tests when applied to macroeconomic processes considered in Stock and Watson (1989), Friedman and Kuttner (1993) and Rothman *et al* (2001). We consider 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 Jan. 1959 - Aug. 2003[9], and seasonally adjusted at the source when applicable.

All variables, except for the rate spread, are differenced in order to control

[9] The sample size is 536 months, before lag and differencing adjustments.

for the likely presence of one positive unit root in each series, as evidenced by standard augmented Dickey-Fuller tests. Evidence suggests the Treasury bill and commercial paper rates are cointegrated of order one such that the spread is $I(0)$. We also consider annual growth rates of all series except the rate spread, $y_t - y_{t-12}$, in order to control for noisiness frequently encountered in monthly macro-time series. In any case, annual growth rates are a fundamental measure of long-run economic growth and stability, and therefore demand investigation for properties of regime nonlinearities.

Results are contained in Tables 5 and 6. For monthly growth series, the STAR and CoSTAR sup-tests and randomized PSTAR test suggest highly significant (below the 1%-level) evidence exists for nonlinearity in money growth, inflation, and fluctuations in the commercial paper rate and the rate spread. The strongest evidence points to LSTAR nonlinearity. The STAR, CoSTAR and PSTAR test were demonstrated to be particularly useful in detecting LSTAR nonlinearity, specifically for the STAR and CoSTAR sup-tests for a sample size of 500 (roughly the size we have in this study), thus the empirical evidence here seems noteworthy. Based on Lemma 1, even if the above macro-processes are not driven by a STAR data generating mechanism an LSTAR model appears to provide a better approximation to the true structure than a linear AR model.

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. The exponential tests suggests nonlinearity in money, and the logistic test suggests nonlinearity in the price series. Moreover, at the 5% level the neural test only finds nonlinearity in the commercial paper rate (the logistic test) and in the rate spread (the exponential test). The RESET test fails to detect nonlinearity in any series.

For the annual growth series, however, the tests portray a somewhat different nature of linearity. The exponential and logistic STAR tests strongly suggest output and the commercial paper rate are nonlinear, respectively, and specifically fail to reject tests of linearity in money and inflation. The neural tests sharply split for the rate spread, favoring the exponential. The PSTAR test never detects STAR nonlinearity: because both tests are consistent against any deviation from the null, it difficult to explain the divergence between the STAR and PSTAR tests. The standard polynomial tests never reject the null of nonlinearity at any conventional level of significance, except for the exponential test on the Treasury bill rate.

6. Concluding Remarks In order to improve on the test currently espoused in the STAR literature, we develop a new STAR model that accounts for multiple transition variables. In particular, we extend the transition function to include vector processes and multivariate, multiplicative transition functions, and we experiment with hybrid polynomial-smooth transition weights. Augmenting consistent moment condition test weights to optimize power against STAR alternatives has the advantage of generating a test statistic which is never degenerate under the alternative, a property that the Bierens test does not have.

Each STAR and randomized PSTAR statistic performs well under the null of linearity, and under various STAR and non-STAR alternatives. Our simulations suggest that it is not merely Hansen's (1996) p-value method that enhances test performance. Of particular note, Hansen's (1996) method neither uniformly nor sufficiently improves the performance of Bierens' (1990) original test sup-statistic in either the exponential or logistic case under the null, and a randomized non-compound PSTAR tests with polynomial terms performs particularly well under all hypotheses, and by construction is much easier to derive.

In our small sample study, it is particularly revealing that the much discussed Taylor expansion method and subsequent polynomial regression test is demonstrably sub-optimal relative to the tests developed here. We allow for the linear component of each model to be randomly constructed, including AR orders and coefficient magnitudes. Previous simulations, by comparison, are rather limited by virtue of fixing the null and alternative structures. In our "double-blind" environment in which we know neither the AR order nor the coefficient magnitudes in advance, the polynomial tests perform well against STAR alternatives, but cannot detect true STAR nonlinearity as frequently as the tests developed here, where typically the new tests provide a substantial margin of improvement. Moreover, and not surprisingly, the polynomial tests proved to be particularly bad at detecting some forms of bilinearity, and the RESET test is more adept at detecting logistic smooth transition nonlinearity than the traditional test method.

Appendix 1: Assumptions

Assumption 1 The data-generating process $\{y_t, x_t\}$ exists on $L_2(\Omega, P, \mathfrak{F}_t)$ where $\mathfrak{F}_t$ denotes a strictly increasing σ-algebra induced by (y_{t-i}, x_{t-i}), $i = 0, 1, ...,$ such that $\mathfrak{F}_{t-1} \subset \mathfrak{F}_t$. The process $\{y_t, x_t\}$ is strictly stationary, ergodic, governed by non-degenerate joint distribution function with non-degenerate marginal distributions, and for some $r > 1$, $E|y_t|^{2r} < \infty$. The regressors x_t are k-vectors, measurable with respect to $\mathfrak{F}_{t-1}$. The innovations ϵ_t form a $(0, \sigma^2)$-martingale difference sequence with respect to $\mathfrak{F}_{t-1}$ under H_0. Under H_1, ϵ_t is a $(0, \sigma^2)$-white noise process for each t. The function $h(x_t, \delta)$ is any bounded, continuous mapping from $\mathbb{R}^k \times \Delta$ to $\mathbb{R}^k$, measurable with respect to $\mathfrak{F}_{t-1}$, such that $P(\inf_{\delta \in \Delta} |h(x_t, \delta)| > 0) = 1$, and $\sup_{\delta \in \Delta} |h(x_t, \delta| < \infty$ with probability one.

