SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

**CREATES Research Paper 2010-36**

# Linearity Testing in Time-Varying Smooth Transition Autoregressive Models under Unknown Degree of Persistency

Robinson Kruse and Rickard Sandberg

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

# Linearity Testing in Time-Varying Smooth Transition Autoregressive Models under Unknown Degree of Persistency

Robinson Kruse* and Rickard Sandberg†

*CREATES, School of Economics and Management, Aarhus University
Building 1322, DK-8000 Aarhus, Denmark

†Department of Economic Statistics, Stockholm School of Economics
P.O. Box 6501, SE-113 83 Stockholm, Sweden

July 26, 2010

**Abstract**

Building upon the work of Vogelsang (1998) and Harvey and Leybourne (2007) we derive tests that are invariant to the order of integration when the null hypothesis of linearity is tested in time-varying smooth transition models. As heteroscedasticity may lead to spurious rejections of the null hypothesis, a White correction is also considered. The asymptotic properties of the tests are studied. Our Monte Carlo simulations suggest that the newly proposed tests exhibit good size and competitive power properties. An empirical application to US inflation data from the Post-Bretton Woods period underlines the empirical usefulness of our tests.

Key Words: Linearity testing, Linear I(0) and (1) models, Non-linear I(0) and I(1) models, White correction.

## 1 Introduction

Ample empirical evidence on the short-comings of AR(I)MA models to capture non-linearities and structural changes in economic time-series have been gathered over the years. The research during the last two or three decades has therefore very much focused on time-series models accommodating both non-linearity and structural change in the dynamics and the deterministic terms. In a sound modelling cycle of such models testing the linearity hypothesis is of obvious

*E-mail: rkruse@creates.au.dk. Helpful comments by Sanne Hiller and financial support from CREATES funded by the Danish National Research Foundation are gratefully acknowledged.

†E-mail: rickard.sandberg@hhs.se. Corresponding author.

interest. A popular method to test linearity is based on Taylor-series approximations of the original model. Thereafter tests are conducted by simple ordinary least squares (OLS), see Luukkonen, Saikkonen, and Teräsvirta (1988), Granger and Teräsvirta (1993), Teräsvirta (1994), Jansen and Teräsvirta (1996), and van Dijk, Teräsvirta, and Franses (2002) to list a few. Other frequently used linearity tests are simulation based, see e.g. Andrews and Ploberger (1994) and Hansen (1996).

It is important, though, to keep in mind that the asymptotic distribution of these linearity tests are not invariant with respect to the order of integration under the null hypothesis of linearity (see Kiliç 2004, and Sandberg 2008 for discussions and examples). This is an undesirable property when they are applied to potentially high and low persistent time-series. Thus, it is not obvious if a linear I(0) or a linear I(1) model should serve as null hypothesis. An obvious remedy employs a unit root pre-test and subsequently works with levels or first-differences of the series. Pre-testing is problematic since unit root tests may exhibit low power. Moreover, the overall significance level is uncontrolled in such a procedure and the multiple testing problem arises.

Our main contribution is to follow up and extend the work on linearity tests in smooth transition autoregressive (STAR) models by Harvey and Leybourne (2007), hereafter H&L, which are invariant to the order of integration. More specifically, we derive invariant linearity tests in the more general time-varying STAR (TV-STAR) model (see e.g. Van Dijk and Franses, 2002 and Lundbergh, Teräsvirta, and van Dijk, 2003).[1] In this sequel we rely upon the seminal work by Vogelsang (1998). In our case, this yields a Wald test statistic which exhibits the same critical values regardless whether a linear I(0) or a linear I(1) model is considered under the null hypothesis. This test is also shown to be consistent against non-linear I(0) or non-linear I(1) TV-STAR models. In addition to the work by H&L, we allow for a linear trend-specification and also consider heteroscedasticity robust linearity tests. Having macroeconomic time-series in mind where both trend and the presence of GARCH effects are notable, these extensions seem very natural.

The TV-STAR model is a natural extension of the STAR model, and it does not only account for a regime-switching behavior (the STAR part) but also parameter instability (structural change). As Perron (2006) points out, structural change is of substantial importance for the modelling of economic time-series. The TV-STAR model is frequently used to resemble the behavior of macroeconomic variables (see e.g. Lundbergh, van Dijk, and Teräsvirta, 2003 for an application to 214 U.S. macroeconomic time-series) for which a debate is still ongoing whether they are best characterized as difference or trend stationary. Due to this dilemma it is common practice to conduct two separate tests; one based on first-differences and another one on de-trended data.[2]

[1] The linearity tests in the TV-STAR model proposed by e.g. van Dijk, Teräsvirta, and Franses (2002) and Lundbergh, Teräsvirta, and van Dijk (2003) are not invariant with respect to the order of integration. The large sample results for their tests are instead based on that the model under the null hypothesis is a linear stationary autoregressive process.

[2] This was also the approach taken on by Lundbergh, Teräsvirta, and van Dijk (2003).

Evidently, such an approach may comprise conflicting results. Therefore, it is essential to derive invariant linearity tests for TV-STAR models.

A few words on the notation in this work: $\Rightarrow$ signifies weak convergence, $\xrightarrow{p}$ denotes convergence in probability, $B(s)$ abbreviates a standard Brownian motion on $[0, 1]$, and integrals of the type $\int_0^1 B(s)ds$ and $\int_0^1 B(s)dB(s)$ are denoted $\int B$ and $\int BdB$ for short.

The rest of the work is organized as follows. Section 2 presents trending non-linear I(0) and I(1) TV-STAR processes, corresponding approximations, and also a hybrid specification regression model. Testing procedures and large sample results for robust and invariant linearity tests are given in Section 3. The finite sample properties of the tests are examined by Monte Carlo experiments in Section 4. An empirical application is given in section 5. Conclusions are drawn in Section 6. Finally, mathematical proofs are provided in the Appendix.

## 2 The Models

### 2.1 Non-linear I(0) and I(1) Models

Consider a stochastic process $\{y_t\}$ generated by

$$y_t \equiv \mu_t + v_t, \quad t = 1, ..., T, \tag{1}$$

where $\mu_t = 0$, $\mu_t = d_0$, and $\mu_t = d_0 + d_1 t$ and are referred to as cases a, b, and c, respectively, and $v_t$ is modelled via a first-order TV-STAR process

$$\begin{aligned} v_t \quad \equiv \quad & [\phi_1 v_{t-1} \{1 - G_1(v_{t-1}; \gamma_1)\} + \phi_2 v_{t-1} G_1(v_{t-1}; \gamma_1)] [1 - G_2(t^*; \gamma_2, c)] \\ & + [\phi_3 v_{t-1} \{1 - G_1(v_{t-1}; \gamma_1)\} + \phi_4 v_{t-1} G_1(v_{t-1}; \gamma_1)] G_2(t^*; \gamma_2, c) + \epsilon_t. \end{aligned} \tag{2}$$

Here, $(\phi_1, ..., \phi_4)'$ is a real-valued parameter vector, the starting value $v_0$ is assumed fixed, $\epsilon_t$ is an error term with properties discussed below, and the bounded non-linear functions $G_1(v_{t-1}; \gamma_1)$ and $G_2(t^*; \gamma_2, c_2)$ are defined by

$$G_1(v_{t-1}; \gamma_1) \equiv \frac{1}{1 + \exp\{-\gamma v_{t-1}\}}, \tag{3}$$

where $\gamma_1 \geq 0$ (an identifying restriction) and

$$G_2(t^*; \gamma_2, c_2) \equiv \frac{1}{1 + \exp\{-\gamma_2 (t^* - c)\}}, \tag{4}$$

where $t^* = t/T$, $\gamma_2 \geq 0$ (another identifying restriction), and $0 \leq c \leq 1$.[3] In (3) and (4), $\gamma_1$ and $\gamma_2$ are parameters which controls for the smoothness, and $c$ is a non-centrality parameter. One

[3]Defining the smooth transition function in (4) in terms of $t^* = t/T$ (rather than $t$) turns out to be convenient when giving the proofs in the Appendix. Yet another advantage is that the speed of transition parameter $\gamma_2$ becomes scale-free.

appealing feature with these functions is that not only smooth non-linearities are captured but also the Heavside function (the step function) and the constant function can be approximated. Specifically, the former function is obtained letting $\gamma_1 \to \infty$ ($\gamma_2 \to \infty$) which yields that $G_1 = 0$ ($= G_2$) when $v_{t-1} < 0$ ($t^* < c$) and $G_1 = 1$ ($= G_2$) when $v_{t-1} \geq 0$ ($t^* \geq c$). On the contrary, the latter function is comprised letting $\gamma_1 \to 0$ ($\gamma_2 \to 0$) which implies that $G_1 = 1/2$ ($= G_2$). We finally notice that $G_1$ ($G_2$) is bounded between 0 and 1 and is a non-decreasing function in $v_{t-1}$ ($t^*$).

The TV-STAR process is preferably interpreted as describing $v_t$ as a STAR process, with the transition variable $v_{t-1}$, at all times. That is, for any fixed $t^* = t_0$ the TV-STAR process accommodates a continuum of regimes for the dynamic root which increases from $\phi_1 + G_2(t_0)(\phi_3 - \phi_1)$ to $\phi_2 + G_2(t_0)(\phi_4 - \phi_2)$ with $v_{t-1}$. Furthermore, in the beginning of the sample these roots equal $\phi_1$ and $\phi_2$ (associated with $G_2(t^*) = 0$) and the corresponding roots at the end of the sample are $\phi_3$ and $\phi_4$ (associated with $G_2(t^*) = 1$). To this end, there are three nested models of particular interest within the framework of the TV-STAR process. First, letting $\phi_1 = \phi_2$ and $\phi_3 = \phi_4$ with $\phi_1 \neq \phi_3$ yields a first-order time-varying autoregressive (TV-AR) process (see e.g. Jansen and Teräsvirta, 1996). Second, instead letting $\phi_1 = \phi_3$ and $\phi_2 = \phi_4$ with $\phi_1 \neq \phi_2$ implies a STAR process. Finally, imposing the restriction $\phi_1 = \phi_2 = \phi_3 = \phi_4$, we arrive at a first-order linear autoregressive process.

To the best of our knowledge, the statistical properties of the TV-STAR process in (2) are not yet fully established in the literature. It is evident, though, that the usual definitions of weak stationarity or geometric ergodicity can not be applied because the TV-STAR process has (for instance) a time-varying variance and can not be written as a time-homogeneous Markow chain. Despite this, it seems that a (heuristic) stability condition for the TV-STAR process can be obtained by combining the stability results for pure TV-AR processes (which in essence states that the roots of the time-varying characteristic equation should be larger than one in modulus at all times) by e.g. Juntunen, Tervo, and Kaipio (1999), with the results for pure STAR processes (that is, the roots associated with the two extreme regimes are less than unity in absolute values) by e.g. Liebscher (2005) or Meitz and Saikkonen (2008). More specifically, because the TV-STAR process for any fixed point in time ($t^* = t_0$) can be expressed as a STAR process with dynamic roots associated to the two extreme regimes given by $\phi_1 + G(t_0)(\phi_3 - \phi_1)$ and $\phi_2 + G(t_0)(\phi_4 - \phi_2)$, it seems reasonable to assume that these roots must lie inside the unit circle for any $t_0 \in [0, 1]$. Thus, we arrive at the following stability condition for the above TV-STAR process.[4]

**Stability Condition:** *If $|\phi_1 + \kappa(\phi_3 - \phi_1)| < 1$ and $|\phi_2 + \kappa(\phi_4 - \phi_2)| < 1$ hold for all $\kappa \in [0, 1]$ in (2), then we say that the TV-STAR process in (2) is stable.*[5]

[4] A more rigorous treatment of the statistical properties of the TV-STAR model is beyond the scope of this work.

[5] It is seen that this condition reduces to the stability condition by Meitz and Saikkonen (2008, p. 463) for pure STAR models letting $\phi_1 = \phi_3$ and $\phi_2 = \phi_4$ with $\phi_1 \neq \phi_2$. Moreover, instead letting letting $\phi_1 = \phi_2$ and $\phi_3 = \phi_4$ with $\phi_1 \neq \phi_3$ the condition reduces to the stability condition by Juntonen, Tervo, and Kaipio (p. 396) for pure

If the TV-STAR process in (2) satisfy the above stability condition we say that the resultant process $\{y_t\}$ in (1) is stable around $\mu_t$, and such a process is referred to as a non-linear I(0) model.

