RESEARCH REPORT

Serial No. 515

September 2015

TESTS FOR TAR MODELS vs. STAR MODELS – A SEPARATE FAMILY OF HYPOTHESES APPROACH

by

Zhaoxing Gao, Shiqing Ling and Howell Tong

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Pokfulam Road, Hong Kong. E-Mail: saas@hku.hk
Tel.: (852) 3917 2466/7 Fax: (852) 2858 9041

Tests for TAR Models vs. STAR Models–a Separate Family of Hypotheses Approach

ZHAOXING GAO[1], SHIQING LING[2] AND HOWELL TONG[3,4]

[1,2]*Hong Kong University of Science and Technology and* [3]*London School of Economics and Political Science* ,[4]*University of Hong Kong*

Abstract. The threshold autoregressive (TAR) model and the smooth threshold autoregressive (STAR) model have been among the most popular parametric nonlinear time series models for the past three decades or so. However, as yet there is no formal statistical test in the literature for one against the other. The two models are fundamentally different in their autoregressive functions, the TAR model being generally discontinuous while the STAR model being smooth (except in the limit of infinitely fast switching). Following the approach initiated by Cox (1961, 1962), we treat the test problem as one of separate families of hypotheses, thus filling a serious gap in the literature. The test statistic under a STAR model is shown to follow asymptotically a chi-squared distribution, and the one under a TAR model expressed as a functional of a chi-squared process. We present numerical results with both simulated and real data to assess the performance of our procedure.

Key Words: Non-nested test, Separate family of hypotheses, STAR model, TAR model.

1 Introduction

Regime switching models are currently a central area of research activities in time series analysis in both the statistical and the econometric literatures. In the latter, important applications relate to many aspects of economics, e.g., business cycles, unemployment rates, exchange rates, prices, interest rates, and others. As far as time series analysis is concerned, the notion of regime switching can be traced to the introduction of the threshold autoregressive (TAR) model, with Tong (1978) and Tong and Lim (1980) being the initiators; see also Tong (2011). In the non-time series context, the idea of smooth regime switching was first introduced by Bacon and Watts

Address for the correspondence: Howell Tong, Department of Statistics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK.

E-mail: howell.tong@gmail.com

(1971). The idea was later systematically incorporated in the time series literature by Chan and Tong (1986) under the name of a smooth threshold autoregressive (STAR) model, as an extension of the TAR model and the exponential autoregressive model of Ozaki (1980). The STAR model was enthusiastically pursued by the econometricians; see, e.g., Luukkonen et al. (1988), Teräsvirta (1994), van Dijk et al. (2002) and Teräsvirta et al. (2010), who changed *smooth threshold* to *smooth transition*, whilst retaining the same acronym, STAR. However, in applications, practitioners typically assume either a TAR model or a STAR model on prior and often arbitrary grounds. Given the fundamentally different switching characteristics (discontinuous vs. smoothly continuous) of the two models, leading to possibly different interpretations, it is clear that there is a definite need for a statistical test to help us make an informed decision on the basis of our data.

This paper aims to fill this long standing gap. It is also prompted by two of the wishes expressed in Cox (1961, 1962), namely time series and continuous hypothesis vs. discontinuous hypothesis. As far as we are aware, our paper represents the first attempt at testing for separate families of hypotheses in nonlinear time series analysis. However, there is an interesting challenge: although the STAR model includes the TAR model as a special case, it does so only in the form of a limiting case with the switching becoming infinitely fast. This renders standard nested tests impotent. In fact, experience in tests for linearity within TAR models (e.g. Chan and Tong (1990)) shows that the standard likelihood ratio test statistic will follow a complicated distribution, which is typically not a chi-squared distribution. In order to develop a test that has sufficient power and is simple to use in practice, we have to adopt an alternative approach to treat this non-standard problem. In this paper, we shall follow the approach of non-nested tests initiated by Cox (1961, 1962). We shall develop non-nested tests for departure from a STAR/TAR model in the direction of a TAR/STAR model, within the context of a separate families of hypotheses. The separate families are defined by disallowing infinitely fast switching in the STAR model. We show that the test statistic under a STAR model follows a chi-squared distribution, asymptotically, and the one under a TAR model is expressed as a functional of a chi-squared process. Numerical studies are carried out on both simulated and real data to assess the performance of our procedure.

This paper is organized as follows. Section 2 presents the STAR and TAR models, and the non-nested testing procedure. Section 3 derives the asymptotic distributions of the proposed non-nested tests and the related algorithm. Section 4 presents a simulation study. Section 5 analyzes two empirical examples. Section 6 gives the proofs of the theorems.

2 The models and the non-nested testing procedure

The time series $\{y_t : t = 0, \pm 1, \pm 2, ...\}$ is said to follow a STAR(p) model if it satisfies the equation

$$y_t = X'_{t-1}\theta_1 + X'_{t-1}\theta_2 G(q_{t-1}, s, r) + \varepsilon_t, \tag{2.1}$$

where $X_t = (1, y_t, ..., y_{t-p+1})'$, $\theta_i = (\phi_{i0}, \phi_{i1}, ..., \phi_{ip})'$, $i = 1, 2$. $q_t \in \mathcal{F}_t^p$, the σ-field generated by $(y_t, y_{t-1}, ..., y_{t-p+1})$ and $\mathcal{F}_t$ is the σ-field generated by $(y_t, y_{t-1}, ...)$, r is the threshold value and $s > 0$ is the switching parameter. Here, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and variance $0 < \sigma^2 < \infty$, and $G(q_{t-1}, s, r)$ is a smooth switching function; for example, the following logistic smooth switching function is a popular choice

$$G(q_{t-1}, s, r) = \frac{1}{1 + e^{-s(q_{t-1} - r)}}, \tag{2.2}$$

and model (2.1) with logistic smooth switching function (2.2) is commonly called an LSTAR model. The true values of the parameters are denoted by θ_{i0}, s_0 and r_0, respectively. A popular nonlinear time series model is the TAR(p) model defined as

$$y_t = X'_{t-1}\theta_1 + X'_{t-1}\theta_2 I(q_{t-1} > r) + \varepsilon_t, \tag{2.3}$$

where $I(\cdot)$ is the indicator function. Figure 1 plots $I(x > 0)$ and $G(x, s, 0)$ for different s with a fixed threshold $r = 0$. This figure highlights the difficulty in distinguishing a TAR model from a STAR model when s is large. To conform with the notion of separate families, we restrict s to lie in a finite interval, namely $s \in [s_1, s_2]$ with $0 < s_1 < s_2 < \infty$. Similar restriction is assumed for s in the general $G(q_{t-1}, s, r)$. Note that a STAR model has one more parameter (namely s) than a TAR model of the same order.

Let $\theta = (\theta'_1, \theta'_2)'$ and $\lambda = (\theta', s, r)'$. We assume that $\theta \in \Theta \subset R^{2p+2}$, $r \in \Gamma \subset R$ and $\lambda \in \Lambda \subset R^{2p+4}$, where Θ, Γ and Λ are compact sets. We first introduce several assumptions as follows.

Assumption 2.1. $\{y_t\}$ *generated by (2.1) or by (2.3) is strictly stationary and ergodic.*

For assumption 2.1 to hold, see the discussions in Chan and Tong (1986) for the STAR model and Chan (1993) for the TAR model.

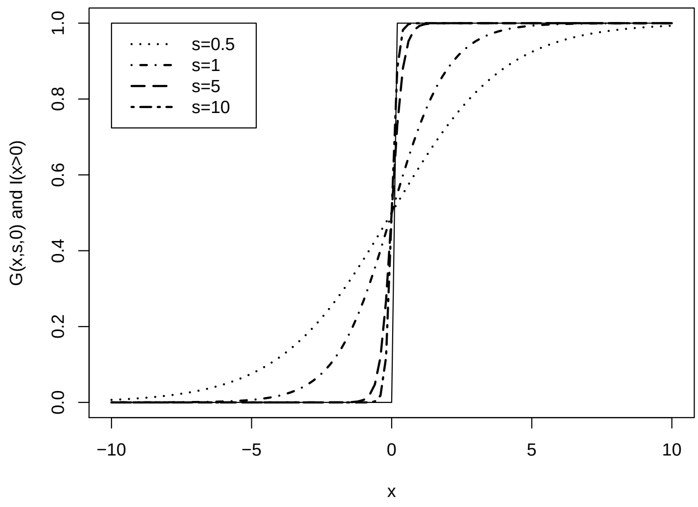

Figure 1: $I(x > 0)$ and $G(x, s, 0)$.