Assumption 2 The conditioning vector $x_t = (1, y_{t-1}, ..., y_{t-p})'$ is bounded in probability element-wise: for each $i = 1...k$, there exists some $0 < M < \infty$ such that $P(|x_{t,i}| < M) = 1$.

Assumption 3 Let Φ denote a compact, convex subset of $\mathbb{R}^k$. There exists a unique element $\phi_0 = \arg\inf_{\phi \in \Phi} E(y_t - \phi' x_t)^2$ where ϕ_0 is in the interior of Φ.

Assumption 4 The following uniform moment bounds hold for each t:

$$\begin{aligned} \sup_{\gamma \in \Gamma} E\left|\epsilon_t^2 F_t(\gamma)^2 x_t x_t'\right| &< \infty; \quad \sup_{\gamma \in \Gamma} E|\epsilon_t^2 F_t(\gamma) x_t x_1'| < \infty \\ \sup_{\gamma \in \Gamma} E|F_t(\gamma) x_t x_1'| &< \infty; \quad E|x_t x_t'| < \infty \end{aligned}$$

where the bounds are understood to be element-wise. Define the probability limits

$$\hat{b}(\gamma) = \frac{1}{n} \sum\nolimits_{t=1}^{n} F_t(\gamma) x_t x_t' \rightarrow b(\gamma); \quad \hat{A} = \frac{1}{n} \sum\nolimits_{t=1}^{n} x_t x_t' \rightarrow A.$$

where $\hat{A}$, for all $n > 1$, and A are nonsingular. Observe that each matrix is symmetric, $k \times k$.

Assumption 1 is standard and essentially restricts dependence, and defines the skeleton $\phi_1' x_{1t}$ as the best $\mathfrak{L}_2$-predictor by the martingale difference property under H_0. Assumption 2 allows for a non-degenerate test statistic by bounding the transition function through x_t. Assumptions 3 and 4 guarantee uniqueness of the underlying parameters, and the existence of the asymptotic covariance matrix for an LM statistic.

Appendix 2: Tables

Table 1

Null, True Order, 5%-level[a]			
n	100	500	1000
CoSTAR_L[c]	.0780[b]	.0240	.0100
CoSTAR_E	.0790	.0240	.0110
STAR_L	.0770	.0120	.0090
STAR_E	.0770	.0160	.0080
PSTAR_L	.0240	.0350	.0410
PSTAR_E	.0290	.0360	.0420
NEURAL_L	.0440	.0380	.0560
NEURAL_E	.0500	.0370	.0430
BIER_han_L	.0280	.0220	.0320
BIER_han_E	.0690	.0580	.0590
BIER_L	.0670	.0610	.0630
BIER_E	.1180	.1090	.1100
POLY_L	.0040	.0000	.0010
POLY_E	.0040	.0000	.0010
RESET	.0380	.0480	.0410
ML-1[d]	.0450	.0540	.0500
ML-2	.0390	.0490	.0540
ML-3	.0510	.0530	.0480

Notes: a. All tests in this study are performed at the 5%-level;
b. Values denote rejection frequencies at the 5%-level;
c. "L" denotes a test against an LSTAR alternative;
"E" denotes a test against an ESTAR alternative;
d. ML-h denotes the ML-test with h-lags.

Table 2

Null, $p^* = \arg\min(AIC^a)$			
n	100	500	1000
CoSTAR_L	.0680	.0260	.0150
CoSTAR_E	.0750	.0320	.0220
STAR_L	.0650	.0170	.0100
STAR_E	.0680	.0150	.0090
PSTAR_L	.0250	.0340	.0390
PSTAR_E	.0290	.0460	.0460
NEURAL_L	.0320	.0500	.0420
NEURAL_E	.0370	.0630	.0350
BIER_han_L	.0190	.0340	.0290
BIER_han_E	.0730	.0620	.0660
BIER_L	.0620	.0720	.0600
BIER_E	.1090	.1190	.1200
POLY_L	.0010	.0030	.0020
POLY_E	.0010	.0030	.0020
RESET	.0450	.0380	.0490
ML-1	.0520	.0700	.0860
ML-2	.0570	.0880	.0910
ML-3	.0640	.1050	.0950
p-differential^b	.3890	.2230	.1720

Notes: a. AR-orders p^* are selected my minimizing the AIC;
b. Sample average of order differential p - p^*.

Table 3.1

H_1, *True Order*, $n = 100$					
	H_1^L	H_1^E	$H_1^{BL_1}$	$H_1^{BL_2}$	$H_1^{BL_3}$
CoSTAR_L	.5120	.2020	.1030	.1380	.1350
CoSTAR_E	.5240	.2010	.1510	.1390	.1420
STAR_L	.5070	.1110	.1030	.1320	.1290
STAR_E	.5200	.1450	.1400	.1360	.1390
PSTAR_L	.2320	.0620	.0560	.0560	.0520
PSTAR_E	.2040	.0580	.0650	.0600	.0590
NEURAL_L	.3880	.2270	.2230	.2200	.1980
NEURAL_E	.3610	.2100	.2360	.2250	.2270
BIER_han_L	.3890	.0820	.1070	.0240	.1220
BIER_han_E	.4910	.3540	.1610	.2010	.1730
BIER_L	.4480	.2380	.1900	.1080	.2260
BIER_E	.5000	.4170	.2380	.2040	.2510
POLY_L	.1030	.0040	.0060	.0190	.0010
POLY_E	.1030	.0040	.0060	.0190	.0010
RESET	.2850	.0460	.0560	.0110	.1360
ML-1	.1490	.0350	.1640	.7750	.1180
ML-2	.1700	.0520	.1550	.8140	.1100
ML-3	.1730	.0700	.1460	.8250	.1110