Having introduced a non-linear I(0) model, it seems that one possibility to define a non-linear I(1) model is via a stable TV-STAR process in first-differences. Hence, we may define a non-linear I(1) model by

$$y_t \equiv \mu_t + v_t, \quad t = 1, ..., T, \tag{5}$$

where

$$\begin{aligned} \Delta v_t \quad \equiv \quad & \left[\psi_1 \Delta v_{t-1} \left\{1 - G_1(\Delta v_{t-1}; \gamma_1)\right\} + \psi_2 v_{t-1} G_1(\Delta v_{t-1}; \gamma_1)\right]\left[1 - G_2(t^*; \gamma_2, c)\right] \\ & + \left[\psi_3 \Delta v_{t-1} \left\{1 - G_1(\Delta v_{t-1}; \gamma_1)\right\} + \psi_4 v_{t-1} G_1(\Delta v_{t-1}; \gamma_1)\right] G_2(t^*; \gamma_2, c) + \epsilon_t, \end{aligned} \tag{6}$$

and $\Delta$ abbreviates the lag-operator.[6] In (6), $(\psi_1, ..., \psi_4)'$ is a real-valued parameter vector, the initial values $v_{-1}$ and $v_0$ are assumed fixed, and the smooth transition functions $G_1$ and $G_2$ are defined as in (3) and (4), respectively, but the transition variable in (3) is now replaced with $\Delta v_{t-1}$. Moreover, the TV-STAR process in first-differences in (6) is stable if its autoregressive parameters satisfy the above Stability Condition. Accordingly, the first-differences of the process in (5) $\{\Delta y_t\}$ is stable around $\Delta \mu_t$, and the resultant model in levels $y_t = \Delta \mu_t + y_{t-1} + \Delta v_t$ is henceforth referred to as a non-linear I(1) model. Towards this end, it is noticed our terminology of non-linear I(0) and I(1) models is somewhat different from that in H&L because they base their work on logistic and exponential STAR processes corresponding to our case b.

## 2.2 Hybrid Regression Specification Models

To facilitate our testing situation we shall approximate our non-linear I(0) and I(1) models by instead using the first-order Taylor-series expansions of the (smooth) transition functions (3) and (4) around zero for the speed of transition parameters. This yields the following approximation to the TV-STAR process in levels

$$v_t = \delta_0 v_{t-1} + \delta_1 t^* v_{t-1} + \delta_2 v_{t-1}^2 + \delta_3 t^* v_{t-1}^2 + \epsilon_t,$$

and the approximation to the TV-STAR process in first-differences is given by

$$\Delta v_t = \lambda_0 \Delta v_{t-1} + \lambda_1 t^* \Delta v_{t-1} + \lambda_2 \left(\Delta v_{t-1}\right)^2 + \lambda_3 t^* \left(\Delta v_{t-1}\right)^2 + \epsilon_t.$$

Next, since we are interested in linear and non-linear I(0) and I(1) alternatives a hybrid specification regression equation can be obtained by combining above approximations into one expression.

TV-AR models . Finally, the stability results for an AR(1) process follows by letting $\phi_1 = \phi_2 = \phi_3 = \phi_4$.

[6] It should be noticed that the definition of $v_t$ in (2) does not imply the definition for $\Delta v_t$ in (6) because $\Delta$ is a linear operator.

To accomplish this we allow us to write

$$y_t = \mu_t + z_t, \quad t = 1, ..., T, \tag{7}$$

where $\mu_t$ is defined as in (1), and

$$\begin{aligned} z_t &= \delta_0 z_{t-1} + \delta_1 t^* z_{t-1} + \delta_2 z_{t-1}^2 + \delta_3 t^* z_{t-1}^2 \\ &+ \lambda_0 \Delta z_{t-1} + \lambda_1 t^* \Delta z_{t-1} + \lambda_2 \left(\Delta z_{t-1}\right)^2 + \lambda_3 t^* \left(\Delta z_{t-1}\right)^2 + \epsilon_t. \end{aligned} \tag{8}$$

Now, consider first maintained linear I(0) and I(1) models by (7) and (8). If $\delta_1 = \delta_2 = \delta_3 = \lambda_1 = \lambda_2 = \lambda_3$,[7] letting $\lambda_0 = 0$ and assuming $\delta_0 \in (-1, 1)$, $z_t$ is a stationary AR(1) process and yields that $y_t$ is a linear I(0) model for the cases $\mathsf{a}$ and $\mathsf{b}$ and linear I(0) model with a drift in case $\mathsf{c}$; instead letting $\delta_0 = 1$ and assuming $\lambda_0 \in (-1, 1)$, $\Delta z_t$ is a stationary AR(1) model implying that $y_t$ is a linear I(1) model possibly with a drift in case $\mathsf{c}$.

Considering next maintained (approximate) non-linear I(0) and I(1) models by (7) and (8). Letting $\lambda_1 = \lambda_2 = \lambda_3 = 0$ and $\delta_k \neq 0$ for at least one $k = 1, 2, 3$, we shall assume that the resultant process for $\{y_t\}$ is a stable non-linear I(1) model. Instead letting $\delta_1 = \delta_2 = \delta_3 = 0$ and $\lambda_k \neq 0$ for at least one $k = 1, 2, 3$, we will assume that the implied process for $\{\Delta y_t\}$ is a non-linear I(0) model.

The model that is used in practice, and which also the subsequent linearity tests are build upon, is a model expressed in terms of observed values That is, we substitute for $z_t = y_t - \mu_t$ and $\Delta z_t = \Delta y_t - \Delta \mu_t$ into (8) to obtain the hybrid regression specification model

$$y_t = \beta_m' x_t^m + \epsilon_t, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}, \tag{9}$$

where

$$\begin{aligned} \beta_{\mathsf{a}} &= (\beta_1, \beta_2, ..., \beta_8)', \\ \beta_{\mathsf{b}} &= (\beta_1, \beta_2, ..., \beta_{10})', \\ \beta_{\mathsf{c}} &= (\beta_1, \beta_2, ..., \beta_{13})', \end{aligned}$$

[7]It may be noticed that these (linearity) restrictions are implications of letting $\gamma_1 = 0$ and $\gamma_2 = 0$ in the smooth transition functions, i.e. if the speed of transition parameters are equated to zero the TV-STAR process yields a linear model.

and

$$
\begin{aligned}
x_t^{\mathsf{a}} &= (y_{t-1}, \Delta y_{t-1}, t^* y_{t-1}, y_{t-1}^2, t^* y_{t-1}^2, t^* \Delta y_{t-1}, (\Delta y_{t-1})^2, t^*(\Delta y_{t-1})^2)', \\
x_t^{\mathsf{b}} &= (1, y_{t-1}, \Delta y_{t-1}, t^*, t^* y_{t-1}, y_{t-1}^2, t^* y_{t-1}^2, t^* \Delta y_{t-1}, (\Delta y_{t-1})^2, t^*(\Delta y_{t-1})^2)', \\
x_t^{\mathsf{c}} &= (1, t^*, t^{*2}, t^{*3}, y_{t-1}, \Delta y_{t-1}, t^* y_{t-1}, t^{*2} y_{t-1}, y_{t-1}^2, t^* y_{t-1}^2, t^* \Delta y_{t-1}, (\Delta y_{t-1})^2, t^*(\Delta y_{t-1})^2)'.
\end{aligned}
$$

# 3 Testing Procedures

## 3.1 The Null Hypotheses of Linearity

The null hypothesis of linearity for the hybrid regression specification model in (9), which does not specify if $y_t$ is a linear I(0) or a linear(1) model, can now for the three cases $\mathsf{a}$, $\mathsf{b}$, and $\mathsf{c}$ be expressed as

$$
\begin{aligned}
\mathsf{H}_0^{\mathsf{a}} &: \beta_3 = \cdots = \beta_8 = 0, \\
\mathsf{H}_0^{\mathsf{b}} &: \beta_4 = \cdots = \beta_{10} = 0, \\
\mathsf{H}_0^{\mathsf{c}} &: \beta_5 = \cdots = \beta_{13} = 0,
\end{aligned}
$$

resulting in the restricted hybrid specification model

$$
y_t = \beta_{m,r}' x_t^{m,r} + \epsilon_t, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}, \tag{10}
$$

where

$$
\begin{aligned}
\beta_{\mathsf{a},r} &= (\beta_1, \beta_2)', \quad x_t^{\mathsf{a},r} = (y_{t-1}, \Delta y_{t-1})', \\
\beta_{\mathsf{b},r} &= (\beta_1, \beta_2, \beta_3)', \quad x_t^{\mathsf{b},r} = (1, y_{t-1}, \Delta y_{t-1})', \\
\beta_{\mathsf{c},r} &= (\beta_1, \beta_2, \beta_3, \beta_4)', \quad x_t^{\mathsf{c},r} = (1, t^*, y_{t-1}, \Delta y_{t-1})',
\end{aligned}
$$

Next, the alternative hypothesis of non-linearity, which does not specify whether $y_t$ is non-linear I(0) or I(1), is simply not the null hypothesis, and can be written as

$$
\mathsf{H}_1^{\mathsf{a}} : \text{at least one of } \beta_3, ..., \beta_8 \neq 0, \tag{11}
$$

$$
\mathsf{H}_1^{\mathsf{b}} : \text{at least one of } \beta_4, ..., \beta_{10} \neq 0, \tag{12}
$$

$$
\mathsf{H}_1^{\mathsf{c}} : \text{at least one of } \beta_5, ..., \beta_{13} \neq 0. \tag{13}
$$

### 3.2 Robust Linearity Tests

The above null hypotheses of linearity in $\mathsf{H}_0^{\mathsf{a}}$, $\mathsf{H}_0^{\mathsf{b}}$, and $\mathsf{H}_0^{\mathsf{c}}$ are tested by the Wald statistic

$$W_T^m \equiv T(RSS_r^m - RSS_u^m)/RSS_u^m, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}, \tag{14}$$

where $RSS_u^m$ and $RSS_r^m$ denote the residual sum of squares from the unrestricted OLS regression $y_t$ on $x_t^m$ in (9) and the restricted OLS regression of $y_t$ on $x_t^{m,r}$ in (10), respectively. Before the large sample properties of the $W_T^m$ statistic are discussed, the following conditions are imposed on the error term $\epsilon_t$.

**Assumption 1** *Let $\{\epsilon_t\}$ be a sequence of independent and identically distributed (i.i.d.) random variables defined on the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathsf{E}\epsilon_t = 0$ and $\mathsf{E}\epsilon_t^2 = \sigma_\epsilon^2$ hold. In addition, assume that $\mathsf{E}\left|\epsilon_t\right|^{8+\delta} < \infty$ for some $\delta > 0$.*

In Assumption 1, the condition $\mathsf{E}\left|\epsilon_t\right|^{8+\delta} < \infty$ is needed in the context of deriving the limiting distribution of $W_T^m$ under a linear I(1) model. In fact, if we only were interested in the asymptotic distribution of $W_T^m$ under a linear I(0) model weaker moment conditions do apply (see e.g. the conditions in Lundbergh, Teräsvirta, and van Dijk, 2003, p. 106).

**Theorem 1** *Consider the regression equation (9) when Assumption 1 holds.*

*(i) Under $\mathsf{H}_0^m$, if $y_t$ is linear I(0), then $W_T^m \Rightarrow W_0^m$, where $W_0^{\mathsf{a}} = \chi^2(6)$, $W_0^{\mathsf{b}} = \chi^2(7)$, and $W_0^{\mathsf{c}} = \chi^2(9)$.*

*(ii) Under $\mathsf{H}_0^m$, if $y_t$ is linear I(1), then $W_T^m \Rightarrow W_1^m$, where $W_1^m = B^m + \chi^2(3)$ and $B^m$ is a matrix function of $B(s)$ which is given in the Appendix. Furthermore, the limiting distribution $W_1^m$ is nuisance parameter free.*

*(iii) Under $\mathsf{H}_1^m$, $W_T^m$ diverges to $+\infty$ at the rate $O_p(T)$ whether $y_t$ is non-linear I(0) or non-linear I(1).*

**Proof.** See the Appendix. ■

As to be expected, the distribution of $W_T^m$ is not invariant to the order of integration under the null hypothesis of linearity. It is interestingly noticed, though, that the test is consistent against both non-linear I(0) and non-linear I(1) specifications. Furthermore, before a remedy the order of integration problem is presented we shall introduce a heteroscedasticity robust version of the Wald statistic in (14). In fact, it is has been shown in the literature that the size properties of the Taylor-series based linearity type of tests are very sensitive to (G)ARCH effects. If such effects are ignored, spurious rejection of the null hypothesis of linearity can occur as often as 70% of the times at a 5% nominal significance level, see e.g. Pavlidis, Paya, and Peel (2009). The robust version of the Wald statistic utilized in this work is the one by White (see White, 1980) and is for our testing situation given by

$$W_{R,T}^{m} \equiv T\left(R_m\hat{\beta}_m\right)'\left[R_m\hat{V}_mR_m'\right]^{-1}\left(R_m\hat{\beta}_m\right), \quad m = \mathsf{a},\mathsf{b},\mathsf{c}, \tag{15}$$

where $\hat{\beta}_m'$ is the OLS estimator of $\beta_m'$ in (9), $R_{\mathsf{a}} = [0_{6\times2} : I_6]$, $R_{\mathsf{b}} = [0_{7\times3} : I_7]$, $R_{\mathsf{c}} = [0_{9\times4} : I_9]$, and the estimated covariance matrix is given by

$$\hat{V}_m = S_m^{-1}\hat{S}_mS_m^{-1},$$

with $S_m = T^{-1}\sum_{t=1}^{T}x_t^mx_t^{m\prime}$, $\hat{S}_m = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2x_t^mx_t^{m\prime}$, and $\hat{\epsilon}_t$ signifies the sample residual from the regression in (9). The large sample properties of the $W_{R,T}^{m}$ is given in the following theorem.