Assumption 2.2. *(i). ε_t and q_t have bounded, continuous and positive densities on R and $E\varepsilon_t^4 < \infty$; (ii). The conditional density of X_t given $q_t = r$ is $f_{X|q}(x|r)$, which is also bounded, continuous and positive on R^{p+1} for all $r \in \Gamma$.*

Assumption 2.2(i) is conventional for the noise ε_t and threshold variable q_t, where the moment condition $E\varepsilon^4 < \infty$ conforms with condition 2 in Chan (1993). Assumption 2.2(ii) implies the existence of the joint density of (X_t', q_t), which is used to establish (6.4).

Assumption 2.3. *(i). $E(\|X_t\|^2|q_t = r) \leq K < \infty$ for all $r \in \Gamma$; (ii). $E(\|X_t\|I(r_1 < q_t \leq r_2)|\mathcal{F}_{t-p}) \leq K\wp_{t-p}|r_2 - r_1|$, where $\wp_{t-p} \in \mathcal{F}_{t-p}$ independent of r_1 and r_2 with $E\wp_{t-p} \leq K < \infty$ for any $r_1 \leq r_2$ in Γ, and $K > 0$ is a constant independent of t and Γ.*

In what follows, we use the notation K as a generic constant whose value can change. By assumption 2.2(ii), assumption 2.3(i) is similar to assumption 1.4 in Hansen (2000), which is a conditional moment condition for $|X_t|$ given q_t, while we only require finite second moment here. Assumption 2.3(ii) is similar to condition (C3) in Chan (1990), while here we use conditional expectation without specifying the form of q_t. When $q_{t-1} = y_{t-d}$ for some $1 \leq d \leq p$, by assumption 2.2, it is not hard to verify assumption 2.3(ii). For example, if $p = 2$ and $d = 2$, then $X_t = (1, y_t, y_{t-1})'$ and $q_t = y_{t-1}$; for the nontrivial term in assumption 2.3(ii) we have

$$E(|y_t|I(r_1 < y_{t-1} \leq r_2|\mathcal{F}_{t-2})$$

$$
\begin{aligned}
&\leq KE[(|\varepsilon_t| + |\varepsilon_{t-1}| + \psi_{t-2})I(r_1 - \phi_{t-2} < \varepsilon_{t-1} \leq r_2 - \phi_{t-2})|\mathcal{F}_{t-2}] \\
&\leq K\kappa_{t-2}E[I(r_1 - \phi_{t-2} < \varepsilon_{t-1} \leq r_2 - \phi_{t-2})|\mathcal{F}_{t-2}] \\
&= K\kappa_{t-2}[F_\varepsilon(r_2 - \phi_{t-2}) - F_\varepsilon(r_1 - \phi_{t-2})] \\
&\leq K\kappa_{t-2}|r_2 - r_1|,
\end{aligned}
$$

where ϕ_{t-2}, ψ_{t-2} and κ_{t-2} are $\mathcal{F}_{t-2}$-measurable functions of the autoregressors, $F_\varepsilon(\cdot)$ is the distribution of ε_t and the last line above is due to Taylor's expansion and the boundedness of the density function of ε_t by assumption 2.2. Define

$$\varepsilon_t(\lambda) = y_t - X'_{t-1}\theta_1 - X'_{t-1}\theta_2 G(q_{t-1}, s, r)$$

and

$$\varepsilon_t(\theta, r) = y_t - X'_{t-1}\theta_1 - X'_{t-1}\theta_2 I(q_{t-1} > r),$$

Denote by $\hat{\lambda}_n$ the least squares estimator (LSE) of λ_0 in model (2.1) and $(\hat{\theta}_n, \hat{r}_n)$ the LSE of (θ_0, r_0) in model (2.3), namely

$$\hat{\lambda}_n = \arg\min_{\lambda\in\Lambda} \sum_{t=1}^{n} \varepsilon_t^2(\lambda), \tag{2.4}$$

$$(\hat{\theta}_n, \hat{r}_n) = \arg\min_{(\theta,r)\in\Theta\times\Gamma} \sum_{t=1}^{n} \varepsilon_t^2(\theta, r). \tag{2.5}$$

We make two assumptions on $\hat{\lambda}_n$ and $(\hat{\theta}_n, \hat{r}_n)$ defined as above.

Assumption 2.4. *Under model (2.1),*

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) = -\Sigma_1^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(\lambda_0)}{\partial \lambda} \varepsilon_t + o_p(1),$$

where $\Sigma_1 = E[\partial\varepsilon_t(\lambda_0)/\partial\lambda \partial\varepsilon_t(\lambda_0)/\partial\lambda']$.

For assumption 2.4 to hold, see the discussion in section 5.2 in van Dijk et al. (2002) on the estimation of STAR model. For general conditions, we refer readers to Klimko and Nelson (1978), Ling and McAleer (2010), among others. When $G(q_{t-1}, s, r)$ is the standard normal distribution function, sufficient conditions are given in Chan and Tong (1986).

Assumption 2.5. *Under model (2.3),* $\hat{r}_n - r_0 = O_p(1/n)$ *and*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\Sigma_2^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(\theta_0, r_0)}{\partial \theta} \varepsilon_t + o_p(1),$$

where $\Sigma_2 = E[\partial\varepsilon_t(\theta_0, r_0)/\partial\theta \partial\varepsilon_t(\theta_0, r_0)/\partial\theta']$.

For assumption 2.5 to hold, we refer to Chan (1993), where V-ergodicity for the time series and discontinuity for the autoregressive function in model (2.3) are discussed.

In the spirit of Cox (1961, 1962), and following Davidson and MacKinnon (1981), MacKinnon et al. (1983), and Bera and McAleer (1989), we can construct a comprehensive or an auxiliary model given by the following linearly weighted competing function:

$$y_t = X'_{t-1}\theta_1 + (1-\delta)X'_{t-1}\theta_2 G(q_{t-1}, s, r) + \delta X'_{t-1}\theta_2 I(q_{t-1} > r) + \varepsilon_t. \tag{2.6}$$

We shall consider testing the hypothesis

$$H_0 : \delta = 0 \text{ against } H_1 : \delta = 1. \tag{2.7}$$

Essentially, we test departure from a STAR model in the direction of a TAR model. Naturally, we can and do also consider testing departure from a TAR model in the direction of a STAR model. Under H_0 and H_1, model (2.6) reduces to model (2.1) and (2.3), respectively. Since model (2.1) and model (2.3) are non-nested, (2.7) is called a non-nested hypothesis.

3 Asymptotic properties of the non-nested tests

We consider the (conditional) quasi-log-likelihood function of model (2.6) as follows.

$$L(\delta, \lambda) = -\frac{1}{2}\sum_{t=1}^{n}[y_t - X'_{t-1}\theta_1 - (1-\delta)X'_{t-1}\theta_2 G(q_{t-1}, s, r) - \delta X'_{t-1}\theta_2 I(q_{t-1} > r)]^2.$$

Denote $D_t(r, s) = G(q_{t-1}, s, r) - I(q_{t-1} > r)$. Under H_0, we obtain the score function and information matrix as follows.

$$\begin{aligned} \frac{\partial L(0, \lambda)}{\partial \delta} &= -\sum_{t=1}^{n}\big\{[y_t - X'_{t-1}\theta_1 - X'_{t-1}\theta_2 G(q_{t-1}, s, r)] \\ &\quad \times [-X'_{t-1}\theta_2 I(q_{t-1} > r) + X'_{t-1}\theta_2 G(q_{t-1}, s, r)]\big\} \\ &= -\sum_{t=1}^{n}\varepsilon_t(\lambda)X'_{t-1}\theta_2 D_t(r, s) \end{aligned} \tag{3.1}$$

and

$$\frac{\partial^2 L(0, \lambda)}{\partial \delta^2} = -\sum_{t=1}^{n}\theta_2' X_{t-1}X'_{t-1}\theta_2 D_t^2(r, s). \tag{3.2}$$

The score based test statistic for testing H_0 against H_1 is defined as

$$T_{1n} = [-\frac{\partial^2 L(0, \hat{\lambda}_n)}{\partial \delta^2}]^{-1}[\frac{\partial L(0, \hat{\lambda}_n)}{\partial \delta}]^2, \tag{3.3}$$

where $\hat{\lambda}_n$ is defined in (2.4). We make one more set of assumptions on the smooth switching function $G(q_{t-1}, s, r)$.