Table 3.2

H_1, *True Order*, $n = 500$					
	H_1^L	H_1^E	$H_1^{BL_1}$	$H_1^{BL_2}$	$H_1^{BL_3}$
CoSTAR_L	.7750	.3710	.3280	.3690	.3810
CoSTAR_E	.7780	.4510	.4160	.4700	.4390
STAR_L	.7740	.3590	.3280	.3450	.3770
STAR_E	.7750	.4420	.4150	.4200	.4350
PSTAR_L	.6750	.4670	.4360	.4300	.4420
PSTAR_E	.6370	.4870	.4400	.4380	.4540
NEURAL_L	.7420	.5190	.4910	.4680	.5010
NEURAL_E	.7060	.5300	.4760	.4600	.5000
BIER_han_L	.6840	.5660	.4900	.0250	.6040
BIER_han_E	.7640	.6070	.5900	.2060	.6310
BIER_L	.6630	.4340	.5970	.0690	.6770
BIER_E	.7140	.6460	.6290	.1960	.6860
POLY_L	.5050	.2160	.2260	.0200	.1960
POLY_E	.5050	.2160	.2260	.0200	.1960
RESET	.6610	.2230	.4930	.0080	.6290
ML-1	.3550	.0840	.4710	.9680	.3490
ML-2	.3960	.1140	.4420	.9830	.3250
ML-3	.4120	.1460	.4220	.9930	.2850

Table 3.3

H_1, *True Order*, $n = 1000$					
	H_1^L	H_1^E	$H_1^{BL_1}$	$H_1^{BL_2}$	$H_1^{BL_3}$
CoSTAR_L	.8600	.5280	.5170	.5520	.5300
CoSTAR_E	.8620	.5810	.5790	.5860	.6100
STAR_L	.8510	.5270	.5130	.5210	.5270
STAR_E	.8610	.5750	.5740	.5790	.6000
PSTAR_L	.7570	.5750	.5980	.5990	.6010
PSTAR_E	.7310	.5750	.5880	.5990	.6110
NEURAL_L	.8030	.6088	.5940	.6040	.6050
NEURAL_E	.7540	.5962	.5940	.5900	.5960
BIER_han_L	.7764	.3980	.6540	.0400	.7390
BIER_han_E	.8310	.7140	.7050	.2530	.7400
BIER_L	.7100	.5420	.7190	.0500	.7760
BIER_E	.7590	.7590	.7470	.2380	.7790
POLY_L	.6750	.4600	.4890	.0220	.4090
POLY_E	.6750	.4600	.4890	.0220	.4090
RESET	.7410	.3400	.6680	.0060	.7670
ML-1	.4290	.1050	.5830	.9910	.5070
ML-2	.4880	.1420	.5690	.9970	.4700
ML-3	.5090	.1910	.5590	.9970	.4450

Table 4.1

H_1, $p^* = \arg\min(AIC)$, $n = 100$					
	H_1^L	H_1^E	$H_1^{BL_1}$	$H_1^{BL_2}$	$H_1^{BL_3}$
CoSTAR_L	.5690	.1230	.0950	.1480	.1010
CoSTAR_E	.5630	.1620	.1250	.1820	.1500
STAR_L	.5670	.1020	.0950	.1170	.0980
STAR_E	.5600	.1480	.1240	.1560	.1390
PSTAR_L	.3700	.0710	.0470	.0640	.0660
PSTAR_E	.3370	.0710	.0360	.0580	.0640
NEURAL_L	.3860	.2150	.1980	.2140	.0960
NEURAL_E	.3440	.2050	.1940	.2150	.0910
BIER_han_L	.3820	.0710	.0850	.0280	.1190
BIER_han_E	.4740	.3340	.1350	.2020	.1970
BIER_L	.4060	.1980	.1840	.0990	.2360
BIER_E	.4580	.3670	.2060	.2170	.2710
POLY_L	.0910	.0070	.0190	.0190	.0010
POLY_E	.0910	.0070	.0190	.0190	.0010
RESET	.3850	.0410	.1220	.0130	.1330
ML-1	.1270	.0490	.5230	.7730	.1140
ML-2	.1490	.0710	.5390	.8210	.1040
ML-3	.1780	.0850	.5530	.8430	.0940
p-differential	.3120	.3340	.3600	.3530	-.1430

Table 4.2

$H_1,\ p^* = \arg\min(AIC),\ n = 500$					
	H_1^L	H_1^E	$H_1^{BL_1}$	$H_1^{BL_2}$	$H_1^{BL_3}$
CoSTAR_L	.7730	.3210	.3400	.3420	.3100
CoSTAR_E	.8100	.3760	.4140	.4230	.4010
STAR_L	.7700	.3150	.3400	.3110	.3030
STAR_E	.8020	.3660	.4100	.3840	.3970
PSTAR_L	.7480	.4450	.4410	.4290	.4320
PSTAR_E	.7110	.4550	.4590	.4420	.4400
NEURAL_L	.7260	.5090	.4940	.4760	.3630
NEURAL_E	.6770	.4920	.5070	.4860	.3750
BIER_han_L	.6990	.2530	.4440	.0320	.6080
BIER_han_E	.7570	.5950	.5220	.2230	.6400
BIER_L	.6460	.4370	.5550	.0630	.6770
BIER_E	.7040	.6540	.5870	.2150	.6870
POLY_L	.4350	.2290	.2290	.0170	.2070
POLY_E	.4350	.2290	.2290	.0170	.2070
RESET	.6320	.2030	.4920	.0040	.6360
ML-1	.3250	.0890	.4620	.9720	.3460
ML-2	.4010	.1440	.4480	.9840	.3230
ML-3	.4310	.1690	.4520	.9930	.3080
p-differential	.0030	.0770	.1590	-.0140	-.1350