**Theorem 2** *Consider the regression equation (9) when Assumption 1 holds.*

*(i) Under $\mathsf{H}_0^m$, if $y_t$ is linear I(0) or linear I(1), then $W_{R,T}^{m} - W_T^m = o_p(1)$.*

*(ii) Under $\mathsf{H}_1^m$, $W_{R,T}^{m}$ diverges to $+\infty$ at the rate $O_p(T)$ whether $y_t$ is non-linear I(0) or non-linear I(1).*

**Proof.** See the Appendix. ■

It should be noticed that under the present assumptions it also follows from part (i) of Theorem 2 that $W_{R,T}^{m} \Rightarrow W_0^m$ if $y_t$ is linear I(0) and $W_{R,T}^{m} \Rightarrow W_1^m$ if $y_t$ is linear I(1), i.e. $W_{R,T}^{m}$ and $W_T^m$ have in this case the same asymptotic distribution. However, even though these tests are asymptotically equivalent it is shown in our Monte Carlo study that $W_{R,T}^{m}$ compares favorable to $W_T^m$ in terms of smaller size distortions in finite samples and the presence of heteroscedastic errors.

### 3.3 Robust and Invariant Linearity Tests

To propose test statistics whose critical values are the same irrespective of a linear I(0) or I(1) processes are considered, we will closely follow the approach suggested by H&L and is outlined below.

Consider first a modified non-robust Wald type of test statistic

$$W_T^{*m} = \exp\{-b_mH_T^m\}W_T^m, \quad m = \mathsf{a},\mathsf{b},\mathsf{c},$$

where $b_m$ is a non-zero constant, $H_T^m$ is a test statistic with pivotal limiting distribution $H^m$ such that $H_T^m \Rightarrow H^m$ when $y_t$ is linear I(1) and $H_T^m \xrightarrow{p} 0$ when $y_t$ is linear I(0). It follows now that under a linear I(0) hypothesis the statistic $W_T^{*m}$ has the same limiting distribution as $W_T^m$ in Theorem 1(i) because $\exp\{-b_mH_T^m\} \xrightarrow{p} 1$. On the other hand, if a linear I(1) model is

**Table 1:** Asymptotic $b$-values

| Significance level | Case $m = \mathsf{a}$ (raw) | Case $m = \mathsf{b}$ (constant) | Case $m = \mathsf{c}$ (linear trend) |
|---:|:---:|:---:|:---:|
| 1% | 0.250 | 0.405 | 0.825 |
| 5% | 0.255 | 0.425 | 0.855 |
| 10% | 0.275 | 0.455 | 0.905 |

**Notes:** Results are based on $T$=100,000 and 100,000 replications.

considered, then $W_T^{*m} \Rightarrow \exp\{-b_m H^m\} W_1^m$ where $W_1^m$ is given in Theorem 1(ii). Taken these results together it becomes evident that we should find a (asymptotic) $b_m$-value such that

$$\Pr\left(W_0^m > c_\alpha\right) = \Pr\left(\exp\{-b_m H^m\} W_1^m\right) > c_\alpha) = \alpha,$$

where $c_\alpha$ denotes the (asymptotic) critical value and $\alpha$ is the significance level of the test. Put differently, if we can find such a $b_m$-value, critical values from a standard chi-square distribution can be used, irrespective of the order of integration, when the null hypothesis of linearity is tested. It should be noticed that this $b_m$-value depends on the desirable significance level $\alpha$ (cf. Table 1).

The $H_T^m$ statistic we choose is the same one as in H&L, i.e. we will use the Dickey-Fuller $t$-statistic, denoted $DF_T^m$, for testing the unit root hypothesis $\pi_1 = 1$ in the regression

$$y_t = \vartheta' d_m + \pi_1 y_{t-1} + \omega_0 \Delta y_{t-1} + \epsilon_t, \;\; m = \mathsf{a}, \mathsf{b}, \mathsf{c}, \tag{16}$$

against the one-sided alternative $\pi_1 < 1$, where $d_{\mathsf{a}} = 0$, $d_{\mathsf{b}} = 1$, and $d_{\mathsf{c}} = (1, t)'$, and $\vartheta$ is a conformable parameter vector. The term $\Delta y_{t-1}$ is included because we allow $\Delta y_t$ to be stationary linear AR(1) process under the null hypothesis. One can now show that $DF_T^m \Rightarrow DF^m$ and $|DF_T^m|^{-1} \Rightarrow |DF^m|^{-1}$ when $y_t$ is linear I(1).[8] It is also a straightforward exercise to show that if $y_t$ is linear I(0), then $DF_T^m \xrightarrow{p} -\infty$ and $|DF_T^m|^{-1} \xrightarrow{p} 0$. Taken these results together, the invariant non-robust linearity test statistics employed in this work is defined by

$$W_T^{*m} \equiv \exp\{-b_m \, |DF_T^m|^{-1}\} W_T^m, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}.$$

To operationalize this test statistic asymptotic $b_m$-values are needed. These values are found by simulations and are for conventional significance levels reported in the Table 1.

The large sample properties under the null and the alternative hypothesis of the $W_T^{*m}$ statistic are summarized in the corollary below.

**Corollary 3** *Consider the regression equation (9) when Assumption 1 holds.*

*(i)* $W_T^{*m} \Rightarrow W_0^m$ *if* $y_t$ *is linear I(0).* $W_T^{*m} \Rightarrow \exp\{-b_m \, |DF^m|^{-1}\} W_1^m$ *if* $y_t$ *is linear I(1).*

*(ii) Under* $\mathsf{H}_1^m$, $W_T^{*m}$ *diverges to* $+\infty$ *at the rate* $O_p(T)$ *whether* $y_t$ *is non-linear I(0) or non-linear I(1).*

[8] Explicit results for the $DF^m$ distribution can in this case be found in e.g. Hamilton (1994 p. 494).

The result (i) is an immediate consequence of the properties of the $W_T^m$ statistic in part (i) and (ii) of Theorem 1. Next, part (ii) follows noticing that if $y_t$ is non-linear I(0) then the $DF_T^m$ statistic is consistent (diverges to $-\infty$) and thus $W_T^{*m} = (1+o_p(1))W_T^m$. If instead $y_t$ is non-linear I(1) then the $DF_T^m$ statistic is $O_p(1)$ (and this $O_p(1)$ term is positive) so $W_T^{*m} = O_p(1)W_T^m$. Combining these results with the result (iii) in Theorem 1 now establishes the results in the corollary.

The robust invariant linearity test in this work is defined by

$$W_{R,T}^{*m} \equiv \exp\{-b_m \, |DF_T^m|^{-1}\} W_{R,T}^m, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}.$$

Due to the results in Theorem 2, the $W_{R,T}^{*m}$ statistic has the same large sample properties as the $W_T^{*m}$ statistic in Corollary 3.

### 3.4 Serially Correlated Errors

Accommodating serially correlated errors for the linearity tests above is most easily accomplished by augmenting (8) with lagged changes in the $\{z_t\}$ sequence, that is

$$\begin{aligned} z_t &= \delta_0 z_{t-1} + \delta_1 t^* z_{t-1} + \delta_2 z_{t-1}^2 + \delta_3 t^* z_{t-1}^2 \\ &\quad + \sum_{j=1}^{p} \lambda_{0,j} \Delta z_{t-1} + \lambda_1 t^* \Delta z_{t-1} + \lambda_2 \left(\Delta z_{t-1}\right)^2 + \lambda_3 t^* \left(\Delta z_{t-1}\right)^2 + \epsilon_t. \end{aligned} \tag{17}$$

where $\epsilon_t$ is a white noise process. To ensure that there are no more than a single unit root, all the values of $r$ satisfying the inverse characteristic equation: $1 - \lambda_{0,1} r - \lambda_{0,2} r^2 + \cdots + \lambda_{0,p} r^p = 0$ must lie outside the unit circle. The unrestricted and restricted hybrid regression equation in (9) is then replaced with

$$y_t = \beta_m' x_t^m + \sum_{j=2}^{p} \xi_j \Delta y_{t-j} + \epsilon_t, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}, \tag{18}$$

and

$$y_t = \beta_{m,r}' x_t^{m,r} + \sum_{j=2}^{p} \xi_j \Delta y_{t-j} + \epsilon_t, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}, \tag{19}$$

respectively. The non-robust version of our linearity test accommodating serially correlated errors, signified as $W_{A,T}^{*m}$, is now defined as

$$W_{A,T}^{*m} \equiv \exp\{-b_m \, |ADF_T^m|^{-1}\} W_{A,T}^m, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c},$$

where $ADF_T^m$ is the traditional augmented DF $t$-test for testing the unit root hypothesis $\pi_1 = 1$ in

$$y_t = \vartheta' d_m + \pi_1 y_{t-1} + \sum_{j=1}^{p} \omega_j \Delta y_{t-j} + \epsilon_t, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c}. \tag{20}$$

against the one-sided alternative $\pi_1 < 1$. Moreover, $W^m_{A,T}$ is defined as $W^m_T$ in (14) but the unrestricted and restricted residual sum of squares are instead obtained from (18) and (19), respectively. It is noticed that the (asymptotic) $b_m$-value is the same as for the case without serially correlated errors and is due to the fact that $ADF^m_T - DF^m_T = o_p(1)$ and $W^m_{A,T} - W^m_T = o_p(1)$. Towards this end, the lag-length $p$ in (19) and (20) is an unknown parameter to be determined. It is noticed that whether we have a linear I(0) or linear I(1) model under the null hypothesis of linearity, $p$ is estimated consistently (see Paulsen, 1984) by the Schwarz (see Schwarz, 1978) information criteria (SIC).

We shall as a final test consider a robust version of the linearity test accommodating serial correlation. This test is abbreviated $W^{*m}_{A,R,T}$ and is defined by

$$W^{*m}_{A,R,T} \equiv \exp\{-b_m \, |ADF^m_T|^{-1}\} W^m_{A,R,T}, \quad m = \mathsf{a}, \mathsf{b}, \mathsf{c},$$

where $W^m_{A,R,T}$ is defined as $W^m_{R,T}$ in (15) but $x^m_t$ and $\hat{\epsilon}_t$ are replaced with the covariates and the sample residuals from (19). One can also show that the $W^m_{A,R,T}$ statistic have the same asymptotic distributions as the $W^m_{A,T}$ statistic, and the (asymptotic) $b_m$-values in Table 1 may therefore once again be used.

# 4 Monte Carlo Study

This section evaluates the small sample performance of the $W^{*m}_T$ statistic by means of a Monte Carlo study. The newly proposed test is compared to the extant one proposed by H&L, hereafter signified as $W_{HL}$. We focus solely on the case $m = \mathsf{b}$ as the H&L test is constructed for this empirically relevant case.[9] Size and power experiments are conducted. Regarding the former one, we allow for three different distributions: Normal, $\chi^2 - 1$, and $t(3)$ as well as for conditional heteroscedasticity via a simple GARCH(1,1) process. These specifications permit skewness, fat tails and volatility clustering which is commonly observed in many economic and financial time-series. In addition to the $W^*_T$ statistic, we also consider the $W^*_{R,T}$ statistic which is expected to be less sensitive to GARCH effects. When the power is evaluated, we focus on the case of normality and homoscedastic errors for simplicity. Moreover, we set the lag length $p$ equal to one. The considered sample sizes are 150 and 300. The number of replications is set equal to 5,000 and the nominal significance level is 5%.

## 4.1 Size Experiments

The data generating process (DGP) is a linear second-order autoregressive model

$$y_t = \rho y_{t-1} + \phi \Delta y_{t-1} + \epsilon_t,$$

[9] Simulation results for the cases of raw ($m = \mathsf{a}$) and trending data ($m = \mathsf{c}$) are available upon request from the authors. For these cases, a direct comparison with the H&L test is not possible.