Assumption 3.1.

$$(i).\ |G(q_{t-1}, s, r)| \leq 1;$$
$$(ii).\ |\frac{\partial G(q_{t-1}, s, r)}{\partial s}| \leq K(|q_{t-1}|^{\alpha_1} + 1) \ and\ |\frac{\partial G(q_{t-1}, s, r)}{\partial r}| \leq K(|q_{t-1}|^{\alpha_2} + 1);$$
$$(iii).\ |\frac{\partial^2 G(q_{t-1}, s, r)}{\partial^2 s}| \leq K(|q_{t-1}|^{\alpha_3} + 1) \ and\ |\frac{\partial^2 G(q_{t-1}, s, r)}{\partial^2 r}| \leq K(|q_{t-1}|^{\alpha_4} + 1);$$
$$(iv).\ |\frac{\partial^2 G(q_{t-1}, s, r)}{\partial r \partial s}| \leq K(|q_{t-1}|^{\alpha} + 1),$$

where α_1, α_2, α_3, α_4, $\alpha \geq 0$ and K is a generic constant independent of t as before.

Assumption 3.1(i) is natural because $G(q_{t-1}, s, r)$ is a switching function between 0 to 1, and assumption 3.1(ii)-(iii) are similar to A1-A2 in Francq et al. (2010). However, here we also need the derivatives with respect to the threshold r. Assumptions 3.1(i)-(ii) are needed for the existence of the limiting distributions in theorems 3.1-3.2, and assumptions 3.1(iii)-(iv) are used to prove (6.12). Elementary calculations show that assumptions 3.1(i)-(iv) hold for the LSTAR model with $\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_3 = 2$, $\alpha_4 = 0$ and $\alpha = 1$.

Define

$$\omega_1 = E\{\theta_{20}' X_{t-1} X_{t-1}' \theta_{20} D_t^2(r_0, s_0)\}$$

and

$$\omega_2 = \omega_1 - \{E X_{t-1}' \theta_{20} D_t(r_0, s_0) \frac{\partial \varepsilon_t(\lambda_0)}{\partial \lambda'}\} \Sigma_1^{-1} \{E X_{t-1}' \theta_{20} D_t(r_0, s_0) \frac{\partial \varepsilon_t(\lambda_0)}{\partial \lambda}\}$$

with their estimators

$$\hat{\omega}_{1n} = \frac{1}{n} \sum_{t=1}^{n} \{\hat{\theta}_{2n}' X_{t-1} X_{t-1}' \hat{\theta}_{2n} D_t^2(\hat{r}_n, \hat{s}_n)\}$$

and

$$\hat{\omega}_{2n} = \hat{\omega}_{1n} - \frac{1}{n} \sum_{t=1}^{n} \{X_{t-1}' \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \frac{\partial \varepsilon_t(\hat{\lambda}_n)}{\partial \lambda'}\} \hat{\Sigma}_{1n}^{-1} \frac{1}{n} \sum_{t=1}^{n} \{X_{t-1}' \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \frac{\partial \varepsilon_t(\hat{\lambda}_n)}{\partial \lambda}\},$$

respectively, where $\hat{\Sigma}_{1n}^{-1} = \sum_{t=1}^{n} [\partial \varepsilon_t(\hat{\lambda}_n)/\partial \lambda \partial \varepsilon_t(\hat{\lambda}_n)/\partial \lambda']/n$. Let $\hat{\sigma}_{0n}^2 = -2L(0, \hat{\lambda}_n)/n$. It is not hard to show that $\hat{\sigma}_{0n}^2 \to_p \sigma^2$ as $n \to \infty$ under H_0. Then we can state the following theorem.

Theorem 3.1. *Under H_0, if assumptions 2.1-2.4 and 3.1 hold, and $E\|X_{t-1}\|^2(|q_{t-1}|^{2\kappa} + 1) < \infty$ with $\kappa = \max(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha)$, then*

$$S_{1n} := \frac{T_{1n}}{\hat{\sigma}_{0n}^2} \frac{\hat{\omega}_{1n}}{\hat{\omega}_{2n}} \longrightarrow_{\mathcal{L}} \chi_1^2,$$

as $n \to \infty$, where χ_1^2 is a chi-squared distribution with one degree of freedom.

Next, we discuss the case when H_1 is true (i.e., $\delta = 1$) and we fix $s > 0$ as a constant in (2.1). Under H_1 , we obtain the score function and information matrix as follows.

$$\begin{aligned}\frac{\partial L(1,\lambda)}{\partial \delta} &= -\sum_{t=1}^{n}\{[y_t - X'_{t-1}\theta_1 - X'_{t-1}\theta_2 I(q_{t-1} > r)] \\ &\quad \times [-X'_{t-1}\theta_2 I(q_{t-1} > r) + X'_{t-1}\theta_2 G(q_{t-1}, s, r)]\} \\ &= -\sum_{t=1}^{n} \varepsilon_t(\theta, r) X'_{t-1}\theta_2 D_t(r, s) \end{aligned} \tag{3.4}$$

and

$$\frac{\partial^2 L(1,\lambda)}{\partial \delta^2} = -\sum_{t=1}^{n} \theta'_2 X_{t-1} X'_{t-1} \theta_2 D_t^2(r, s). \tag{3.5}$$

For a given $s > 0$, the score based test statistic for testing H_1 against H_0 is defined as

$$T_{2n}(s) =[-\frac{\partial^2 L(1,\hat{\theta}_n, s, \hat{r}_n)}{\partial \delta^2}]^{-1}[\frac{\partial L(1,\hat{\theta}_n, s, \hat{r}_n)}{\partial \delta}]^2, \tag{3.6}$$

where $\hat{\theta}_n$ and $\hat{r}_n$ are defined in (2.5). In (3.6), we have a nuisance parameter s, which is not identified under H_1. In the spirit of Francq et al. (2010), here we assume $s \in [1/\bar{s}, \bar{s}]$ for an $\bar{s} > 0$ instead of $[s_1, s_2]$. Let $D[1/\bar{s}, \bar{s}]$ be the Skorokhod space and $\Longrightarrow$ be the weak convergence. We have the following theorem.

Theorem 3.2. *Under H_1, if assumptions 2.1-2.3, 2.5 and 3.1 hold, and $E\|X_{t-1}\|^2(|q_{t-1}|^{2\alpha_1} + 1) < \infty$, then,*

$$(a) \quad \frac{1}{\sqrt{n}} \frac{\partial L(1,\hat{\theta}_n, s, \hat{r}_n)}{\partial \delta} \Longrightarrow \sigma Z(s) \quad in \quad D[1/\bar{s}, \bar{s}],$$

$$(b) \quad \sup_{s\in[1/\bar{s},\bar{s}]} \left| -\frac{1}{n}\frac{\partial^2 L(1,\hat{\theta}_n, s, \hat{r}_n)}{\partial \delta^2} - \omega(s)\right| \to_p 0,$$

as $n \to \infty$, where $\omega(s) = E\{\theta'_{20} X_{t-1} X'_{t-1}\theta_{20} D_t^2(r_0, s)\}$, $Z(s)$ is Gaussian process with $EZ(s) = 0$ and $EZ(s)Z(\tau) = E\{\theta'_{20} X_{t-1} X'_{t-1}\theta_{20} D_t(r_0, s) D_t(r_0, \tau)\} - \{EX'_{t-1}\theta_{20} D_t(r_0, s)\partial\varepsilon_t(\theta_0, r_0)/\partial\theta'\}$ $\Sigma_2^{-1}\{EX'_{t-1}\theta_{20} D_t(r_0, \tau)\partial\varepsilon_t(\theta_0, r_0)/\partial\theta\}$.