Table 4.3

$H_1,\ p^* = \arg\min(AIC),\ n = 1000$					
	H_1^L	H_1^E	$H_1^{BL_1}$	$H_1^{BL_2}$	$H_1^{BL_3}$
CoSTAR_L	.7930	.4720	.4900	.5480	.4810
CoSTAR_E	.8100	.5390	.5690	.5710	.5690
STAR_L	.7910	.4660	.4890	.5050	.4800
STAR_E	.8070	.5300	.5680	.5650	.5670
PSTAR_L	.7800	.5840	.5960	.5900	.5840
PSTAR_E	.7440	.6220	.6080	.5990	.5940
NEURAL_L	.7680	.5720	.5760	.5840	.4960
NEURAL_E	.7090	.5700	.5660	.5660	.4780
BIER_han_L	.7430	.4040	.6250	.0420	.7350
BIER_han_E	.7970	.7170	.6800	.2160	.7410
BIER_L	.7160	.5560	.7120	.0690	.7800
BIER_E	.7700	.7350	.7260	.2200	.7830
POLY_L	.5840	.4860	.5240	.0140	.4000
POLY_E	.5840	.4860	.5240	.0140	.4000
RESET	.6990	.3680	.6660	.0020	.7600
ML-1	.3820	.1190	.5850	.9910	.4640
ML-2	.4490	.1850	.5740	.9980	.4340
ML-3	.5160	.2280	.5670	.9980	.4240
p-differential	-.0280	-.0030	.0950	-.2000	-.1220

Table 5

Monthly Growth Series						
	Δm	Δy	Δp	Δr_b	Δr_p	$r_b - r_p$
CoSTAR_L	.0000[a,b]	1.000[c]	.0000	.4200	.0150	.0000
CoSTAR_E	.0000	1.000	.0000	.8700	.0080	.0000
STAR_L	.0000	1.000	.0000	.6700	.0200	.0000
STAR_E	.0000	1.000	.0000	.9500	.0100	.0000
PSTAR_L	.0082	.4481	.0784	.3745	.0139	.0295
PSTAR_E	.0676	.3890	.4117	.5400	.0500	.1688
NEURAL_L	.0921	.5621	.4432	.3444	.0076	.0327
NEURAL_E	.2154	.2715	.1108	.5181	.0695	.5619
BIER_han_L	.3800	.2800	.4400	.1400	.0500	.1700
BIER_han_E	.4800	.1900	.3900	.1100	.6600	.1000
BIER_L	.0106	.0439	.0362	.1115	.0027	.0137
BIER_E	.0012	.0136	.0227	.0225	.7061	.0440
POLY_L	.0793	.4401	.0525	.1143	.0144	.5239
POLY_E	.0019	.3310	.1059	.2018	.0090	.1277
RESET	.3933	.3103	.2875	.2953	.9339	.2066

Notes: a. Values denote p-values;
b. p-values less than .00005 are imputed as .0000;
c. p-values greater than .99995 are imputed as 1.000.

Table 6

Annual Growth Series					
	Δm	Δy	Δp	Δr_b	Δr_p
CoSTAR_L	.8600	.8900	.9100	.5600	.0120
CoSTAR_E	.9700	.0000	.8700	.0900	.1080
STAR_L	.9400	1.000	1.000	.8700	.0300
STAR_E	1.000	.0000	1.000	.1200	.1100
PSTAR_L	.7992	.6492	.7700	.1834	.2703
PSTAR_E	.7565	.9162	.8304	.1841	.3567
NEURAL_L	.8663	.7439	.6095	.0675	.2532
NEURAL_E	.8839	.9437	.6396	.2198	.4081
BIER_han_L	.5600	.1900	.3300	.2600	.3500
BIER_han_E	.3200	.5400	.7500	.4700	.3300
BIER_L	.3583	.0997	.2054	.0556	.0434
BIER_E	.3095	.2209	.1524	.0489	.1886
POLY_L	.9040	.1989	.3213	.1066	.1218
POLY_E	.8301	.3573	.4343	.0556	.1841
RESET	.7921	.9912	.6648	.6120	.4044

Appendix 3: Formal Proofs

Proof of Lemma 1. The proof follows almost directly from Lemma 1 of Bierens (1990), Stinchcombe and White (1994), or Theorem 1 of Bierens and Ploberger (1997). We provide details for completeness. Lemma 1 of Bierens (1990) states if $P\left(E\left[\epsilon|x\right]=0\right)<1$, then the set

$$S=\{\gamma\in\Gamma:E[\epsilon F(x,\gamma)]=0\},\ i=1...k \tag{39}$$

has Lebesgue measure zero where $F(x,\gamma)$ denotes the exponential function. Stinchcombe and White (1998) prove the result holds for essentially any $\mathfrak{F}$-measurable analytic function $F(x,\gamma)$, including the logistic and sin + cos