**Table 2:** Size Experiments with different Distributions

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Normally distributed errors | | | | | | | | |
| | $T = 150$ | | | | $T = 300$ | | | |
| $\phi/\rho$ | 0.0 | 0.8 | 0.9 | 1.0 | 0.0 | 0.8 | 0.9 | 1.0 |
| -0.5 | 0.071 | 0.041 | 0.044 | 0.076 | 0.060 | 0.035 | 0.033 | 0.063 |
| | 0.128 | 0.043 | 0.045 | 0.050 | 0.123 | 0.042 | 0.043 | 0.049 |
| 0.0 | 0.056 | 0.039 | 0.041 | 0.081 | 0.049 | 0.034 | 0.032 | 0.067 |
| | 0.052 | 0.049 | 0.054 | 0.058 | 0.052 | 0.047 | 0.049 | 0.054 |
| 0.5 | 0.057 | 0.044 | 0.041 | 0.093 | 0.051 | 0.038 | 0.034 | 0.071 |
| | 0.055 | 0.053 | 0.056 | 0.061 | 0.051 | 0.049 | 0.050 | 0.056 |
| $\chi^2(1) - 1$ distributed errors | | | | | | | | |
| | $T = 150$ | | | | $T = 300$ | | | |
| $\phi/\rho$ | 0.0 | 0.8 | 0.9 | 1.0 | 0.0 | 0.8 | 0.9 | 1.0 |
| -0.5 | 0.086 | 0.050 | 0.053 | 0.096 | 0.071 | 0.042 | 0.042 | 0.077 |
| | 0.137 | 0.051 | 0.053 | 0.087 | 0.131 | 0.046 | 0.049 | 0.073 |
| 0.0 | 0.057 | 0.042 | 0.042 | 0.097 | 0.055 | 0.038 | 0.035 | 0.079 |
| | 0.064 | 0.051 | 0.055 | 0.093 | 0.061 | 0.047 | 0.050 | 0.078 |
| 0.5 | 0.051 | 0.033 | 0.031 | 0.107 | 0.050 | 0.030 | 0.026 | 0.075 |
| | 0.061 | 0.044 | 0.047 | 0.089 | 0.058 | 0.041 | 0.040 | 0.068 |
| Student-$t(3)$ distributed errors | | | | | | | | |
| | $T = 150$ | | | | $T = 300$ | | | |
| $\phi/\rho$ | 0.0 | 0.8 | 0.9 | 1.0 | 0.0 | 0.8 | 0.9 | 1.0 |
| -0.5 | 0.089 | 0.042 | 0.055 | 0.079 | 0.075 | 0.035 | 0.043 | 0.073 |
| | 0.151 | 0.043 | 0.051 | 0.076 | 0.134 | 0.038 | 0.052 | 0.071 |
| 0.0 | 0.057 | 0.047 | 0.047 | 0.099 | 0.055 | 0.043 | 0.039 | 0.078 |
| | 0.062 | 0.058 | 0.058 | 0.087 | 0.064 | 0.054 | 0.056 | 0.078 |
| 0.5 | 0.056 | 0.046 | 0.043 | 0.114 | 0.048 | 0.039 | 0.034 | 0.080 |
| | 0.061 | 0.060 | 0.060 | 0.098 | 0.057 | 0.055 | 0.051 | 0.088 |

**Notes:** The DGP is given by $y_t = \rho y_{t-1} + \phi \Delta y_{t-1} + \epsilon_t$. Reported numbers are simulated rejection frequencies of the $W_T^*$ test (upper entries) and the $W_{HL}$ test (lower entries).

where the autoregressive parameters $\rho$ and $\phi$ take the following values: $\rho = \{0.0, 0.8, 0.9, 1.0\}$ and $\phi = \{-0,5, 0.0, 0.5\}$. This means that we consider I(0) and I(1) models under the null hypothesis of linearity. The error term $\epsilon_t$ is either standard normally distributed, follows a skewed $\chi^2 - 1$ distribution or a fat-tailed Student-$t$ distribution with three degrees of freedom, i.e. $t(3)$. Moreover, we allow for a GARCH(1,1) process

$$\begin{aligned} \epsilon_t &= \eta_t \sqrt{h_t}, \\ h_t &= \omega_0 + \omega_1 \epsilon_{t-1}^2 + \omega_2 h_{t-1}, \end{aligned}$$

where $\eta_t$ is either standard normally, $\chi^2 - 1$ or $t(3)$-distributed. The parameters are chosen to resemble the typically observed behaviour of volatility clustering: $(\omega_0, \omega_1, \omega_2) = (0.1, 0.3, 0.6)$.

Results are reported in Tables 2 and 3 for homoscedastic and heteroscedastic errors, respectively. As an exception to other experiments, a third sample size of $T = 500$ is included here

**Table 3:** Size Experiments with GARCH Effects

| Performance of $W_T^*$, $W_{R,T}^*$ and $W_{HL}$ under heteroscedastic errors | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T = 150$ | | | | $T = 300$ | | | | $T = 500$ | | | |
| $\phi/\rho$ | 0.0 | 0.8 | 0.9 | 1.0 | 0.0 | 0.8 | 0.9 | 1.0 | 0.0 | 0.8 | 0.9 | 1.0 |
| -0.5 | 0.203 | 0.239 | 0.242 | 0.253 | 0.303 | 0.316 | 0.301 | 0.287 | 0.412 | 0.396 | 0.365 | 0.333 |
| | 0.203 | 0.252 | 0.242 | 0.247 | 0.198 | 0.207 | 0.194 | 0.194 | 0.203 | 0.175 | 0.157 | 0.161 |
| | 0.278 | 0.339 | 0.342 | 0.287 | 0.385 | 0.445 | 0.436 | 0.362 | 0.505 | 0.533 | 0.521 | 0.409 |
| 0.0 | 0.297 | 0.229 | 0.215 | 0.248 | 0.400 | 0.303 | 0.280 | 0.278 | 0.500 | 0.393 | 0.358 | 0.309 |
| | 0.315 | 0.240 | 0.213 | 0.240 | 0.266 | 0.185 | 0.161 | 0.177 | 0.231 | 0.163 | 0.139 | 0.141 |
| | 0.383 | 0.367 | 0.352 | 0.291 | 0.499 | 0.473 | 0.465 | 0.375 | 0.589 | 0.556 | 0.543 | 0.393 |
| 0.5 | 0.296 | 0.201 | 0.185 | 0.217 | 0.397 | 0.284 | 0.253 | 0.219 | 0.495 | 0.378 | 0.332 | 0.257 |
| | 0.337 | 0.217 | 0.194 | 0.214 | 0.289 | 0.186 | 0.159 | 0.145 | 0.264 | 0.170 | 0.139 | 0.127 |
| | 0.359 | 0.343 | 0.328 | 0.257 | 0.480 | 0.457 | 0.435 | 0.325 | 0.559 | 0.536 | 0.510 | 0.364 |
| Performance of $W_{R,T}^*$ under homoscedastic errors | | | | | | | | | | | | |
| | $T = 150$ | | | | $T = 300$ | | | | $T = 500$ | | | |
| $\phi/\rho$ | 0.0 | 0.8 | 0.9 | 1.0 | 0.0 | 0.8 | 0.9 | 1.0 | 0.0 | 0.8 | 0.9 | 1.0 |
| -0.5 | 0.135 | 0.130 | 0.113 | 0.149 | 0.096 | 0.082 | 0.073 | 0.101 | 0.074 | 0.070 | 0.054 | 0.081 |
| 0.0 | 0.189 | 0.118 | 0.114 | 0.158 | 0.120 | 0.089 | 0.077 | 0.106 | 0.096 | 0.070 | 0.062 | 0.086 |
| 0.5 | 0.181 | 0.129 | 0.107 | 0.164 | 0.122 | 0.095 | 0.078 | 0.103 | 0.107 | 0.077 | 0.065 | 0.093 |

**Notes:** The DGP is given by $y_t = \rho y_{t-1} + \phi \Delta y_{t-1} + \epsilon_t$, where $\epsilon_t = \eta_t \sqrt{h_t}, \eta_t \overset{iid}{\sim} N(0,1)$ with $h_t = 0.1 + 0.3\epsilon_{t-1}^2 + 0.6h_{t-1}$. Reported numbers are simulated rejection frequencies of the $W_T^*$ test (upper entries), the heteroscedasticity-robust $W_{R,T}^*$ test (middle entries) and the $W_{HL}$ test (lower entries), see upper panel. The lower panel reports only the performance of the heteroscedasticity-robust $W_{R,T}^*$ test. For the performance of the other two tests, see Table 2.

as well for the purpose of illustration. The results in Table 2 suggest the $W_T^{*m}$ test is generally correctly sized with some minor discrepancies. Positive deviations mainly occur in the case of an I(1) DGP. It can also be seen that the H&L test performs in general as good as the newly proposed one. Thus, it appears that both our test and the $W_{HL}$ test are robust to skewed and fat-tailed distributions. The results for heteroscedastic errors (lower panel of Table 3) reveal the following:

First, all three test statistics ($W_T^*$, $W_{R,T}^*$, and $W_{HL}$) are over-sized. Even though the $W_{R,T}^*$ statistic performs similar to the $W_T^*$ statistic for $T = 150$, the size-distortions are substantially mitigated for the $W_{R,T}^*$ statistic when the sample size is increased. It can also be seen that the $W_{HL}$ test is for all cases considered even more sensitive to GARCH effects than the $W_T^*$ test. Second, when no GARCH effects are present, the heteroscedasticity-robust test statistic is over-sized, but the magnitude of distortions decline rapidly with an increasing sample size. Finally, it appears that the heteroscedasticity-robust test is suitable for sample sizes larger than or equal to $T = 300$, while it should be used with caution in smaller sample sizes.

**Table 4:** Power Experiments for I(0) Data

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| STAR Model, $\phi_1 = \phi_3$, $\phi_2 = \phi_4$ | | | | | | | | | |
| | | | | $T = 150$ | | | $T = 300$ | | |
| $\phi_1$ | $\phi_2$ | | | $W_T^*$ | $W_T^0$ | $W_{HL}$ | $W_T^*$ | $W_T^0$ | $W_{HL}$ |
| 0.5 | 0.7 | | | 0.070 | 0.085 | 0.087 | 0.096 | 0.128 | 0.134 |
| 0.3 | 0.7 | | | 0.127 | 0.182 | 0.209 | 0.249 | 0.346 | 0.394 |
| 0.5 | 0.9 | | | 0.111 | 0.177 | 0.264 | 0.236 | 0.362 | 0.551 |
| 0.3 | 0.9 | | | 0.189 | 0.298 | 0.467 | 0.417 | 0.594 | 0.797 |
| TV-AR Model, $\phi_1 = \phi_2$, $\phi_3 = \phi_4$ | | | | | | | | | |
| | | | | $T = 150$ | | | $T = 300$ | | |
| $\phi_1$ | | $\phi_3$ | | $W_T^*$ | $W_T^0$ | $W_{HL}$ | $W_T^*$ | $W_T^0$ | $W_{HL}$ |
| 0.5 | | 0.7 | | 0.098 | 0.127 | 0.058 | 0.167 | 0.244 | 0.062 |
| 0.3 | | 0.7 | | 0.284 | 0.405 | 0.089 | 0.620 | 0.755 | 0.108 |
| 0.5 | | 0.9 | | 0.384 | 0.559 | 0.121 | 0.812 | 0.921 | 0.172 |
| 0.3 | | 0.9 | | 0.713 | 0.866 | 0.193 | 0.986 | 0.997 | 0.333 |
| TV-STAR Model, $\phi_4 - \phi_3 = \phi_2 - \phi_1$ | | | | | | | | | |
| | | | | $T = 150$ | | | $T = 300$ | | |
| $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ | $W_T^*$ | $W_T^0$ | $W_{HL}$ | $W_T^*$ | $W_T^0$ | $W_{HL}$ |
| 0.5 | 0.7 | 0.7 | 0.9 | 0.170 | 0.259 | 0.150 | 0.398 | 0.555 | 0.291 |
| 0.1 | 0.5 | 0.5 | 0.9 | 0.571 | 0.711 | 0.432 | 0.933 | 0.968 | 0.773 |
| -0.3 | 0.3 | 0.3 | 0.9 | 0.883 | 0.942 | 0.714 | 0.998 | 1.000 | 0.968 |

## 4.2 Power Experiments

As a next step the empirical power of the tests are analyzed and evaluated. To this end, three different DGPs are considered. They are similar to the cases (b), (c) and (f) in Lundbergh, Teräsvirta, and van Dijk (2003) which are a STAR, a TV-AR and a TV-STAR process, respectively. The STAR process is given by

$$\begin{aligned} y_t &= \phi_1 y_{t-1} G_1(y_{t-1};\gamma_1) + \phi_2 y_{t-1}\{1 - G_1(y_{t-1};\gamma_1)\} + \epsilon_t, \\ \Delta y_t &= \psi_1 \Delta y_{t-1} G_1(\Delta y_{t-1};\gamma_1) + \psi_2 \Delta y_{t-1}\{1 - G_1(\Delta y_{t-1};\gamma_1)\} + \epsilon_t, \end{aligned}$$

for the I(0) and I(1) case, respectively. The second DGP is a time-varying AR model

$$\begin{aligned} y_t &= \phi_1 y_{t-1}[1 - G_2(t^*;\gamma_2, c_2)] + \phi_3 y_{t-1} G_2(t^*;\gamma_2, c_2) + \epsilon_t, \\ \Delta y_t &= \psi_1 \Delta y_{t-1}[1 - G_2(t^*;\gamma_2, c_2)] + \psi_3 \Delta y_{t-1} G_2(t^*;\gamma_2, c_2) + \epsilon_t. \end{aligned}$$

The third DGP is a TV-STAR process where the autoregressive parameters are restricted in the following way: $\phi_4 - \phi_3 = \phi_2 - \phi_1$ for the I(0) case and $\psi_4 - \psi_3 = \psi_2 - \psi_1$ for the I(1) case, respectively.