Remark 3.1. *With the weak convergence of part (a), since $\omega(s)$ and $EZ(s)Z(\tau)$ involve neither derivatives of any order with respect to r nor second-order derivatives with respective to s, and $\varepsilon_t(\theta, r)$ is linear in θ, the moment condition in Theorem 3.2 is slightly weaker than that in Theorem 3.1.*

Following Hansen (1996) and Francq et al. (2010), among others, we use the supremum statistic $\sup_{s\in[1/\bar{s},\bar{s}]} T_{2n}(s)/\hat{\sigma}^2_{1n}$ as our test statistic, where $\hat{\sigma}^2_{1n} = -2L(1,\hat{\theta}_n, s, \hat{r}_n)/n$, which does not depend on s. It is not hard to show that $\hat{\sigma}_{1n}^2 \to_p \sigma^2$ as $n \to \infty$ under H_1. By Theorem 3.2 and the continuous mapping theorem, it follows that

$$S_{2n} := \sup_{s\in[1/\bar{s},\bar{s}]} \frac{T_{2n}(s)}{\hat{\sigma}_{1n}^2} \longrightarrow_{\mathcal{L}} \sup_{s\in[1/\bar{s},\bar{s}]} \frac{Z^2(s)}{\omega(s)},$$

which is the limiting distribution of our test statistic. Following Hansen (1996), Francq et al. (2010) and using (6.19), (6.29) and Glivenko-Cantelli theorem, we can show that the following algorithm can be used to simulate the quantiles of the distribution of $\sup_{s\in[1/\bar{s},\bar{s}]} \frac{Z^2(s)}{\omega(s)}$ conditional on the data $\{y_1, ..., y_n\}$.

Algorithm 1. *For $i = 1, ..., N$:*

- *(i) generate a $N(0,1)$ sample $\varepsilon_1^{(i)}, ..., \varepsilon_n^{(i)}$;*
- *(ii) set*

$$Z_n^{(i)}(s) = -\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} X_{t-1}'\hat{\theta}_{2n} D_t(\hat{r}_n, s)\varepsilon_t^{(i)} + [\frac{1}{n^{3/2}} \sum_{t=p+1}^{n} X_{t-1}'\hat{\theta}_{2n} D_t(\hat{r}_n, s) \times \frac{\partial\varepsilon_t(\hat{\theta}_n, \hat{r}_n)}{\partial\theta'}]\hat{\Sigma}_{2n}^{-1} \sum_{t=p+1}^{n} \varepsilon_t^{(i)} \frac{\partial\varepsilon_t(\hat{\theta}_n, \hat{r}_n)}{\partial\theta}$$

  and

$$\hat{\omega}_n(s) = \frac{1}{n} \sum_{t=p+1}^{n} \{\hat{\theta}_{2n}' X_{t-1} X_{t-1}' \hat{\theta}_{2n} D_t^2(\hat{r}_n, s)\};$$

- *(iii) compute $\sup_{s\in[1/\bar{s},\bar{s}]} \frac{[Z_n{}^{(i)}(s)]^2}{\hat{\omega}_n(s)}$, denoted by $S^{(i)}$,*

where $\hat{\Sigma}_{2n} = \sum_{t=p+1}^{n} [\partial\varepsilon_t(\hat{\theta}_n, \hat{r}_n)/\partial\theta \partial\varepsilon_t(\hat{\theta}_n, \hat{r}_n)/\partial\theta']/n$. Conditional on $\{y_1, ..., y_n\}$, the sequence $\{S^{(i)}, i = 1, ..., N\}$ constitutes an independent and identically distributed sample of the random variable $\sup_{s\in[1/\bar{s},\bar{s}]} \frac{Z^2(s)}{\omega(s)}$. The $(1-\alpha)$-quantile of the distribution of $\sup_{s\in[1/\bar{s},\bar{s}]} \frac{Z^2(s)}{\omega(s)}$ can be approximated by the empirical $(1-\alpha)$-quantile of the artificial sample $\{S^{(i)}, i = 1, ..., N\}$, denoted by c_α. The rejection region of the test at the nominal level α is

$$\left\{ \sup_{s\in[1/\bar{s},\bar{s}]} \frac{Z^2(s)}{\omega(s)} > c_\alpha \right\}.$$

4 Simulation studies

First we examine the performance of the statistic S_{1n} and S_{2n} in finite samples through Monte Carlo experiments. In the experiments, the sample sizes (n) are 400, 800, 1500, 3000 and 5000,

Table 1: Testing H_0 against H_1.

			n				
		α	400	800	1500	3000	5000
size	$s_0 = 2$	0.1	0.136	0.096	0.116	0.102	0.102
		0.05	0.084	0.056	0.046	0.048	0.054
		0.01	0.038	0.0124	0.006	0.010	0.008
size	$s_0 = 5$	0.1	0.100	0.108	0.098	0.102	0.084
		0.05	0.054	0.064	0.046	0.050	0.036
		0.01	0.008	0.010	0.006	0.008	0.014
size	$s_0 = 10$	0.1	0.112	0.108	0.102	0.104	0.100
		0.05	0.046	0.044	0.072	0.048	0.044
		0.01	0.010	0.010	0.018	0.006	0.008
power		0.1	0.516	0.592	0.664	0.830	0.912
		0.05	0.460	0.526	0.610	0.792	0.900
		0.01	0.378	0.390	0.482	0.702	0.844

and the number of replications is 500 for each case. The null hypothesis H_0 is the STAR(1) model with $(\theta_0', r_0) = (-0.9, -0.4, 2, 0.9, 0.8)$ and $s_0 = 2$, 5 and 10, respectively, and the smooth switching function is given by (2.2) with $q_{t-1} = y_{t-1}$. The alternative hypothesis H_1 is a TAR(1) model with $q_{t-1} = y_{t-1}$ and parameters (θ_0', r_0) as before. We set the significance levels at 0.01, 0.05 and 0.1; the corresponding critical values for χ_1^2 are 6.635, 3.841 and 2.706, respectively. We use the package `tsDyn` in `R` software and `lstar` function to fit the logistic STAR model when testing H_0 against From Table 1, it can be seen that the size becomes closer to the nominal level in each case as the sample size increases. Table 1 also shows that the power increases with the sample size. Generally speaking, we require a sample size in excess of 1500 for decent power. H_1. The results are summarized in Table 1.

When Testing H_1 against H_0, we set $\bar{s} = 15$, 30 and 45 in Theorem 3.2. For each $\bar{s}$, under H_1, we consider the cases with $s_0 = 2$, 5 and 10, respectively. We first simulate the critical values by Algorithm 1 in section 3 with $N = 10000$. For each sample size n, conditional on one data set we simulate the critical values c_α with $\alpha = 0.1$, 0.05 and 0.01. Table 2 summarizes the results when $\bar{s} = 15$. Since the results for $\bar{s} = 30$ and 45 are similar, they are not reported here. From Table 2, we can see that at each level, the critical values for the different sample sizes are very close to one another. As a result, we shall adopt their average at each level as the critical value

Table 2: Simulated critical values c_α when testing H_1 against H_0 with $\bar{s} = 15$.

data	s_0	α	n 400	800	1500	3000	5000	average
TAR		0.1	1.84	1.74	1.82	1.75	1.78	1.786
		0.05	2.62	2.49	2.58	2.47	2.53	2.538
		0.01	4.53	4.25	4.60	4.75	4.44	4.514
LSTAR	$s_0 = 2$	0.1	1.39	1.41	1.47	1.32	1.46	1.410
		0.05	2.00	2.03	2.09	1.89	2.08	2.058
		0.01	3.55	3.59	3.60	3.41	3.83	3.596
LSTAR	$s_0 = 5$	0.1	1.65	1.70	1.68	1.72	1.70	1.690
		0.05	2.32	2.48	2.42	2.52	2.48	2.444
		0.01	4.05	4.32	4.14	4.65	4.36	4.304
LSTAR	$s_0 = 10$	0.1	1.80	1.73	1.78	1.81	1.76	1.776
		0.05	2.59	2.48	2.54	2.57	2.58	2.552
		0.01	4.55	4.51	4.49	4.48	4.33	4.472

at that level. Strictly speaking, we should simulate the critical value for each data set and for each sample size n when verifying the efficacy of our test. However, in view of the closeness of the critical values for different sample sizes, we suggest that taking their average as the critical value is a practical way to apply our test. Thus, Table 3 summarizes the simulated critical values with $\bar{s} = 15$, 30 and 45, respectively. For each $\bar{s}$, we choose $s_0 = 1$, 2, 5, 10 and 15 respectively in the LSTAR model.