By Assumptions 2 and 3, x and $h(x,\delta)$ are $\mathfrak{F}$-measurable, $P(\inf_{\delta\in\Delta}|h(x,\delta)|$ $>0)=1$ and $\sup_{\delta\in\Delta}|h(x,\delta)|<\infty$ with probability one. Thus, if

$$P\left(E\left[\epsilon|x\right]=0\right)=1 \tag{40}$$

such that the null is true, then

$$\begin{aligned} P\left(\sup\nolimits_{\delta\in\Delta}E\left[h(x,\delta)\epsilon|x\right]=0\right) &= P\left(\sup\nolimits_{\delta\in\Delta}h(x,\delta)E\left[\epsilon|x\right]=0\right) \\ &= P\left(E\left[\epsilon|x\right]=0\right)=1. \end{aligned} \tag{41}$$

Under the alternative, we likewise deduce

$$\begin{aligned} P\left(\sup\nolimits_{\delta\in\Delta}E\left[h(x,\delta)\epsilon|x\right]=0\right) &= P\left(\sup\nolimits_{\delta\in\Delta}h(x,\delta)E\left[\epsilon|x\right]=0\right) \\ &= P\left(E\left[\epsilon|x\right]=0\right)<1. \end{aligned} \tag{42}$$

Both (41) and (42) imply we may simply redefine ϵ as $\sup_{\delta\in\Delta}|h(x,\delta)|\epsilon$ in (39), and apply Lemma 1 of Bierens (1990), or the generalized result of Stinchcombe and White (1998), to the scalar components $\sup_{\delta\in\Delta}|h_i(x,\delta)|\epsilon$. We deduce each set S_i has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^k$. ■

Proof of Theorem 2. Consider (i). Let x be the $k\times n$ design matrix $(x_1,...,x_n)$, and let $y=(y_1,...,y_n)$ and $\epsilon=(\epsilon_1,...,\epsilon_n)$ be $1\times n$ row vectors. By standard least squares algebra

$$\begin{aligned} \hat{\phi}_1 &= (xx')^{-1}xy' \\ \hat{\phi}_1' &= \phi_1'+\epsilon x'(xx')^{-1} \end{aligned} \tag{43}$$

Thus, the score evaluated under the null hypothesis of linearity reduces to

$$\begin{aligned}
\hat{s}(0,\gamma) &= \frac{1}{n}\sum\nolimits_{t=1}^{n}(y_t - \hat{\phi}_1' x_t)x_t F_t \qquad (44)\\
&= \frac{1}{n}\sum\nolimits_{t=1}^{n} x_t F_t(\epsilon_t - \epsilon x'(xx')^{-1}x_t)\\
&= \frac{1}{n}\sum\nolimits_{t=1}^{n} x_t F_t\left(\epsilon_t - \frac{1}{n}\sum\nolimits_{s=1}^{n}\epsilon_s x_s'(xx')^{-1}x_t\right)\\
&= \frac{1}{n}\sum\nolimits_{t=1}^{n} x_t F_t\left(\epsilon_t - x_t'\hat{A}^{-1}\frac{1}{n}\sum\nolimits_{s=1}^{n}\epsilon_s x_s\right)\\
&= \frac{1}{n}\sum\nolimits_{t=1}^{n} x_t F_t \epsilon_t - \frac{1}{n}\sum\nolimits_{t=1}^{n} x_t x_t' F\hat{A}^{-1}\frac{1}{n}\sum\nolimits_{t=1}^{n}\epsilon_t x_t\\
&= \frac{1}{n}\sum\nolimits_{t=1}^{n}\left(F_t I_k - \hat{b}\hat{A}^{-1}\right)x_t\epsilon_t\\
&= \frac{1}{n}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t
\end{aligned}$$

say, where

$$\begin{aligned}
\hat{g}_t(\gamma) &= F_t(\gamma)I_k - \hat{b}(\gamma)\hat{A}^{-1} \qquad (45)\\
\hat{b}(\gamma) &= \frac{1}{n}\sum\nolimits_{t=1}^{n} x_t x_t' F_t(\gamma)\\
\hat{A}^{-1} &\equiv \frac{1}{n}\sum\nolimits_{t=1}^{n} x_t x_t',
\end{aligned}$$

and where I_k denotes the k-dimensional identity matrix. By Assumption 1, ϵ_t is a martingale difference sequence under the null, hence

$$E\left(\epsilon_t|\mathfrak{F}_{t-1}\right) = 0. \qquad (46)$$

Because x_t and F_t are $\mathfrak{F}_{t-1}$-measurable, it follows that $\hat{g}_t(\gamma)x_t$ is $\mathfrak{F}_{t-1}$-measurable, hence $\hat{g}_t(\gamma)x_t\epsilon_t$ forms a martingale difference sequence:

$$E\left([\hat{g}_t(\gamma)x_t]_i\,\epsilon_t|\mathfrak{F}_{t-1}\right) = 0,\ i = 1...k. \qquad (47)$$

Therefore, by Assumptions 1-4, the Slutsky Theorems and the martingale central limit theorem, cf. McLeish (1974), the sequence $(1/\sqrt{n})\sum_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t$ converges in law jointly to a Gaussian random vector. In particular,

$$\sqrt{n}\hat{s}(0,\gamma) = \frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t \Longrightarrow N(0,V(\gamma)), \qquad (48)$$

pointwise in Γ for some covariance matrix, $V(\gamma)$.