The exact parameter constellations for autoregressive parameters $\phi_1$, $\phi_2$, $\phi_3$, $\phi_4$ and $\psi_1$, $\psi_2$, $\psi_3$, $\psi_4$ are given in Tables 4 and 5.[10] Following Lundbergh, Teräsvirta, and van Dijk (2003),

[10]The results for the opposite direction of autoregressive parameters, i.e. decreasing instead of increasing from

**Table 5:** Power Experiments for I(1) Data

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| STAR Model, $\psi_1 = \psi_3$, $\psi_2 = \psi_4$ | | | | | | | | | |
| | | | | $T = 150$ | | | $T = 300$ | | |
| $\psi_1$ | $\psi_2$ | | | $W_T^*$ | $W_T^1$ | $W_{HL}$ | $W_T^*$ | $W_T^1$ | $W_{HL}$ |
| 0.5 | 0.7 | | | 0.152 | 0.082 | 0.099 | 0.145 | 0.121 | 0.117 |
| 0.3 | 0.7 | | | 0.220 | 0.185 | 0.164 | 0.282 | 0.353 | 0.252 |
| 0.5 | 0.9 | | | 0.291 | 0.181 | 0.219 | 0.328 | 0.358 | 0.357 |
| 0.3 | 0.9 | | | 0.373 | 0.284 | 0.328 | 0.422 | 0.592 | 0.573 |
| TV-AR Model, $\psi_1 = \psi_2$, $\psi_3 = \psi_4$ | | | | | | | | | |
| | | | | $T = 150$ | | | $T = 300$ | | |
| $\psi_1$ | | $\psi_3$ | | $W_T^*$ | $W_T^1$ | $W_{HL}$ | $W_T^*$ | $W_T^1$ | $W_{HL}$ |
| 0.5 | | 0.7 | | 0.154 | 0.127 | 0.077 | 0.169 | 0.243 | 0.066 |
| 0.3 | | 0.7 | | 0.291 | 0.402 | 0.102 | 0.481 | 0.763 | 0.108 |
| 0.5 | | 0.9 | | 0.410 | 0.565 | 0.159 | 0.654 | 0.919 | 0.216 |
| 0.3 | | 0.9 | | 0.626 | 0.862 | 0.242 | 0.833 | 0.998 | 0.391 |
| TV-STAR Model, $\psi_4 - \psi_3 = \psi_2 - \psi_1$ | | | | | | | | | |
| | | | | $T = 150$ | | | $T = 300$ | | |
| $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $W_T^*$ | $W_T^1$ | $W_{HL}$ | $W_T^*$ | $W_T^1$ | $W_{HL}$ |
| 0.5 | 0.7 | 0.7 | 0.9 | 0.230 | 0.248 | 0.141 | 0.351 | 0.536 | 0.243 |
| 0.1 | 0.5 | 0.5 | 0.9 | 0.507 | 0.697 | 0.361 | 0.782 | 0.968 | 0.677 |
| -0.3 | 0.3 | 0.3 | 0.9 | 0.771 | 0.949 | 0.651 | 0.940 | 1.000 | 0.926 |

$\gamma_1 = 5$ and $\gamma_2 = 25$. The break point $c_2$ is specified as 0.5. Innovations $\epsilon_t$ are drawn from the standard normal distribution.

These three DGPs allows us to study the empirical power properties in empirically relevant settings. The first DGP is a pure STAR model without structural change, and it is here expected that the H&L test performs somewhat better than the newly proposed one. The reason for this expectation is the fact that the H&L test is designed to detect non-linearity of this certain type. The second DGP is a time-varying AR model with a smooth structural change in the AR parameter. As such, the H&L test is not designed to direct power against this DGP whereas, to some extents, our test is. Regarding the third DGP, our test is expected to perform better than the H&L test, although the latter one may have satisfactory power if the structural changes are less pronounced.

Similarly to H&L, the following $W_T^0$ and $W_T^1$ versions of linearity tests are considered as benchmark tests. The former one assumes an I(0) model, while the latter one assumes an I(1) model. In particular, the $W_T^0$ is carried out by running the following OLS regression in levels

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 t^* + \beta_4 t^* y_{t-1} + \beta_5 y_{t-1}^2 + \beta_6 t^* y_{t-1}^2 + \epsilon_t.$$

The null hypothesis of linearity is then given by $H_0 : \beta_3 = \ldots = \beta_6 = 0$. The corresponding Wald statistic is asymptotically distributed as a $\chi^2$ random variable with four degrees of freedom.

($\phi_1$ over to $\phi_4$ and $\psi_1$ over to $\psi_4$, are not reported to save space. They are more or less symmetric and are available from authors upon request.

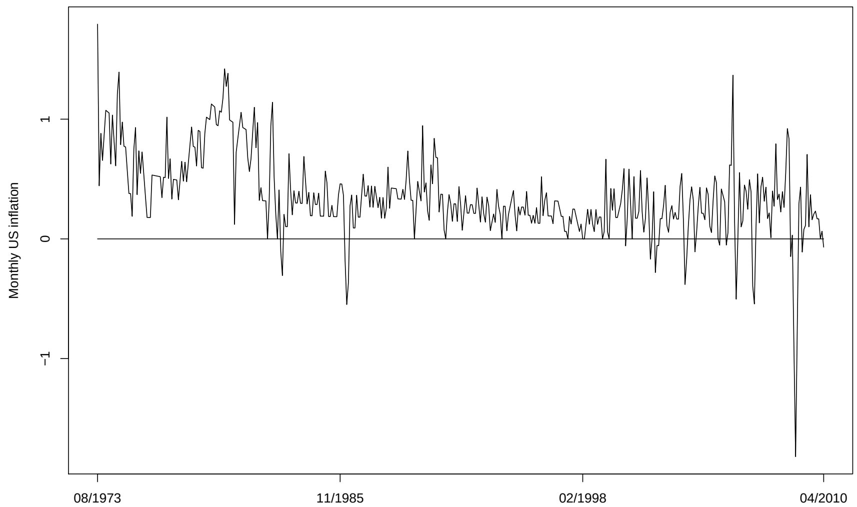


Figure 1: Monthly US inflation, August 1973 – April 2010.

Analogously, the $W_T^1$ statistic is based on an OLS regression in first differences

$$\Delta y_t = \beta_1 + \beta_2 \Delta y_{t-1} + \beta_3 t^* + \beta_4 t^* \Delta y_{t-1} + \beta_5 (\Delta y_{t-1})^2 + \beta_6 t^* (\Delta y_{t-1})^2 + \epsilon_t.$$

Again, the null hypothesis of linearity is given by $H_0 : \beta_3 = \ldots = \beta_6 = 0$ and the corresponding Wald statistic is asymptotically distributed as a $\chi^2$ random variable with four degrees of freedom.

Results for the I(0) and I(1) cases are reported in Table 4 and 5, respectively. The upper, middle and lower entry in each cell reports the empirical power of the $W_T^*$ test, the $W_T^0$ and the $W_{HL}$ test, respectively. The results in Table 4 for the STAR model suggest that the $W_{HL}$ test indeed dominates the $W_T^*$ and the $W_T^0$ test. The difference in power increases with the sample size and the degree of non-linearity measured by the distance of AR parameters across regimes. Nonetheless, the $W_T^*$ test exhibits non-trivial power to detect STAR dynamics. The results in the case of a TV-AR model in Table 5 show that the $W_T^*$ test outperforms the $W_{HL}$ test, and it appears that the extant $W_{HL}$ test is less useful in detecting a smooth structural change in the autoregressive parameters. Moreover, the $W_T^*$ test performs often as good as the $W_T^0$ benchmark test which underlines its usefulness. For the TV-STAR model, the expected dominance of the $W_T^*$ test over the $W_{HL}$ test is confirmed. The $W_T^*$ test exhibits satisfactory power.

The results in Table 5 for I(1) DGPs confirm the previous conclusions in general. Some remarks on the minor differences to the results for I(0) DGPs are in order. In general, the $W_T^*$ is more powerful when applied to I(0) time-series than to I(1). This result is in line with H&L. It may also be noticed that for the pure STAR model the $W_T^*$ test performs in fact better even better than the $W_T^1$ test for $T = 150$. This result is can be attributed to the relatively small sample size and the fact that both tests are misspecified.

**Table 6:** Empirical Results for US Inflation

| | $W_T^*$ | $W_{R,T}^*$ | $W_T^0$ | $W_T^1$ | $W_{HL}$ |
|---|---|---|---|---|---|
| Test statistic | 88.203 | 53.540 | 55.118 | 4.607 | 8.021 |
| Critical value (5%) | 14.067 | 14.067 | 9.488 | 9.488 | 9.488 |

**Notes:** The lag length $p$ is chosen via SIC.

# 5 Empirical Application

This section contains an empirical application to US inflation data. Monthly inflation based on CPI data obtained from the FREDII database is used. The sample spans from the Post-Bretton Woods period up to the most recent observation, i.e. August 1973–April 2010 with $T = 441$. The data is displayed in Figure 1.

Visual inspection of the time-series suggests that it is presumably characterized by serial correlation, non-linearity, and heteroscedasticity. Moreover, it exhibits a few relatively extreme observations leading to a mild degree of excess kurtosis (4.54). Therefore, it serves as an interesting time-series for the application of $W_T^*$ and $W_{R,T}^*$ statistics. In the following, the case $m = \mathsf{b}$ is considered, as it fits most with the nature of data shown in Figure 1. Table 6 reports the results for the test statistics $W_T^*$, $W_{R,T}^*$, $W_T^0$, $W_T^1$ and $W_{HL}$ together with their critical values at a nominal significance level of five percent.

Evidence for TV-STAR dynamics is found by our $W_T^*$ test, but not by the $W_{HL}$ test. When considering the heteroscedasticity-robust version $W_{R,T}^*$, evidence for non-linearity is less strong, but still significant. This underlines the practical usefulness of the $W_T^*$ test and the importance to account for heteroscedasticity via the $W_{R,T}^*$ test. The test statistics $W_T^0$ and $W_T^1$ assume an I(0) and an I(1) model, respectively. Only the test assuming stationarity rejects the null of linearity.

In the case of a rejection one is left inconclusive about the degree of integration. The same problem is encountered in H&L and they advocate the use of the non-parametric test by Harris, McCabe, and Leybourne (2003). This test is based on sample autocovariances and tests the null hypothesis of stationarity against a unit root. Define $a_{t,k} = \tilde{y}_t \tilde{y}_{t-k}$, where $\tilde{y}_t$ denotes the deviation of $y_t$ from its mean $\bar{y} \equiv T^{-1} \sum_{t=1}^{T} y_t$. The test statistic is then given by

$$S_T = T^{-1/2} \frac{\sum_{t=k+1}^{T} a_{t,k}}{\hat{\omega}(a_{t,k})} \xrightarrow{d} N(0,1)$$

where $\hat{\omega}(a_{t,k})^2$ is the Bartlett kernel-based long run variance estimator of $a_{t,k}$. More specifically,

$$\hat{\omega}(a_{t,k})^2 = \hat{\gamma}_0(a_{t,k}) + 2 \sum_{j=1}^{l} \left(1 - \frac{j}{l}\right) \hat{\gamma}_j(a_{t,k})$$

$$\hat{\gamma}_j(a_{t,k}) = T^{-1} \sum_{t=j+k+1}^{T} a_{t,k} a_{t-j,k}$$

The limiting distribution of $S_T$ is standard normal in the case of globally stationary processes. The test rejects the null hypothesis of stationarity for large values of $S_T$. In our case this test is not directly applicable since our non-linear I(0) model is not globally stationary as discussed in Section 2.1.[11] In further simulation studies (not reported here) it appears, however, that the standard normal distribution yields a fairly good approximation to the limiting distribution of the $S_T$ statistic when a stable TV-STAR process is considered. Thus, we proceed by using the $S_T$ statistic to make inference about the order of integration under the alternative hypothesis.

The values of $k$ and $l$ are $T^{2/3}$ and $12(T/100)^{1/4}$ (rounded to the nearest integer), respectively. The value of the test statistic $S_T$ equals 1.557 and is not significant at the nominal five percent level. The null hypothesis of stationarity is therefore not rejected. Together with the outcomes of the linearity tests, it is concluded that US inflation can be characterized by a non-linear and stationary TV-STAR model.

# 6 Concluding Remarks

In this work we derive an invariant test for the linearity hypothesis against a TV-STAR alternative. Our test is invariant in the sense that critical values from a standard chi-square distribution are applicable irrespectively whether a linear I(0) or I(1) model is considered under the null hypothesis. The true degree of integration has not to be known, pre-specified or pre-tested. Another contribution of this work is to suggest an alternative test which is additionally robust to heteroscedasticity which is often found in economic data. The robustness to heteroscedasticity is achieved by using a White correction for the estimated covariance matrix.