Based on the critical values in Table 3, we use 500 replications in this experiment for each case and Tables 4–6 report the sizes and powers when testing H_1 against H_0 for $\bar{s} = 15$, 30 and 45, respectively. From Tables 4–6, we can see that the sizes are very close to their nominal levels. We can also see that the power increases with the sample size. For each $\bar{s}$, the power is initially lower when $s_0 = 1$, 2 than that when $s_0 = 5$, 10 and 15, but when the sample size is larger than 1500, all the powers are quite high and even close to 1 when $n \geq 3000$. It is also noted that, when $\bar{s}$ becomes larger, the power seems to decrease slightly at each corresponding slot. Moreover, Tables 4–6 show lower power at $s_0 = 1$ and 2 than at 5, 10 and 15, The explanation for this and the above observation rests with $\tilde{s}_n := \{s : \sup_{s\in[1/\bar{s},\bar{s}]} T_{2n}(s)/\hat{\sigma}^2_{1n}\}$, which, as an estimator of s_0, depends on s_0, n and $\bar{s}$ in a fairly complex manner. Table 7 shows the relation when $n = 400$. It shows the mean of 500 estimators for each s_0. In view of Figure 1, a larger estimator $\tilde{s}_n$ will give

Table 3: Simulated critical values c_α when testing H_1 against H_0.

data	$\bar{s}$	s_0	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
TAR	15		1.786	2.538	4.514
TAR	30		2.153	3.128	5.320
TAR	45		2.370	3.374	6.052
LSTAR	15	1	1.166	1.723	3.103
LSTAR	15	2	1.410	2.058	3.596
LSTAR	15	5	1.690	2.444	4.304
LSTAR	15	10	1.776	2.552	4.472
LSTAR	15	15	1.744	2.495	4.410
LSTAR	30	1	1.783	2.597	4.585
LSTAR	30	2	1.870	2.659	4.693
LSTAR	30	5	2.076	2.941	5.171
LSTAR	30	10	2.181	3.092	5.415
LSTAR	30	15	2.177	3.110	5.397
LSTAR	45	1	2.024	2.886	4.995
LSTAR	45	2	2.224	3.177	5.489
LSTAR	45	5	2.362	3.333	5.760
LSTAR	45	10	2.347	3.306	5.745
LSTAR	45	15	2.286	3.248	5.577

Table 4: Testing H_1 against H_0 when $\bar{s} = 15$.

	data	s_0	α	n 400	800	1500	3000	5000
size	TAR		0.1	0.170	0.154	0.156	0.146	0.160
			0.05	0.070	0.072	0.082	0.080	0.086
			0.01	0.014	0.020	0.020	0.018	0.012
power	LSTAR	$s_0 = 1$	0.1	0.582	0.704	0.748	0.904	0.982
			0.05	0.442	0.532	0.640	0.832	0.950
			0.01	0.198	0.272	0.378	0.638	0.864
power	LSTAR	$s_0 = 2$	0.1	0.382	0.604	0.776	0.886	0.962
			0.05	0.224	0.432	0.642	0.822	0.930
			0.01	0.070	0.170	0.364	0.606	0.820
power	LSTAR	$s_0 = 5$	0.1	0.956	1	1	1	1
			0.05	0.916	1	1	1	1
			0.01	0.720	0.996	1	1	1
power	LSTAR	$s_0 = 10$	0.1	0.910	0.996	1	1	1
			0.05	0.856	0.994	1	1	1
			0.01	0.622	0.970	1	1	1
power	LSTAR	$s_0 = 15$	0.1	0.786	0.990	1	1	1
			0.05	0.690	0.976	1	1	1
			0.01	0.442	0.926	1	1	1

Table 5: Testing H_1 against H_0 when $\bar{s} = 30$.

	data	s_0	α	n				
				400	800	1500	3000	5000
size	TAR		0.1	0.138	0.160	0.180	0.118	0.150
			0.05	0.064	0.078	0.100	0.060	0.080
			0.01	0.012	0.010	0.014	0.008	0.016
power	LSTAR	$s_0 = 1$	0.1	0.584	0.664	0.752	0.892	0.962
			0.05	0.402	0.500	0.622	0.804	0.934
			0.01	0.146	0.198	0.324	0.552	0.784
power	LSTAR	$s_0 = 2$	0.1	0.390	0.520	0.668	0.774	0.864
			0.05	0.220	0.378	0.552	0.672	0.768
			0.01	0.060	0.126	0.270	0.444	0.578
power	LSTAR	$s_0 = 5$	0.1	0.962	1	1	1	1
			0.05	0.888	0.998	1	1	1
			0.01	0.640	0.996	1	1	1
power	LSTAR	$s_0 = 10$	0.1	0.868	0.998	1	1	1
			0.05	0.802	0.996	1	1	1
			0.01	0.534	0.956	1	1	1
power	LSTAR	$s_0 = 15$	0.1	0.786	0.980	1	1	1
			0.05	0.638	0.952	1	1	1
			0.01	0.342	0.842	1	1	1

Table 6: Testing H_1 against H_0 when $\bar{s} = 45$.

	data	s_0	α	n: 400	800	1500	3000	5000
			0.1	0.152	0.162	0.142	0.164	0.168
size	TAR		0.05	0.066	0.070	0.080	0.086	0.082
			0.01	0.020	0.010	0.012	0.014	0.018
			0.1	0.588	0.692	0.816	0.914	0.960
power	LSTAR	$s_0 = 1$	0.05	0.420	0.492	0.676	0.818	0.920
			0.01	0.148	0.182	0.354	0.538	0.778
			0.1	0.330	0.496	0.628	0.746	0.778
power	LSTAR	$s_0 = 2$	0.05	0.182	0.322	0.470	0.600	0.668
			0.01	0.034	0.096	0.238	0.380	0.442
			0.1	0.930	1	1	1	1
power	LSTAR	$s_0 = 5$	0.05	0.832	1	1	1	1
			0.01	0.518	0.986	1	1	1
			0.1	0.842	0.996	1	1	1
power	LSTAR	$s_0 = 10$	0.05	0.728	0.994	1	1	1
			0.01	0.410	0.930	1	1	1
			0.1	0.716	0.978	1	1	1
power	LSTAR	$s_0 = 15$	0.05	0.564	0.962	1	1	1
			0.01	0.284	0.826	0.998	1	1

Table 7: The realized estimator $\tilde{s}_n$ for different true value s_0 under H_1 when $n = 400$.

$\bar{s}$	s_0: 0.5	1	2	5	8	10	15	20
15	13.37	13.23	9.00	6.54	8.65	9.73	11.38	12.06
30	24.13	24.07	20.69	6.75	9.23	10.66	13.83	16.5
45	32.64	32.83	31.56	7.83	9.38	10.94	15.10	18.05
100	56.65	55.57	58.6	20.04	16.74	16.99	21.29	25.72

rise to less difference between the smooth function and the indicator function and hence a lower power, and a smaller one will give higher power. The result in Table 7 conforms to the ones we obtained in Tables 4–6.

5 Real data examples

In this section, we re-visit two real data sets to illustrate our tests. Now, Teräsvirta et al. (2010) fitted (on p. 390) an LSTAR model to the Wolf's sunspot numbers (1700 to 1979) and van Dijk et al. (2002) fitted a similar model to the U.S. unemployment rate. Later, Ekner and Nejstgaard (2013) examined the profile likelihoods of the switching parameter of the above two examples, after an appropriate reparametrization.