The covariance matrix $V(\gamma)$ will be the point-wise probability limit of

$$n\hat{s}(0,\gamma)\hat{s}(0,\gamma)' = \left[\frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t\right]\left[\frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t\right]', \qquad (49)$$

provided the limit exists. In particular, by the martingale difference property of the innovations ϵ_t (and $\hat{g}_t(\gamma)x_t\epsilon_t$) under H_0, we deduce

$$
\begin{aligned}
& \left[\frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t\right]\left[\frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t\right]' \qquad (50)\\
= \quad & \frac{1}{n}\sum\nolimits_{t=1}^{n}\epsilon_t^2\hat{g}_t(\gamma)x_tx_t'\hat{g}_t(\gamma)' + o_p(1)\\
= \quad & \frac{1}{n}\sum\nolimits_{t=1}^{n}\epsilon_t^2\left(F_tI_k - \hat{b}(\gamma)\hat{A}^{-1}\right)x_tx_t'\left(F_tI_k - \hat{b}(\gamma)\hat{A}^{-1}\right)' + o_p(1),
\end{aligned}
$$

where the term $o_p(1)$ contains the cross-products $\epsilon_s\epsilon_t$, $s \neq t$, and follows from ergodicity and the martingale-difference property of ϵ_t, cf. Assumption 1. By stationarity and Assumption 4, we obtain the probability limits

$$
\hat{b}(\gamma) \to b(\gamma), \quad \hat{A} \to A, \qquad (51)
$$

pointwise in γ, and therefore by standard properties of functional probability limits we deduce by the weak law of large numbers and Assumptions 1-4

$$
\begin{aligned}
& \left[\frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t\right]\left[\frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n}\hat{g}_t(\gamma)x_t\epsilon_t\right]' \qquad (52)\\
= \quad & \frac{1}{n}\sum\nolimits_{t=1}^{n}\epsilon_t^2\left(F_tI_k - \hat{b}(\gamma)\hat{A}^{-1}\right)x_tx_t'\left(F_tI_k - \hat{b}(\gamma)\hat{A}^{-1}\right)' + o_p(1)\\
\to \quad & E\left[\epsilon_t^2\left\{F_tI_k - b(\gamma)A^{-1}\right\}x_tx_t'\left\{F_tI_k - A^{-1}b(\gamma)\right\}\right].
\end{aligned}
$$

Finally, consider (ii). Under H_1 the innovations ϵ_t form a white noise process, hence $\hat{\phi}_1$ is a consistent estimator of the k-vector ϕ_1. Thus, (22) follows from Assumptions 1-4, Lemma 1 and the law of large numbers. In particular,

$$
\frac{1}{n}\sum\nolimits_{t=1}^{n}(y_t - \hat{\phi}_1'x_t)x_{i,t}F_t(\gamma) \to E\left(\epsilon_tx_{i,t}F_t(\gamma)\right) \neq 0 \qquad (53)
$$

for each $i = 1...k$ with probability one. Therefore, $\eta(\gamma) = E\left(\epsilon_tx_tF_t\right) \neq 0$, *a.s.* ■

Proof of Lemma 3. Because x_t does not contain a constant term, x_t is a p-vector. Consider any $\gamma \in S^*$, and notice that

$$
\begin{aligned}
V(\gamma) \quad &= \quad E\left[\epsilon_t^2\left\{F_t(\gamma)I_k - b(\gamma)A^{-1}\right\}x_tx_t'\left\{F_t(\gamma)I_k - A^{-1}b(\gamma)\right\}\right] \qquad (54)\\
&= \quad E\left[z_t(\gamma)z_t(\gamma)'\epsilon_t^2\right]
\end{aligned}
$$

where we define the p-vector $z_t(\gamma)$ as

$$
z_t(\gamma) = \left\{F_t(\gamma)I_k - b(\gamma)A^{-1}\right\}x_t. \qquad (55)
$$

For simplicity, we will drop the argument γ from $z_t(\gamma)$. Recall that $F_t(\gamma) \equiv F_t(x_t, \gamma)$ is $\mathfrak{F}_{t-1}$-measurable.

Step 1 ($r'V(\gamma)r = 0$): Because $V(\gamma)$ is non-positive definite for every $\gamma \in S^*$, there exists a p-vector $r \in \mathbb{R}^p$, $r \neq 0$, such that[10]

$$r'V(\gamma)r = 0, \tag{56}$$

which reduces to

$$\begin{aligned} r'V(\gamma)r &= r'E\left[z_t z_t' \epsilon_t^2\right] r \\ &= r'E\left[z_t z_t{}' E\left(\epsilon_t^2 | x_t\right)\right] r \\ &= E\left[r' z_t z_t{}' r E\left(\epsilon_t^2 | x_t\right)\right] \\ &= E\left[\left(\sum\nolimits_{i=1}^p r_i z_{t,i}\right)^2 E\left(\epsilon_t^2 | x_t\right)\right] = 0. \end{aligned} \tag{57}$$

By Assumption 5, $E\left(\epsilon_t^2 | x_t\right) > 0$ with probability one, hence the equality holds *if and only if*

$$\sum\nolimits_{i=1}^p r_i z_{t,i} = 0, \; a.s., \tag{58}$$

for every $r \in \mathbb{R}^p$, $r \neq 0$.