The empirical properties of both tests are evaluated by means of a Monte Carlo study. The results suggest that our tests are correctly sized even if the error distribution exhibits skewness and fat tails. Moreover, the problem of spurious rejections due to neglected heteroscedasticity is mostly remedied by applying the test including the White correction. The power experiments reveal that our test is powerful and competitive with respect to the extant one by Harvey and Leybourne (2007).

In our application to US inflation data from the Post-Bretton Woods period evidence in favour of a non-linear I(0) model is found whereas the test by Harvey and Leybourne (2007) instead lends support to a linear I(0) model. This might be explained by the fact that the Harvey and Leybourne (2007) test is less powerful against time-series subject to non-linearities and structural changes. Another insight of our application is that heteroscedasticity is of importance as the evidence for non-linearity and structural change is reduced, but still significant, when the White correction is applied.

[11] The limiting distribution of the $S_T$ statistic is expected to fairly complicated when the DGP is TV-STAR process, and establishing a complete expression for this distribution is beyond the scope of this work.

# Mathematical Appendix

The proofs in this appendix are only given for the case $\mathsf{a}$. The proofs for the cases $\mathsf{b}$ and $\mathsf{c}$ are similar and therefore omitted. Moreover, the proof of Theorem 1 below closely follows the proof of Theorem 1 in H&L, but it is noticed that our large sample results are different from those derived in H&L for the simple reason that different models are studied. To this end, all summations in this appendix go from 1 to $T$ and the short-hand notion used is $\sum_t$.

**Proof of Theorem 1.** (i) Standard.

Next, the proof of (ii). We notice first that the DGP under the null hypothesis is given by $y_t = y_{t-1} + \epsilon_t$, with starting values $y_{-1}$ and $y_0$ assumed to be known (either fixed or stochastic). Without loss of generality we may set $y_{-1} = y_0 = 0$. Furthermore, it proves convenient to modify some of the covariates in $x_t^{\mathsf{a}}$ in (9) and also re-order them according to levels and first-differences as follows (it is evident that $RSS_u$ is not affected by this manipulation):

$$x_t = (y_{t-1}, t^* y_{t-1}, y_{t-1}^2, t^* y_{t-1}^2, \Delta y_{t-1}, t^* \Delta y_{t-1}, a_{t-1}, b_{t-1})',$$

where $a_{t-1} \equiv (\Delta y_{t-1})^2 - m_2$ and $b_{t-1} \equiv t^* (\Delta y_{t-1})^2 - m_2/2$ with $m_2 = \mathsf{E}(\Delta y_{t-1})^2$.[12] Next, stack $x_t$ into the matrix $X_u$ and $x_t^{\mathsf{a},r}$ in (10) into $X_r$, and by $\hat{\epsilon}_u$ and $\hat{\epsilon}_r$ denote the sample counterparts of $\epsilon = (\epsilon_1, \epsilon_2, ..., \epsilon_T)$ in (9) and (10), respectively. The Wald statistic in (14) can now be written as

$$\begin{aligned} W_T^{\mathsf{a}} &= \frac{\hat{\epsilon}_r'\hat{\epsilon}_r - \hat{\epsilon}_u'\hat{\epsilon}_u}{\hat{\epsilon}_u'\hat{\epsilon}_u/T} \\ &= \frac{\epsilon' X_u (X_u' X_u)^{-1} X_u' \epsilon - \epsilon' X_r (X_r' X_r)^{-1} X_r' \epsilon}{\hat{\epsilon}_u'\hat{\epsilon}_u/T} \\ &= \frac{\left(\epsilon' X_u \gamma_u^{-1}\right) \left[\gamma_u^{-1} \left(X_u' X_u\right) \gamma_u^{-1}\right]^{-1} \left(\gamma_u^{-1} X_u' \epsilon\right)}{\hat{\epsilon}_u'\hat{\epsilon}_u/T} \\ &\quad - \frac{\left(\epsilon' X_r \gamma_r^{-1}\right) \left[\gamma_r^{-1} \left(X_r' X_r\right) \gamma_r^{-1}\right]^{-1} \left(\gamma_r^{-1} X_r' \epsilon\right)}{\hat{\epsilon}_u'\hat{\epsilon}_u/T}, \end{aligned} \tag{A.1}$$

where $\gamma_u = diag\{T, T, T^{3/2}, T^{3/2}, T^{1/2}, T^{1/2}, T^{1/2} T^{1/2}\}$ and $\gamma_r = diag\{T, T^{1/2}\}$ are scaling matrices.

In order to derive the limiting distribution of $W_T^{\mathsf{a}}$ we start with examining the large sample properties of $\gamma_u^{-1} \left(X_u' X_u\right) \gamma_u^{-1}$ and $\gamma_r^{-1} \left(X_r' X_r\right) \gamma_r^{-1}$ in (A.1). Hence, consider first the partition

$$X_u' X_u = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}' & X_{22} \end{bmatrix},$$

[12] Thereby $T^{-1} \sum_t a_{t-1} \xrightarrow{p} 0$ and $T^{-1} \sum_t b_{t-1} \xrightarrow{p} 0$ as long as $\mathsf{E}\epsilon_t^2 < \infty$.

where the sub-matrices are given by

$$X_{11} = \begin{bmatrix} \sum_t y_{t-1}^2 & \sum_t t^* y_{t-1}^2 & \sum_t y_{t-1}^3 & \sum_t t^* y_{t-1}^3 \\ & \sum_t t^{*2} y_{t-1}^2 & \sum_t t^* y_{t-1}^3 & \sum_t t^{*2} y_{t-1}^3 \\ & & \sum_t y_{t-1}^4 & \sum_t t^* y_{t-1}^4 \\ & & & \sum_t t^{*2} y_{t-1}^4 \end{bmatrix},$$

$$X_{12} = \begin{bmatrix} \sum_t \Delta y_{t-1} y_{t-1} & \sum_t t^* \Delta y_{t-1} y_{t-1} & \sum_t y_{t-1} a_{t-1} & \sum_t y_{t-1} b_{t-1} \\ \sum_t t^* \Delta y_{t-1} y_{t-1} & \sum_t t^{*2} \Delta y_{t-1} y_{t-1} & \sum_t t^* y_{t-1} a_{t-1} & \sum_t t^{*2} y_{t-1} b_{t-1} \\ \sum_t \Delta y_{t-1} y_{t-1}^2 & \sum_t t^* \Delta y_{t-1} y_{t-1}^2 & \sum_t y_{t-1}^2 a_{t-1} & \sum_t y_{t-1}^2 b_{t-1} \\ \sum_t t^* \Delta y_{t-1} y_{t-1}^2 & \sum_t t^{*2} \Delta y_{t-1} y_{t-1}^2 & \sum_t t^* y_{t-1}^2 a_{t-1} & \sum_t t^* y_{t-1}^2 b_{t-1} \end{bmatrix},$$

$$X_{22} = \begin{bmatrix} \sum_t (\Delta y_{t-1})^2 & \sum_t t^* (\Delta y_{t-1})^2 & \sum_t \Delta y_{t-1} a_{t-1} & \sum_t \Delta y_{t-1} b_{t-1} \\ & \sum_t t^{*2} (\Delta y_{t-1})^2 & \sum_t t^* \Delta y_{t-1} a_{t-1} & \sum_t t^* \Delta y_{t-1} b_{t-1} \\ & & \sum_t a_{t-1}^2 & \sum_t a_{t-1} b_{t-1} \\ & & & \sum_t b_{t-1}^2 \end{bmatrix},$$

and also write

$$X_r' X_r = \begin{bmatrix} \sum_t y_{t-1}^2 & \sum_t y_{t-1} \Delta y_{t-1} \\ & \sum_t (\Delta y_{t-1})^2 \end{bmatrix}.$$

where $X_{11}$, $X_{22}$, and $X_r' X_r$ are symmetric matrices. Next, the moment conditions in Assumption 1 and the fact that $\{\epsilon_t\}$ is an i.i.d. sequence assert that we can use the results in Hansen (1992, Theorem 4.1 and 4.2), He and Sandberg (2006, Lemma A1), Sandberg (2009, Theorem 1), and a law of large numbers for martingale difference sequences (see e.g. White, 2000 Theorem 3.76) to deduce that

$$\gamma_u^{-1} \left( X_u' X_u \right) \gamma_u^{-1} \Rightarrow \begin{bmatrix} \sigma_u B_u \sigma_u & 0_{4\times4} \\ 0_{4\times4} & Z_u \end{bmatrix}, \tag{A.2}$$

$$\gamma_r^{-1} \left( X_r' X_r \right) \gamma_r^{-1} \Rightarrow \begin{bmatrix} \sigma_r B_r \sigma_r & 0 \\ 0 & Z_r \end{bmatrix}, \tag{A.3}$$

converge jointly as $T \to \infty$, where the sub-matrices are given by

$$B_u = \begin{bmatrix} \int B^2 & \int sB^2 & \int B^3 & \int sB^3 \\ & \int s^2 B^2 & \int sB^3 & \int s^2 B^3 \\ & & \int B^4 & \int sB^4 \\ & & & \int s^2 B^4 \end{bmatrix},$$

$$B_r = \int B^2,$$

and

$$
\begin{aligned}
Z_u &= \begin{bmatrix} m_2 & m_2/2 & m_3 & m_3/2 \\ & m_2/3 & m_3/2 & m_3/3 \\ & & m_4 - m_2^2 & m_4/2 - m_2^2/2 \\ & & & m_4/3 - m_2^2/4 \end{bmatrix}, \\
Z_r &= m_2,
\end{aligned}
$$

where $B_u$ and $Z_u$ are symmetric matrices, $\sigma_u = diag\{\sigma_\epsilon, \sigma_\epsilon, \sigma_\epsilon^2, \sigma_\epsilon^2,\}$, $\sigma_r = \sigma_\epsilon$, and $m_i = \mathsf{E}(\Delta y_{t-1})^i = \mathsf{E}(\epsilon_{t-1}^i)$. In particular, $m_2 = \sigma_\epsilon^2$.

Considering next the limiting results for $\gamma_u^{-1} X_u' \epsilon$ and $\gamma_r^{-1} X_r' \epsilon$ in (A.1) Under Assumption 1, once more using the results in Hansen (1992, Theorem 4.1 and 4.2), He and Sandberg (2006, Lemma A1), Sandberg (2009, Theorem 1 and Corollary 1), and also a central limit theorem for martingale difference sequences (see e.g. White, 2000 Corollary 5.26), it follows that

$$
\gamma_u^{-1} X_u' \epsilon \Rightarrow \sigma_\epsilon \begin{bmatrix} \sigma_u \tilde{B}_u \\ \tilde{Z}_u \end{bmatrix}, \tag{A.4}
$$

$$
\gamma_r^{-1} X_r' \epsilon \Rightarrow \sigma_\epsilon \begin{bmatrix} \sigma_r \tilde{B}_r \\ \tilde{Z}_r \end{bmatrix}, \tag{A.5}
$$

converge jointly as $T \to \infty$, where the sub-vectors are given by

$$
\begin{aligned}
\tilde{B}_u &= \begin{bmatrix} \int BdB \\ \int sBdB \\ \int B^2 dB \\ \int sB^2 dB \end{bmatrix}, \\
\tilde{B}_r &= \int BdB,
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{Z}_u &= \begin{bmatrix} N(0, m_2) \\ N(0, m_2/3) \\ N(0, m_4 - m_2^2) \\ N(0, m_4/3 - m_2^2/4) \end{bmatrix}, \\
\tilde{Z}_r &= N(0, m_2).
\end{aligned}
$$

Here, $\tilde{Z}_u \sim MVN(0, Z_u)$, and $\tilde{Z}_u$ and $\tilde{Z}_r$ are independent of $B(s)$. Finally, combining the results in (A.2)-(A.5), also noticing that $\hat{\epsilon}_u' \hat{\epsilon}_u / T \xrightarrow{p} \sigma_\epsilon^2$ holds under the null hypothesis, the continuous mapping theorem entails

$$
\begin{aligned}
& W_T^{\mathsf{a}} \Rightarrow \\
& \sigma_\epsilon^{-2}\left(\sigma_\epsilon\begin{bmatrix}\sigma_u\tilde{B}_u\\ \tilde{Z}_u\end{bmatrix}\right)'\begin{bmatrix}\sigma_u B_u\sigma_u & 0_{4\times 4}\\ 0_{4\times 4} & Z_u\end{bmatrix}^{-1}\left(\sigma_\epsilon\begin{bmatrix}\sigma_u\tilde{B}_u\\ \tilde{Z}_u\end{bmatrix}\right) \\
& -\sigma_\epsilon^{-2}\left(\sigma_\epsilon\begin{bmatrix}\sigma_r\tilde{B}_r\\ \tilde{Z}_r\end{bmatrix}\right)'\begin{bmatrix}\sigma_r B_r\sigma_r & 0\\ 0 & Z_r\end{bmatrix}^{-1}\left(\sigma_\epsilon\begin{bmatrix}\sigma_r\tilde{B}_r\\ \tilde{Z}_r\end{bmatrix}\right) \\
= \; & \tilde{B}_u' B_u^{-1}\tilde{B}_u - \tilde{B}_r^2 B_r^{-1} + \tilde{Z}_u' Z_u^{-1}\tilde{Z}_u - \tilde{Z}_r^2 Z_r^{-1}. \qquad \text{(A.6)}
\end{aligned}
$$