The first data set consists of the Wolf's annual sunspot numbers, which are available at the Belgian web page of Solar Influences Data Analysis Center.*. Teräsvirta et al. (2010) fitted an LSTAR model to the sunspot numbers for the period 1700-1979. Following Ghaddar and Tong (1981), they used the square-root transformed sunspot numbers, namely $y_t = 2\{(1+z_t)^{1/2} - 1\}$, where z_t is the original sunspot number. Ekner and Nejstgaard (2013) reproduced the LSTAR model as well as fitted a TAR model as follows (standard deviations in parentheses):†

$$\begin{aligned} H_0: \quad y_t = \; & \underset{(0.08)}{1.46}y_{t-1} - \underset{(0.13)}{0.76}y_{t-2} + \underset{(0.05)}{0.17}y_{t-7} + \underset{(0.04)}{0.11}y_{t-9} \\ & +(\underset{(0,85)}{2.65} - \underset{(0.13)}{0.54}y_{t-1} + \underset{(0.18)}{0.75}y_{t-2} - \underset{(0.11)}{0.47}y_{t-3} \\ & +\underset{(0.11)}{0.32}y_{t-4} - \underset{(0.07)}{0.26}y_{t-5} - \underset{(0.05)}{0.24}y_{t-8} + \underset{(0.06)}{0.17}y_{t-10})\hat{G}(y_{t-2}, 5.46/\hat{\sigma}_{y_{t-2}}, 7.88) \end{aligned} \quad (5.1)$$

and

$$\begin{aligned} H_1: \quad y_t = \; & \underset{(0.08)}{1.43}y_{t-1} - \underset{(0.14)}{0.77}y_{t-2} + \underset{(0.05)}{0.17}y_{t-7} + \underset{(0.05)}{0.12}y_{t-9} \\ & +(\underset{(0,70)}{2.69} - \underset{(0.11)}{0.45}y_{t-1} + \underset{(0.18)}{0.69}y_{t-2} - \underset{(0.11)}{0.48}y_{t-3} \\ & +\underset{(0.11)}{0.36}y_{t-4} - \underset{(0.07)}{0.27}y_{t-5} - \underset{(0.05)}{0.21}y_{t-8} + \underset{(0.05)}{0.14}y_{t-10})I(y_{t-2} > 6.39), \end{aligned} \quad (5.2)$$

*http://www.sidc.oma.be/sunspot-data/

†There are very minor differences between three of the estimated parameters, most probably due to rounding from two decimal places to one in Teräsvirta et al. (2010).

Table 8: Testing (5.1) against (5.2) (NR=not rejected)

	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	p-value
Decision	NR	NR	NR	0.764

Table 9: Testing (5.2) against (5.1).

	$\bar{s}$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	p-value
Decision	15	NR	NR	NR	0.964
	30	NR	NR	NR	0.958
	45	NR	NR	NR	0.962

where $\hat{\sigma}_{y_{t-2}}$ is the standard deviation of $q_{t-1} = y_{t-2}$, $\hat{\sigma}^2_{0n} = 3.414$ and $\hat{\sigma}^2_{1n} = 3.410$. From the data, we obtain $\hat{\sigma}_{y_{t-2}} = 5.57$, giving $\hat{s}_n = 0.98$. When testing H_0 (i.e., (5.1)) against H_1 (i.e., (5.2)), the results are summarized in Table 8. From Table 8, we can see that we do not reject (5.1) at each of the three levels and the $p-$value is 0.764. Then we test H_1 against H_0 and we choose $\bar{s} = 15$, 30, and 45, respectively. The results are summarized in Table 9. From Table 9, we can see that we again do not reject (5.2) at each of the three levels and for each $\bar{s}$, and the $p-$values are 0.964, 0.958 and 0.962, respectively. Tables 8 and 9 suggest that given a sample size of only 280 and the fairly large number of parameters (14 for (5.1) and 13 for (5.2)), neither test seems to enjoy sufficient power to detect departure from one model in the direction of the other. However, the difference between the near-unity $p-$values in Table 9 as against the $p-$value of 0.764 in Table 8 suggests that, if properly reformulated as Bayesian posterior odds, it can lend credence to the conclusion of Ekner and Nejstgaard (2013), which finds from their profile likelihood analysis that 'the global maximum is actually the TAR model' whereas the STAR model adopted by Teräsvirta et al. (2010) is only a local maximum.

In the second example, we re-examine the monthly seasonally unadjusted unemployment rate for U.S. males aged 20 and over for the period 1968:6-1989:12, to which van Dijk et al. (2002) fitted an LSTAR model.[‡] Ekner and Nejstgaard (2013) re-examined the above LSTAR model as well as fitted a TAR model as follows (standard deviations in parentheses).

[‡] The series is constructed from data on the unemployment level and labor force for the particular sub-population. These two series are published together with Gauss programs used to estimate their model at http://swopec.hhs.se/hastef/abs/hastef0380.htm.

$$
\begin{aligned}
H_0: \quad \Delta y_t \;=\; & \underset{(0.07)}{0.479} + \underset{(0.07)}{0.645}D_{1,t} - \underset{(0.10)}{0.342}D_{2,t} - \underset{(0.09)}{0.68}D_{3,t} - \underset{(0.11)}{0.725}D_{4,t} - \underset{(0.10)}{0.649}D_{5,t} \\
& -\underset{(0.09)}{0.317}D_{6,t} - \underset{(0.09)}{0.410}D_{7,t} - \underset{(0.09)}{0.501}D_{8,t} - \underset{(0.09)}{0.554}D_{9,t} - \underset{(0.07)}{0.306}D_{10,t} \\
& +[-\underset{(0.01)}{0.040}y_{t-1} - \underset{(0.08)}{0.146}\Delta y_{t-1} - \underset{(0.06)}{0.101}\Delta y_{t-6} + \underset{(0.06)}{0.097}\Delta y_{t-8} - \underset{(0.06)}{0.123}\Delta y_{t-10} \\
& +\underset{(0.07)}{0.129}\Delta y_{t-13} - \underset{(0.06)}{0.103}\Delta y_{t-15}] \times [1 - \hat{G}(\Delta_{12} y_{t-1}, 23.15/\hat{\sigma}_{\Delta_{12} y_{t-1}}, 0.274)] \\
& +[-\underset{(0.01)}{0.011}y_{t-1} + \underset{(0.08)}{0.225}\Delta y_{t-1} + \underset{(0.08)}{0.307}\Delta y_{t-2} - \underset{(0.07)}{0.119}\Delta y_{t-7} - \underset{(0.09)}{0.155}\Delta y_{t-13} \\
& -\underset{(0.09)}{0.215}\Delta y_{t-14} - \underset{(0.09)}{0.235}\Delta y_{t-15}] \times \hat{G}(\Delta_{12} y_{t-1}, 23.15/\hat{\sigma}_{\Delta_{12} y_{t-1}}, 0.274)
\end{aligned} \tag{5.3}
$$

and

$$
\begin{aligned}
H_1: \quad \Delta y_t \;=\; & \underset{(0.07)}{0.473} + \underset{(0.07)}{0.644}D_{1,t} - \underset{(0.10)}{0.343}D_{2,t} - \underset{(0.09)}{0.675}D_{3,t} - \underset{(0.11)}{0.721}D_{4,t} - \underset{(0.10)}{0.641}D_{5,t} \\
& -\underset{(0.09)}{0.308}D_{6,t} - \underset{(0.09)}{0.410}D_{7,t} - \underset{(0.08)}{0.505}D_{8,t} - \underset{(0.09)}{0.546}D_{9,t} - \underset{(0.07)}{0.295}D_{10,t} \\
& +[-\underset{(0.01)}{0.040}y_{t-1} - \underset{(0.08)}{0.14}\Delta y_{t-1} - \underset{(0.06)}{0.094}\Delta y_{t-6} + \underset{(0.06)}{0.092}\Delta y_{t-8} - \underset{(0.06)}{0.116}\Delta y_{t-10} \\
& +\underset{(0.07)}{0.136}\Delta y_{t-13} - \underset{(0.06)}{0.106}\Delta y_{t-15}] \times I(\Delta_{12} y_{t-1} \le 0.268) \\
& +[-\underset{(0.01)}{0.012}y_{t-1} + \underset{(0.08)}{0.227}\Delta y_{t-1} + \underset{(0.08)}{0.307}\Delta y_{t-2} - \underset{(0.07)}{0.094}\Delta y_{t-7} - \underset{(0.09)}{0.146}\Delta y_{t-13} \\
& -\underset{(0.09)}{0.211}\Delta y_{t-14} - \underset{(0.09)}{0.216}\Delta y_{t-15}] \times I(\Delta_{12} y_{t-1} > 0.268)
\end{aligned} \tag{5.4}
$$