Now, define $D_t = (d_{t,i,j})_{i,j=1}^p \equiv F_t(\gamma) I_k - b(\gamma) A^{-1}$. Then $z_t = D_t x_t$. Recalling $x_t = (y_{t-1}, ..., y_{t-p})'$ we deduce (58) holds for every $r \neq 0$ *if and only if*

$$\begin{aligned} \sum\nolimits_{i=1}^p r_i z_{t,i} &= \left[\sum\nolimits_{i=1}^p r_i \left(\sum\nolimits_{j=1}^p d_{t,i,j} x_{t,j}\right)\right] \\ &= \left[\sum\nolimits_{i=1}^p r_i \left(\sum\nolimits_{j=1}^p d_{t,i,j} y_{t-j}\right)\right] \\ &= \sum\nolimits_{i=1}^p y_{t-i} \left(\sum\nolimits_{j=1}^p d_{t,i,j} r_j\right) \\ &= \sum\nolimits_{i=1}^p y_{t-i} e_{t,i}(r) = 0, \; a.s. \end{aligned} \tag{59}$$

where we define $e_{t,i}(r) \equiv \sum_{j=1}^p d_{t,i,j} r_j$. Notice $e_{t,i}(r)$ is $\mathfrak{F}_{t-1}$-measurable because

$$\begin{aligned} d_{t,i,j} &= -\left[b(\gamma)A^{-1}\right]_{i,j}, \;\; i \neq j \\ &= F_t(\gamma) - \left[b(\gamma)A^{-1}\right]_{i,j}, \;\; i = j, \end{aligned} \tag{60}$$

and $F_t(\gamma)$ is $\mathfrak{F}_{t-1}$-measurable.

Step 2 ($e_{t,i}(r) = 0$): For the next step of the proof, we demonstrate $e_{t,i}(r) = 0$ with probability one for each $i = 1...p$ and arbitrary $r \in \mathbb{R}^p$, $r \neq 0$, by exploiting standard metric projection theory for Hilbert spaces.

[10] Clearly

$$\begin{aligned} r'V(\gamma)r &= r'E\left[z_t z_t{}' E\left(\epsilon_t^2 | x_t\right)\right] r \\ &= E\left[r' z_t z_t{}' r E\left(\epsilon_t^2 | x_t\right)\right] \\ &= E\left[\left(\sum\nolimits_{i=1}^k r_i z_{t,i}\right)^2 E\left(\epsilon_t^2 | x_t\right)\right] \geq 0 \end{aligned}$$

therefore is suffices to consider only the indefinite case with equality.

Denote by $\mathfrak{L}_t$ the space $L_2(\Omega, \mathfrak{F}_t, Q)$. Consider the process $\{\tilde{y}_{t-i}\} \equiv \{y_{t-i}e_{t,i}\}$, construct the L_2-projection $P(\tilde{y}_{t-i}|\mathfrak{L}_{t-i-1})$ for each $\tilde{y}_{t-i}$ into $\mathfrak{L}_{t-i-1}$, and deduce an orthogonal projection error

$$\begin{aligned} \tilde{y}^*_{t-i} &\equiv P(\tilde{y}_{t-i}|\mathfrak{L}_{t-i-1}) \\ u_{t-i} &\equiv \tilde{y}_{t-i} - \tilde{y}^*_{t-i} \perp \mathfrak{L}_{t-i-1}. \end{aligned} \tag{61}$$

By L_2-orthogonality of the projection error, we know

$$E(u_{t-i}w_{t-i-1}) = 0$$

for every element $w_{t-i-1} \in \mathfrak{L}_{t-i-1}$. Multiply $\sum_{i=1}^{p} y_{t-i}e_{t,i}(r)$ by u_{t-1}, and take the expectation. From (59) we then deduce

$$\sum\nolimits_{i=1}^{p} y_{t-i}e_{t,i}(r)u_{t-1} = \sum\nolimits_{i=1}^{p} \tilde{y}_{t-i}u_{t-1} = 0, \ a.s., \tag{62}$$

thus, with probability one we have

$$\begin{aligned} 0 &= E\left(\sum\nolimits_{i=1}^{p} \tilde{y}_{t-i}u_{t-1}\right) \\ &= E(\tilde{y}_{t-i}u_{t-1}) + \sum\nolimits_{i=2}^{p} E(\tilde{y}_{t-i}u_{t-1}) \\ &= E(\tilde{y}_{t-1}u_{t-1}) + 0. \end{aligned} \tag{63}$$

The third line follows from orthogonality: $u_{t-1} \perp \mathfrak{L}_{t-2}$ implies $E(\tilde{y}_{t-i}u_{t-1}) = 0$, $i = 2...p$, because each $\tilde{y}_{t-i} \in \mathfrak{L}_{t-i} \subseteq \mathfrak{L}_{t-2}$ for $i \geq 2$. Hence,

$$\begin{aligned} 0 &= E(\tilde{y}_{t-1}u_{t-1}) \\ &= E\left(\left(\tilde{y}_{t-1} - \tilde{y}^*_{t-1}\right)u_{t-1}\right) + E\left(\tilde{y}^*_{t-1}u_{t-1}\right) \\ &= E\left(u^2_{t-1}\right) + E\left(\tilde{y}^*_{t-1}u_{t-1}\right) \\ &= E\left(u^2_{t-1}\right), \end{aligned} \tag{64}$$

where we again exploit orthogonality: $E\left(\tilde{y}^*_{t-1}u_{t-1}\right) = 0$ because $\tilde{y}^*_{t-1} \in \mathfrak{L}_{t-2}$ and $u_{t-1} \perp \mathfrak{L}_{t-2}$.