Here, $\tilde{B}_u' B_u^{-1}\tilde{B}_u - \tilde{B}_r^2 B_r^{-1} \equiv B^{\mathsf{a}}$, and it is also straightforward to show that $\tilde{Z}_u' Z_u^{-1}\tilde{Z}_u - \tilde{Z}_r^2 Z_r^{-1}$ is a $\chi^2(3)$ variate.[13] Finally, by (A.6) it becomes evident that the limiting distribution of $W_T^{\mathsf{a}}$ is nuisance parameter free.[14]

The proof of (iii). The following equivalent expression for the Wald statistic is used

$$
W_T^{\mathsf{a}} = \frac{\left(R_{\mathsf{a}}\hat{\beta}_{\mathsf{a}}\right)'\left(R_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}R_{\mathsf{a}}'\right)^{-1}\left(R_{\mathsf{a}}\hat{\beta}_{\mathsf{a}}\right)}{\hat{\epsilon}_u'\hat{\epsilon}_u/T},
$$

where the matrix $X_{\mathsf{a}}^*$ contains the stacked $x_t^{\mathsf{a}}$, and $R_{\mathsf{a}}$ and $\hat{\beta}_{\mathsf{a}}$ are defined as in (15). Next, partition $X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^*$ as

$$
X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*} = \begin{bmatrix} X_{11}^* & X_{12}^* \\ X_{12}^{*\prime} & X_{22}^* \end{bmatrix},
$$

where

$$
\begin{aligned}
X_{11}^* &= \begin{bmatrix} \sum_t y_{t-1}^2 & \sum_t \Delta y_{t-1}y_{t-1} \\ & \sum_t \left(\Delta y_{t-1}\right)^2 \end{bmatrix}, \\
X_{12}^{*\prime} &= \begin{bmatrix}
\sum_t t^* y_{t-1}^2 & \sum_t t^*\Delta y_{t-1}y_{t-1} \\
\sum_t y_{t-1}^3 & \sum_t \Delta y_{t-1}y_{t-1}^2 \\
\sum_t t^* y_{t-1}^3 & \sum_t t^*\Delta y_{t-1}y_{t-1}^2 \\
\sum_t t^*\Delta y_{t-1}y_{t-1} & \sum_t t^*\left(\Delta y_{t-1}\right)^2 \\
\sum_t \left(\Delta y_{t-1}\right)^2 y_{t-1} & \sum_t \left(\Delta y_{t-1}\right)^3 \\
\sum_t t^*\left(\Delta y_{t-1}\right)^2 y_{t-1} & \sum_t t^*\left(\Delta y_{t-1}\right)^3
\end{bmatrix},
\end{aligned}
$$

and

$$
X_{22}^* = \begin{bmatrix} X_{0,22}^* & X_{1,22}^* \\ X_{1,22}^{*\prime} & X_{2,22}^* \end{bmatrix},
$$

[13] The expression $\tilde{B}_u' B_u^{-1}\tilde{B}_u$ corresponds to the limiting distribution for the linearity test in the TV-STAR model under a unit root assumption by Sandberg (2008) and his expression (2.9) for $i = 3$. In addition, the expression $\tilde{B}_r^2 B_r^{-1}$ corresponds to the square of the Dickey-Fuller unit root $t$-statistic based on a mean-zero AR(1) process.

[14] Corresponding results for $W_T^{\mathsf{b}}$ and $W_T^{\mathsf{c}}$ are available upon request from the authors.

with sub-matrices

$$X_{0,22}^{*} = \begin{bmatrix} \sum_t t^{*2}y_{t-1}^2 & \sum_t t^{*}y_{t-1}^3 & \sum_t t^{*2}y_{t-1}^3 & \sum_t t^{*2}\Delta y_{t-1}y_{t-1} \\ & \sum_t y_{t-1}^4 & \sum_t t^{*}y_{t-1}^4 & \sum_t t^{*}\Delta y_{t-1}y_{t-1}^2 \\ & & \sum_t t^{*2}y_{t-1}^4 & \sum_t t^{*2}\Delta y_{t-1}y_{t-1}^2 \\ & & & \sum_t t^{*2}\left(\Delta y_{t-1}\right)^2 \end{bmatrix},$$

$$X_{1,22}^{*} = \begin{bmatrix} \sum_t t^{*}\left(\Delta y_{t-1}\right)^2 y_{t-1} & \sum_t t^{*2}\left(\Delta y_{t-1}\right)^2 y_{t-1} \\ \sum_t \left(\Delta y_{t-1}\right)^2 y_{t-1}^2 & \sum_t t^{*}\left(\Delta y_{t-1}\right)^2 y_{t-1}^2 \\ \sum_t t^{*}\left(\Delta y_{t-1}\right)^2 y_{t-1}^2 & \sum_t t^{*2}\left(\Delta y_{t-1}\right)^2 y_{t-1}^2 \\ \sum_t t^{*}\left(\Delta y_{t-1}\right)^3 & \sum_t t^{*2}\left(\Delta y_{t-1}\right)^3 \end{bmatrix},$$

$$X_{1,22}^{*} = \begin{bmatrix} \sum_t \left(\Delta y_{t-1}\right)^4 & \sum_t t^{*}\left(\Delta y_{t-1}\right)^4 \\ & \sum_t t^{*2}\left(\Delta y_{t-1}\right)^4 \end{bmatrix},$$

where $X_{11}^{*}$ and $X_{0,22}^{*}$ are symmetric matrices. Consider next the partition $\hat{\beta}_{\mathsf{a}} = \left(\hat{\beta}_L', \hat{\beta}_{NL}'\right)'$ where $\hat{\beta}_L = (\hat{\beta}_1, \hat{\beta}_2)'$ and $\hat{\beta}_{NL} = (\hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_7, \hat{\beta}_8)'$ are the OLS estimators of $\beta_L = (\beta_1, \beta_2)'$ and $\beta_{NL} = (\beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8)'$. This allows us to write $W_T^{\mathsf{a}}$ as

$$\begin{aligned} W_T^{\mathsf{a}} &= \frac{\hat{\beta}_{NL}'\left[X_{22}^{*} - X_{12}^{*\prime}\left(X_{11}^{*}\right)^{-1}X_{12}^{*}\right]\hat{\beta}_{NL}}{\hat{\epsilon}_u'\hat{\epsilon}_u/T} \\ &= \frac{\hat{\beta}_{NL}'\Delta_2^{-1}Q_\Delta\Delta_2^{-1}\hat{\beta}_{NL}}{\hat{\epsilon}_u'\hat{\epsilon}_u/T} \end{aligned}$$

where results on the inverse of partitioned matrices are used to obtain the first equality and $Q_\Delta$ in the second equality is given by

$$Q_\Delta = \Delta_2 X_{22}^{*}\Delta_2 - (\Delta_1 X_{12}^{*}\Delta_2)'(\Delta_1 X_{11}^{*}\Delta_1)^{-1}(\Delta_1 X_{12}^{*}\Delta_2),$$

where $\Delta_1$ and $\Delta_2$ are diagonal scaling matrices, explored in detail below, chosen such that $Q_\Delta$ is tight ($O_p(1)$) and positive definite.

Consider now the case when $y_t$ is non-linear I(0). Letting $\Delta_1 = T^{-1/2}I_2$ and $\Delta_2 = T^{-1/2}I_6$ implies that $Q_\Delta$ is $O_p(1)$ and positive definite, and it also follows that $\Delta_2^{-1}\hat{\beta}_{NL}$ is $O_p(T^{1/2})$ since $\hat{\beta}_{NL} \xrightarrow{p} (\beta_3, \beta_4, \beta_5, 0, 0, 0)'$ where at least one of the parameters $\beta_3$, $\beta_4$, and $\beta_5$ are different from zero. Hence, the numerator of $W_T^{\mathsf{a}}$ is $O_p(T)$. Next, the consistency of the OLS estimators implies that $\hat{\epsilon}_u'\hat{\epsilon}_u/T \xrightarrow{p} \sigma_\epsilon^2$ and $W_T^{\mathsf{a}}$ is thereby $O_p(T)$.

Consider next the case when $y_t$ is non-linear I(1). Letting $\Delta_1 = diag\{T^{-1}, T^{-1/2}\}$ and $\Delta_2 = diag\{T^{-1}, T^{-3/2}, T^{-3/2}, T^{-1/2}, T^{-1/2}, T^{-1/2}\}$ ensures that $Q_\Delta$ is tight and positive definite, and it follows that $\Delta_2^{-1}\hat{\beta}_{NL}$ is $O_p(T^{1/2})$ because $\hat{\beta}_{NL} \xrightarrow{p} (0, 0, 0, \beta_6, \beta_7, \beta_8)'$ where at least one of the parameters $\beta_6$, $\beta_7$, and $\beta_8$ are different from zero. thus, the numerator of $W_T^{\mathsf{a}}$ is $O_p(T)$. Finally, the OLS estimators are also in this case consistent yielding $\hat{\epsilon}_u'\hat{\epsilon}_u/T \xrightarrow{p} \sigma_\epsilon^2$ and $W_T^{\mathsf{a}}$ is $O_p(T)$. ■

**Proof of Theorem 2.** The proof of (i) when if $y_t$ is linear I(0) is, more or less, standard. Next, the proof of (i) when $y_t$ is linear I(1). The expression for $W_T^{\mathsf{a}}$ (as given in the part (iii) of Theorem 1) and $W_{R,T}^{\mathsf{a}}$ can be written as (see e.g. Hamilton, 1994 p. 525)

$$W_T^{\mathsf{a}} = \left(R_{\mathsf{a}}\gamma_{\mathsf{a}}\hat{\beta}_{\mathsf{a}}\right)' \left(R_{\mathsf{a}}\left(\hat{\epsilon}_u'\hat{\epsilon}_u/T\right)\gamma_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}\gamma_{\mathsf{a}}R_{\mathsf{a}}'\right)^{-1}\left(R_{\mathsf{a}}\gamma_{\mathsf{a}}\hat{\beta}_{\mathsf{a}}\right),$$

and

$$W_{R,T}^{\mathsf{a}} = \left(R_{\mathsf{a}}\gamma_{\mathsf{a}}\hat{\beta}_{\mathsf{a}}\right)' \left[R_{\mathsf{a}}\gamma_{\mathsf{a}}\left(\hat{V}_{\mathsf{a}}/T\right)\gamma_{\mathsf{a}}R_{\mathsf{a}}'\right]^{-1}\left(R_{\mathsf{a}}\gamma_{\mathsf{a}}\hat{\beta}_{\mathsf{a}}\right),$$

where $\gamma_{\mathsf{a}} = diag\{T, T^{1/2}, T, T^{3/2}, T^{3/2}, T^{1/2}, T^{1/2}, T^{1/2}\}$. Thus, to prove that $W_{R,T}^{\mathsf{a}} - W_T^{\mathsf{a}} = o_p(1)$ we only have to show that the modified covariance matrices $\left(\hat{\epsilon}_u'\hat{\epsilon}_u/T\right)\gamma_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}\gamma_{\mathsf{a}}$ and $\gamma_{\mathsf{a}}\left(\hat{V}_{\mathsf{a}}/T\right)\gamma_{\mathsf{a}}$ are stochastically equicontinuous. As such, the weak convergence result for $\gamma_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}\gamma_{\mathsf{a}}$ is, more or less, already derived in Theorem 1 and expression (A.2) and will here be signified $\left(\int B_{\mathsf{a}}\right)^{-1}$ for short.[15] It follows that

$$\left(\hat{\epsilon}_u'\hat{\epsilon}_u/T\right)\gamma_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}\gamma_{\mathsf{a}} \Rightarrow \sigma_\epsilon^2\left(\int B_{\mathsf{a}}\right)^{-1}.$$

Consider next the weak convergence results for $\gamma_{\mathsf{a}}\left(\hat{V}_{\mathsf{a}}/T\right)\gamma_{\mathsf{a}}$. Write

$$\gamma_{\mathsf{a}}\left(\hat{V}_{\mathsf{a}}/T\right)\gamma_{\mathsf{a}} = \gamma_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}\gamma_{\mathsf{a}}\left(\gamma_{\mathsf{a}}^{-1}\left[\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right]\gamma_{\mathsf{a}}^{-1}\right)\gamma_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}\gamma_{\mathsf{a}},$$

where it is only the middle component (the modified $\hat{S}_{\mathsf{a}}$-term) on the r.h.s. that must be further examined. The consistency of the OLS estimators yields that $\hat{\epsilon}_t = \epsilon_t + o_p(1)$, and it is straightforward to show that

$$\gamma_{\mathsf{a}}^{-1}\left[\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right]\gamma_{\mathsf{a}}^{-1} = \gamma_{\mathsf{a}}^{-1}\left[\sum\nolimits_t \epsilon_t^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right]\gamma_{\mathsf{a}}^{-1} + o_p(1). \tag{A.7}$$