where $\Delta y_t = y_t - y_{t-1}$, $\Delta_{12} y_t = y_t - y_{t-12}$, $\hat{\sigma}^2_{0n} = 0.03407$ and $\hat{\sigma}^2_{1n} = 0.03412$, and $D_{i,t}$ is monthly dummy variable where $D_{i,t} = 1$ if observation t corresponds to month i and $D_{i,t} = 0$ otherwise. From the data, we obtain $\hat{\sigma}_{\Delta_{12} y_{t-1}} = 1.35$, giving $\hat{s}_n = 17.15$. The results of testing H_0 (i.e., (5.3)) against H_1 (i.e., (5.4)) are summarized in Table 10. From Table 10, we can see that we reject (5.3) at 0.1 significance level and do not reject it at the 0.05 and 0.01 levels, and the $p-$value is 0.075. Then we test H_1 against H_0 and choose $\bar{s} = 15$, 30 and 45, respectively. The results are summarized in Table 11. From Table 11, we can see that we do not reject (5.4) at any of the three levels for each $\bar{s}$, and the $p-$value is 0.99 for each $\bar{s}$. Overall, the results tend to suggest that a TAR model is more plausible than a STAR model. The same conclusion was drawn by Ekner and Nejstgaard (2013), who found that for the STAR model, the profile likelihood of the s parameter is rather flat and the maximum occurs at a rather large value of s; they concluded that 'a large and imprecise estimate of s implies that the LSTAR model is effectively a TAR

Table 10: Testing (6.3) against (6.4). (NR=not rejected)

	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	p-value
Decision	**rejected**	NR	NR	0.075

Table 11: Testing (6.4) against (6.3).

	$\bar{s}$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	p-value
Decision	15	NR	NR	NR	0.99
	30	NR	NR	NR	0.99
	45	NR	NR	NR	0.99

model.' However, the very large number of parameters for both models tends to suggest some model over-parametrization.

6 Proofs of Theorems 3.1-3.2

To prove Theorem 3.1, we need the following basic lemma.

Lemma 6.1. *$\{X_t\}$ is a strictly stationary and ergodic process, $f(X_t, \theta)$ is a measurable function with respect to X_t and $\theta \in \Theta$, which is a compact set in R^d for some integer $d > 0$.*

(i) If $E \sup_{\theta\in\Theta} |f(X_t, \theta)| < \infty$ and $Ef(X_t, \theta)$ is continuous in θ, then, for any $\epsilon > 0$, there exists an $\eta > 0$ such that

$$\lim_{n\to\infty} P\Big(\sup_{\|\theta-\theta_0\|\leq\eta} \frac{1}{n} \Big|\sum_{t=1}^{n} [f(X_t, \theta) - f(X_t, \theta_0)]\Big| \geq \epsilon\Big) = 0; \tag{6.1}$$

(ii) If $f(X_t, \theta)$ satisfies assumption 2.3 with $\|X_t\|$ and Γ replaced by $|f(X_t, \theta)|$ and $[0, \frac{M}{\sqrt{n}}]$ for any $\theta \in \Theta$ and $M > 0$, respectively, and $q_t \in \mathcal{F}_t^p$, which has bounded, continuous and positive density $f_q(x)$ on R, then, for any $\epsilon > 0$ and $\theta_0 \in \Theta$,

$$\lim_{n\to\infty} P\Big(\sup_{0\leq r\leq \frac{M}{\sqrt{n}}} \frac{1}{\sqrt{n}} \Big|\sum_{t=1}^{n} f(X_t, \theta_0) I(0 < q_t \leq r)\varepsilon_t\Big| \geq \epsilon\Big) = 0, \tag{6.2}$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence independent of $\mathcal{F}_t$ with mean zero and finite variance.

Proof. (i). Let

$$H_t(\eta) = \sup_{\|\theta-\theta_0\|\leq\eta} |f(X_t, \theta) - f(X_t, \theta_0)|.$$

Since $E\sup_{\theta\in\Theta}|f(X_t,\theta)|<\infty$ and $Ef(X_t,\theta)$ is continuous in θ, for any $\epsilon>0$, there exists an $\eta>0$ small enough, such that $EH_t(\eta)<\epsilon/2$. As $H_t(\eta)$ is strictly stationary and ergodic, by ergodic theorem, we have

$$\lim_{n\to\infty}P(\frac{1}{n}\sum_{t=1}^{n}H_t(\eta)\geq\epsilon)\leq\lim_{n\to\infty}P(\frac{1}{n}|\sum_{t=1}^{n}[H_t(\eta)-EH_t(\eta)]|\geq\frac{\epsilon}{2})=0.$$

Thus, (6.1) holds.

(ii). As the interval $[0,M]$ is compact, for any small $\delta>0$, there is a finite integer $N>0$ such that $0=M_0\leq M_1\leq...\leq M_N=M$ with $|M_i-M_{i-1}|\leq\delta$, $i=1,...,N$. Then,

$$\begin{aligned}
&P(\sup_{0\leq r\leq\frac{M}{\sqrt{n}}}\frac{1}{\sqrt{n}}|\sum_{t=1}^{n}f(X_t,\theta_0)I(0<q_t\leq r)\varepsilon_t|\geq\epsilon)\\
\leq&P(\sup_{1\leq i\leq N}\sup_{\frac{M_{i-1}}{\sqrt{n}}\leq r\leq\frac{M_i}{\sqrt{n}}}\frac{1}{\sqrt{n}}|\sum_{t=1}^{n}f(X_t,\theta_0)I(0<q_t\leq r)\varepsilon_t|\geq\epsilon)\\
\leq&P(\sup_{1\leq i\leq N}\sup_{\frac{M_{i-1}}{\sqrt{n}}\leq r\leq\frac{M_i}{\sqrt{n}}}\frac{1}{\sqrt{n}}|\sum_{t=1}^{n}f(X_t,\theta_0)I(\frac{M_{i-1}}{\sqrt{n}}<q_t\leq r)\varepsilon_t|\geq\epsilon/2)\\
&+\sum_{i=1}^{N}P(\frac{1}{\sqrt{n}}|\sum_{t=1}^{n}f(X_t,\theta_0)I(0<q_t\leq\frac{M_{i-1}}{\sqrt{n}})\varepsilon_t|\geq\epsilon/2)\\
\leq&\left\{\sum_{i=1}^{N}P\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\leq\frac{M_i}{\sqrt{n}})\right.\right.\right.\\
&\left.\left.-E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\leq\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-1})\right]\geq\frac{\epsilon}{2(p+1)}\right)+\cdots\\
&+\sum_{i=1}^{N}P\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\leq\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p+1})\right.\right.\\
&\left.\left.\left.-E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\leq\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p})\right]\geq\frac{\epsilon}{2(p+1)}\right)\right\}\\
&+P(\sup_{1\leq i\leq N}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\leq\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p})\geq\frac{\epsilon}{2(p+1)})\\
&+\sum_{i=1}^{N}P(\frac{1}{\sqrt{n}}|\sum_{t=1}^{n}f(X_t,\theta_0)I(0<q_t\leq\frac{M_{i-1}}{\sqrt{n}})\varepsilon_t|\geq\epsilon/2)\\
\triangleq&\Pi_{1n}+\Pi_{2n}+\Pi_{3n}.
\end{aligned}\tag{6.3}$$

For any random variable Z, if the joint density of (Z,q_t) exists, we have

$$\frac{d}{dr}E[ZI(q_t\leq r)]=E[Z|q_t=r]f_q(r),$$

then, for any r_1, $r_2\in\Gamma$ with $r_1<r_2$, by Taylor's expansion,

$$|E[ZI(r_1<q_r\leq r_2)]|=|E[Z|q_t=r^*]f_q(r^*)|\,|r_2-r_1|,\tag{6.4}$$

where r^* lies between r_1 and r_2.