Clearly $E\left(u^2_{t-1}\right) = 0$ *if and only if* $u_{t-1} = 0$ with probability one. By the assumption the σ-fields are strictly increasing we have $\mathfrak{L}_{t-2} \subset \mathfrak{L}_{t-1}$, hence the only way for the L_2-projection error to be identically zero with probability one *for any* $r \neq 0$ is if $\tilde{y}_{t-1} = y_{t-1}e_{t,1}(r)$ is induced by $\mathfrak{F}_{t-2}$ with probability one: see, e.g., Brockwell and Davis (1987). Because y_{t-1} is $\mathfrak{F}_{t-1}$-measurable with a non-degenerate distribution, however, we deduce $\sigma(y_{t-1}e_{t,1}(r)) \in \mathfrak{F}_{t-2}$ is possible *if and only if* $e_{t,1}(r) = 0$ with probability one for any $r \in \mathbb{R}^k$, $r \neq 0$.

Notice that $y_{t-1}e_{t,1}(r) = e \neq 0$, with probability one for some non-zero constant e, is ruled out because $e_{t,1}(r)$ is a function of r, y_{t-1} is $\mathfrak{F}_{t-1}$-measurable and governed by a nondegenerate marginal distribution, and A^{-1} is a non-zero matrix: it is straightforward to show that if $y_{t-1}e_{t,1}(r_0) = e \neq 0$ for some $r_0 \neq 0$, then there exists an $r \neq r_0$, $r \neq 0$, such that $y_{t-1}e_{t,1}(r) \neq e$, and therefore $y_{t-1}e_{t,1}(r)$ cannot be non-zero constant-valued.

The proof that the remaining $e_{t,i}(r) = 0$ with probability one follows in an identical manner: for each $e_{t,k}(r)$, impose $e_{t,i}(r) = 0$ for $i = 1...k-1$, multiply $\sum_{i=k}^{p} y_{t-i}e_{t,i}(r)$ by u_{t-k}, and inmate the logic above in order to deduce $e_{t,k}(r)$ $= 0$.

Step 3 ($F_t = c$):

The identity $e_{t,i}(r) = 0$, *a.s.*, for each $i = 1...p$, implies

$$e_{t,i}(r) = \sum\nolimits_{j=1}^{p} d_{t,i,j}r_j = 0, \text{ } a.s., \tag{65}$$

for every $r \in \mathbb{R}^p$, $r \neq 0$. Because $r \neq 0$ can be chosen arbitrarily, we deduce $\sum_{j=1}^{p} d_{t,i,j}r_j = 0$ with probability one for any $r \neq 0$ *if and only if* $d_{t,i,j} = 0$ for each $i, j = 1...p$. For example, choose $r = (0, ..., 1, ...0)'$ where the 1 is placed in the i^{th} row, $i = 1...p$. Then

$$e_{t,i}(r) = \sum\nolimits_{j=1}^{p} d_{t,i,j}r_j = d_{t,i,i} = F_t(\gamma) - \left[b(\gamma)A^{-1}\right]_{i,i} = 0, \text{ } a.s., \tag{66}$$

hence

$$F_t(\gamma) = \left[b(\gamma)A^{-1}\right]_{i,i} = 0, \text{ } a.s., \tag{67}$$

This implies the $\mathfrak{F}_{t-1}$-measurable $F_t(\gamma)$ is constant-valued with probability one for any $\mathfrak{F}_{t-1}$-measurable x_t, because $b(\gamma)$ and A contain only constants (by the assumption of stationarity in Assumption 1).

Because $\gamma \in \Gamma$ is bounded, and due to the boundedness of x_t with probability one, $x_t \neq 0$ *a.s.*, and by the assumption that $x_t = v_t = (y_{t-1}, ..., y_{t-p})'$ does not contain a constant term, $F_t(x_t, \gamma)$ is constant-valued with probability one *if any only if* $\gamma = 0$ *a.s.* Therefore, for any $\gamma \in S^*$ it must be the case that γ $= 0$, hence S^* has Lebesgue measure zero. ■

References

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

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

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

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

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

[6] Brockwell, P.J. and R.A. Davis, 1987, Time Series: Theory and Methods (Springer: New York).

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

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

[9] Corradi, V., and N. Swanson, 2002, A Consistent Test for Nonlinear Out of Sample Prediction Accuracy, Journal of Econometrics 110, 353-381.

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

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

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

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

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

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

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

[17] Friedman, B., and K. Kuttner, 1992, Money, Income, Prices, and Interest Rates, American Economics Review 82, 472-492.

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

[19] Granger, C.W.J. and T. Teräsvirta, 1996, Modeling Nonlinear Economic Relationships (Oxford University Press).

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

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

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

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

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

[25] Jennrich, R.I., Asymptotic Properties of Nonlinear Least Squares Estimators, Annals of Mathematical Statistics 40, 633-643.

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

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

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

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

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

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

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

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

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

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

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

[37] Skalin, J., 1998, Testing Linearity against Smooth Transition Autoregression Using A Parametric Bootstrap, Working Paper Series in Economics and Finance No. 276, Dept. of Economic Statistics, Stockholm School of Economics.

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

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

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

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

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

[43] Tsay, R., 1989, Testing and Modeling Threshold Autoregressive Processes, Journal of the American Statistical Association 84, 231-240.

[44] 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).

[45] Wooldridge, J., 1991, On the Application of Robust, Regression- Based Diagnostics to Models of Conditional Means and Conditional Variances, Journal of Econometrics 47, 5-46.

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

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