Furthermore, write

$$\gamma_{\mathsf{a}}^{-1}\left[\sum\nolimits_t \epsilon_t^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right]\gamma_{\mathsf{a}}^{-1} = \gamma_{\mathsf{a}}^{-1}\left[\sum\nolimits_t \sigma_\epsilon^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right]\gamma_{\mathsf{a}}^{-1} + \gamma_{\mathsf{a}}^{-1}\left[\sum\nolimits_t (\epsilon_t^2 - \sigma_\epsilon^2) x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right]\gamma_{\mathsf{a}}^{-1}, \tag{A.8}$$

where $\gamma_{\mathsf{a}}^{-1}\left[\sum_t \sigma_\epsilon^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right]\gamma_{\mathsf{a}}^{-1} \Rightarrow \sigma_\epsilon^2 \int B_{\mathsf{a}}$ and because the condition

$$\lim_{n\to\infty}\sup_t \mathsf{E}\left|\mathsf{E}\left(\epsilon_t^2|\mathcal{F}_{t-n}\right) - \mathsf{E}\epsilon_t^2\right| = 0$$

is trivially fulfilled under the present assumptions ($\mathcal{F}_{t-n}$ is the sigma-algebra generated by $\epsilon_{t-n}, \epsilon_{t-n-1}, \ldots$) we can apply Theorem 3.3 of Hansen (1992) to obtain

$$\sup_{0\le s\le 1}\left|\sum\nolimits_{t=1}^{[sT]}\gamma_{\mathsf{a}}^{-1}\left(x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right)\gamma_{\mathsf{a}}^{-1}(\epsilon_t^2 - \sigma_\epsilon^2)\right| \xrightarrow{p} 0,$$

[15] A complete expression for the weak convergence result of $\gamma_{\mathsf{a}}\left[X_{\mathsf{a}}^{*\prime}X_{\mathsf{a}}^{*}\right]^{-1}\gamma_{\mathsf{a}}$ is straightforward to derive using Theorem 1 in Sandberg (2009), and is available upon request from the authors.

and the second term on the r.h.s. in (A.8) is thus $o_p(1)$. It follows that $\sum_t \epsilon_t^2 \gamma_{\mathsf{a}}^{-1} (x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}) \gamma_{\mathsf{a}}^{-1} \Rightarrow \sigma_\epsilon^2 \int B_{\mathsf{a}}$, and the relationship in (A.7) yields $\gamma_{\mathsf{a}}^{-1} \left[\sum_t \hat{\epsilon}_t^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right] \gamma_{\mathsf{a}}^{-1} \Rightarrow \sigma_\epsilon^2 \int B_{\mathsf{a}}$, and the claim of stochastic equicontinuity between the two modified covariance matrices now follows since

$$\begin{aligned}(\hat{\epsilon}_u' \hat{\epsilon}_u / T) \gamma_{\mathsf{a}} \left[X_{\mathsf{a}}^{*\prime} X_{\mathsf{a}}^*\right]^{-1} \gamma_{\mathsf{a}} - \gamma_{\mathsf{a}} \left(\hat{V}_{\mathsf{a}}/T\right) \gamma_{\mathsf{a}} \quad &\Rightarrow \quad \sigma_\epsilon^2 \left(\int B_{\mathsf{a}}\right)^{-1} - \left(\int B_{\mathsf{a}}\right)^{-1} \left(\sigma_\epsilon^2 \int B_{\mathsf{a}}\right) \left(\int B_{\mathsf{a}}\right)^{-1} \\ &= \quad 0.\end{aligned}$$

The proof of (ii). Notice first that

$$\begin{aligned} R_{\mathsf{a}} \left(\hat{V}_{\mathsf{a}}/T\right) R_{\mathsf{a}}' &= R_{\mathsf{a}} \left(X_{\mathsf{a}}^{*\prime} X_{\mathsf{a}}^*\right)^{-1} \left[\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{\mathsf{a}} x_t^{\mathsf{a}\prime}\right] \left(X_{\mathsf{a}}^{*\prime} X_{\mathsf{a}}^*\right)^{-1} R_{\mathsf{a}}' \\ &= \left(X_{22}^* - X_{12}^{*\prime} \left(X_{11}^*\right)^{-1} X_{12}^*\right)^{-1} \left[\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right] \left(X_{22}^* - X_{12}^{*\prime} \left(X_{11}^*\right)^{-1} X_{12}^*\right)^{-1}, \end{aligned}$$

where $x_t^{*\mathsf{a}} = (t^* y_{t-1}, y_{t-1}^2, t^* y_{t-1}^2, t^* \Delta y_{t-1}, (\Delta y_{t-1})^2, t^* (\Delta y_{t-1})^2)$. Using this result implies that $W_{R,T}^{\mathsf{a}}$ can be written as

$$\begin{aligned} W_{R,T}^{\mathsf{a}} &= \hat{\beta}_{NL}' \left(X_{22}^* - X_{12}^{*\prime} \left(X_{11}^*\right)^{-1} X_{12}^*\right) \left[\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right]^{-1} \left(X_{22}^* - X_{12}^{*\prime} \left(X_{11}^*\right)^{-1} X_{12}^*\right) \hat{\beta}_{NL} \\ &= \hat{\beta}_{NL}' \Delta_2^{-1} Q_\Delta^R \Delta_2^{-1} \hat{\beta}_{NL}, \end{aligned}$$

where

$$\begin{aligned} Q_\Delta^R &= \left[\Delta_2 X_{22}^* \Delta_2 - (\Delta_1 X_{12}^* \Delta_2)' (\Delta_1 X_{11}^* \Delta_1)^{-1} (\Delta_1 X_{12}^* \Delta_2)\right] \times \left[\Delta_2 \left(\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right) \Delta_2\right]^{-1} \\ &\quad \times \left[\Delta_2 X_{22}^* \Delta_2 - (\Delta_1 X_{12}^* \Delta_2)' (\Delta_1 X_{11}^* \Delta_1)^{-1} (\Delta_1 X_{12}^* \Delta_2)\right] \\ &= Q_\Delta \times \left[\Delta_2 \left(\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right) \Delta_2\right]^{-1} \times Q_\Delta. \end{aligned}$$

Here, $Q_\Delta$ is defined as in the proof of Theorem (iii) and in the subsequent discussion we shall prove that the same choices of the diagonal scaling matrices $\Delta_1$ and $\Delta_2$ as those in the proof of Theorem 1(iii) ensure that $W_{R,T}^{\mathsf{a}}$ is $O_p(T)$ whether a non-linear I(0) or I(1) model is considered. Hence, having the matrices $\Delta_1$ and $\Delta_2$ specified as in part (iii) of Theorem 1, it only remains to show that $\Delta_2 \left(\sum_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right) \Delta_2$ is $O_p(1)$ since $Q_\Delta$ is $O_p(1)$ and $\hat{\beta}_{NL}' \Delta_2^{-1}$ is $O_p(T^{1/2})$ as before.

First, when $y_t$ is non-linear I(0), we have that

$$\begin{aligned} \Delta_2 \left(\sum\nolimits_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right) \Delta_2 &= T^{-1} \sum\nolimits_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime} \\ &\xrightarrow{p} \mathsf{E} \sum\nolimits_t \epsilon_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}, \end{aligned}$$

where the existence and finiteness of $\mathsf{E}\sum_t \epsilon_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}$ is asserted by the present assumption and $\Delta_2\left(\sum_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right)\Delta_2$ is thereby $O_p(1)$. Consider finally the case when $y_t$ is non-linear I(1). It is straightforward to show that $\Delta_2\left(\sum_t \hat{\epsilon}_t^2 x_t^{*\mathsf{a}} x_t^{*\mathsf{a}\prime}\right)\Delta_2$ converges weakly under the present moment condition to a matrix of stochastic integrals and is thus $O_p(1)$.[16]

## References


Andrews, D. W., and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

Granger, C. W., and T. Teräsvirta (1993): *Modelling Nonlinear Economic Relationsships*. Oxford University Press.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, New Jersey.

Hansen, B. E. (1992): "Convergence to stochastic integrals for dependent heterogeneous processes," *Econometric Theory*, 8, 489–501.

——— (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413–430.

Harris, D., B. McCabe, and S. Leybourne (2003): "Some limit theory for autocovariances whose order depends on the sample size," *Econometric Theory*, 19, 829–864.

Harvey, D. I., and S. J. Leybourne (2007): "Testing for time series linearity," *Econoemtrics Journal*, 10, 149–165.

He, C., and R. Sandberg (2006): "Dickey-Fuller Type of Tests Against Non-Linear Dynamic Models," *Oxford Bulletin of Economic and Statistics*, 68, 835–861.

Jansen, E. S., and T. Teräsvirta (1996): "Testing Parameter Constancy and Super Exogeneity in Econometric Equations," *Oxford Bulletin of Economics and Statistics*, 58, 735–763.

Juntunen, M., J. Tervo, and J. P. Kaipio (1999): "Stabilization of Subba Rao-Liporace models," *Circuits Systems Signal Process*, 18, 395–406.

Kiliç, R. (2004): "Linearity tests and stationarity," *Econometrics Journal*, 7, 55–62.

Liebscher, E. (2005): "Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes," *Journal of Time Series Analysis*, 26, 669–689.

Lundbergh, S., T. Teräsvirta, and D. van Dijk (2003): "Time-Varying Smooth Transition Autoregressive Models," *Journal of Business and Economic Statistics*, 21, 104–121.


[16] The proof of this claim is similar to the proof of that the l.h.s. in (A.7) converges weakly to a matrix of stochastic integrals ($\sigma_\epsilon^2 \int B_{\mathsf{a}}$) and is therefore omitted.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

Meitz, M., and P. Saikkonen (2008): "Stability of nonlinear AR-GARCH models," *Journal of Time Series Analysis*, 29, 453–475.

Paulsen, J. (1984): "Order determination of multivariate autoregressive time series with unit roots," *Journal of Time Series Analysis*, 5, 115–127.

Pavlidis, E., I. Paya, and D. Peel (2009): "Specifying Smooth TRansition Regression Models in the Presence of Conditional Heteroscedsaticity of Unknown Form," Lancaster University Manangement School, Working Paper No. 2009/009.

Perron, P. (2006): "Dealing with Structural Breaks," *Palgrave Handbook of Econometrics, vol.1: Econometric Theory*, pp. 278–352.

Sandberg, R. (2008): "Critical values for linearity tests in time-varying smooth transition auturegressive type of models when data are highly persistent," *Econometrics Journal*, 11, 638–647.

——— (2009): "Convergence to stochastic power integrals for dependent heterogeneous processes," *Econometric Theory*, 25, 739–747.

Schwarz, G. (1978): "Estimating the dimension of a model," *Annals of Statistics*, 6, 461–464.

Teräsvirta, T. (1994): "Specification, estimation and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89, 208–18.

van Dijk, D., T. Teräsvirta, and P. Franses (2002): "Smooth Transition Autorgeressive Models-A survey of Recent Developments," *Journal of Business and Economic Statistics*, 21, 1–47.

Vogelsang, T. J. (1998): "Trend function hypothesis testing in the presence of serial correlation," *Econometrica*, 66, 123–148.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

——— (2000): *Asymptotic Theory for Econometricians*. Academic Press.

■

# Research Papers 2010

2010-22: Ivan Nourdin, Giovanni Peccati and Mark Podolskij: Quantitative Breuer-Major Theorems

2010-23: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Bootstrapping Density-Weighted Average Derivatives

2010-24: Søren Johansen and Morten Ørregaard Nielsen: Likelihood inference for a fractionally cointegrated vector autoregressive model

2010-25: Tom Engsted and Bent Nielsen: Testing for rational bubbles in a co-explosive vector autoregression

2010-26: Robinson Kruse: On European monetary integration and the persistence of real effective exchange rates

2010-27: Sanne Hiller and Robinson Kruse: Milestones of European Integration: Which matters most for Export Openness?

2010-28: Robinson Kruse: Forecasting autoregressive time series under changing persistence

2010-29: Nikolaus Hautsch and Mark Podolskij: Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence

2010-30: Martin M. Andreasen: Non-linear DSGE Models and The Central Difference Kalman Filter

2010-31: Morten Ørregaard Nielsen and Per Frederiksen: Fully Modified Narrow-Band Least Squares Estimation of Weak Fractional Cointegration

2010-32: Mogens Bladt and Michael Sørensen: Simple simulation of diffusion bridges with application to likelihood inference for diffusions

2010-33: Fernando Baltazar-Larios and Michael Sørensen: Maximum likelihood estimation for integrated diffusion processes

2010-34: Leonidas Tsiaras: The Forecast Performance of Competing Implied Volatility Measures: The Case of Individual Stocks

2010-35: Leonidas Tsiaras: Dynamic Models of Exchange Rate Dependence Using Option Prices and Historical Returns

2010-36: Robinson Kruse and Rickard Sandberg: Linearity Testing in Time-Varying Smooth Transition Autoregressive Models under Unknown Degree of Persistency