By (6.4) and assumption 2.2-2.3, we have the following three inequalities in order,

$$\begin{aligned}
&E\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-j})\right.\right.\\
&\left.\left.-E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-j-1})\right]\right)^2\\
\le&2E|f(X_t,\theta_0)\varepsilon_t|^2I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})\\
\le&K\frac{\delta}{\sqrt{n}},
\end{aligned}\tag{6.5}$$

$$\begin{aligned}
&E[\sup_{1\le i\le N}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p})]\\
=&E[\sup_{1\le i\le N}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\wp_{t-p}\frac{M_i-M_{i-1}}{\sqrt{n}}]E|\varepsilon_t|\\
\le&\delta\{\frac{1}{n}\sum_{t=1}^{n}E[\wp_{t-p}]\}E|\varepsilon_t|\\
\le&K\delta,
\end{aligned}\tag{6.6}$$

and

$$\begin{aligned}
&E(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}f(X_t,\theta_0)I(0<q_t\le\frac{M_{i-1}}{\sqrt{n}})\varepsilon_t)^2\\
=&Ef(X_t,\theta_0)^2\varepsilon_t^2I(0<q_t\le\frac{M_{i-1}}{\sqrt{n}})\\
\le&K\frac{M_{i-1}}{\sqrt{n}},
\end{aligned}\tag{6.7}$$

where $j=0,1,...,p-1$, $\wp_{t-p}$ is defined in assumption 2.2 and $K>0$ is a generic constant independent of t.

By Markov inequality and (6.5)-(6.7), we have

$$\Pi_{1n}+\Pi_{2n}+\Pi_{3n}\le\sum_{i=1}^{N}\frac{Kp\delta}{\sqrt{n}[\epsilon/(2(p+1))]^2}+\frac{K\delta}{[\epsilon/(2(p+1))]}+\sum_{i=1}^{N}\frac{M_{i-1}}{\sqrt{n}(\epsilon/2)^2}\to 0,\tag{6.8}$$

as $n\to\infty$ and $\delta\to 0$. Then, (6.2) follows from (6.3) and (6.8). $\square$

Proof of Theorem 3.1. Under H_0, by Taylor's expansion, we have

$$\varepsilon_t(\hat{\lambda}_n)=\varepsilon_t(\lambda_0)+\frac{\partial\varepsilon_t(\lambda_{nt})}{\partial\lambda}(\hat{\lambda}_n-\lambda_0)$$

$$= \varepsilon_t + \frac{1}{\sqrt{n}} \frac{\partial \varepsilon_t(\lambda_{nt})}{\partial \lambda'} \sqrt{n}(\hat{\lambda}_n - \lambda_0), \tag{6.9}$$

where λ_{nt} lies between $\hat{\lambda}_n$ and λ_0 for each t. Then, it follows that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \frac{\partial L(0, \hat{\lambda}_n)}{\partial \delta} =& - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \varepsilon_t \\
& - \frac{1}{n} \sum_{t=1}^{n} X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \frac{\partial \varepsilon_t(\lambda_{nt})}{\partial \lambda'} \sqrt{n}(\hat{\lambda}_n - \lambda_0) \\
=& - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \varepsilon_t \\
& - \frac{1}{n} \sum_{t=1}^{n} X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \frac{\partial \varepsilon_t(\hat{\lambda}_n)}{\partial \lambda'} \sqrt{n}(\hat{\lambda}_n - \lambda_0) + R_n,
\end{aligned} \tag{6.10}$$

where

$$\begin{aligned}
R_n =& \frac{1}{n} \sum_{t=1}^{n} X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \left(\frac{\partial \varepsilon_t(\hat{\lambda}_n)}{\partial \lambda'} - \frac{\partial \varepsilon_t(\lambda_{nt})}{\partial \lambda'} \right) \sqrt{n}(\hat{\lambda}_n - \lambda_0) \\
=& \frac{1}{n^{3/2}} \sum_{t=1}^{n} X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \sqrt{n}(\hat{\lambda}_n - \lambda_{nt})' \frac{\partial^2 \varepsilon_t(\lambda^*_{nt})}{\partial \lambda \partial \lambda'} \sqrt{n}(\hat{\lambda}_n - \lambda_0),
\end{aligned} \tag{6.11}$$

where λ^*_{nt} lies between $\hat{\lambda}_n$ and λ_{nt} for each t. By assumptions 2.1-2.4 and the definition of λ_{nt} in (6.9), $\sqrt{n}(\hat{\lambda}_n - \lambda_0) = O_p(1)$, $\sup_{t \le n} \sqrt{n}|\hat{\lambda}_n - \lambda_{nt}| \le \sqrt{n}|\hat{\lambda}_n - \lambda_0| = O_p(1)$. For any matrix or vector $A = (a_{ij})$, we introduce the notation $|A| = (|a_{ij}|)$ in this proof. Then, by assumption 3.1(iii)-(iv),

$$\begin{aligned}
|R_n| \le& \sqrt{n} |(\hat{\lambda}_n - \lambda_0)'| \frac{1}{n^{3/2}} \sum_{t=1}^{n} |X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n)| \left| \frac{\partial^2 \varepsilon_t(\lambda^*_{nt})}{\partial \lambda \partial \lambda'} \right| |\sqrt{n}(\hat{\lambda}_n - \lambda_0)| \\
\le& \sqrt{n} |(\hat{\lambda}_n - \lambda_0)'| \frac{K}{n^{3/2}} \sum_{t=1}^{n} |X'_{t-1} \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n)| |M(X_{t-1}, q_{t-1})| |\sqrt{n}(\hat{\lambda}_n - \lambda_0)|,
\end{aligned}$$

where $M(X_{t-1}, q_{t-1})$ is defined as

$$M(X_{t-1}, q_{t-1}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P(X_{t-1}, q_{t-1}) \end{pmatrix}_{(2p+4)\times(2p+4)},$$

where

$$P(X_{t-1}, q_{t-1}) = \begin{pmatrix} \mathbf{0} & |X_{t-1}||q_{t-1}|^{\alpha_1} & |X_{t-1}||q_{t-1}|^{\alpha_2} \\ |X'_{t-1}||q_{t-1}|^{\alpha_1} & \|X_{t-1}\||q_{t-1}|^{\alpha_3} & \|X_{t-1}\||q_{t-1}|^{\alpha} \\ |X'_{t-1}||q_{t-1}|^{\alpha_1} & \|X_{t-1}\||q_{t-1}|^{\alpha} & \|X_{t-1}\||q_{t-1}|^{\alpha_4} \end{pmatrix}_{(p+3)\times(p+3)}.$$

By assumption 2.4 and Lemma 6.1(i) it is not hard to show that

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}|X_{t-1}'\hat{\theta}_{2n}D_t(\hat{r}_n,\hat{s}_n)||M(X_{t-1},q_{t-1})| = o_p(1).$$

Thus,

$$R_n = o_p(1). \tag{6.12}$$

Now, we look at the first term on the right-hand side of (6.10). Let $\xi = (\theta_2', s, r)'$ and $g_t(\xi) = X_{t-1}'\theta_2 G(q_{t-1}, s, r)$, by Taylor's expansion, assumption 2.4 and Lemma 6.1(i), we can show that, for some ξ_n^* lying between $\hat{\xi}_n$ and ξ_0,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_t(\hat{\xi}_n)\varepsilon_t =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_t(\xi_0)\varepsilon_t + [\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t(\xi_n^*)}{\partial \xi'}\varepsilon_t]\sqrt{n}(\hat{\xi}_n-\xi_0)\\ =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_t(\xi_0)\varepsilon_t + o_p(1). \end{aligned}\tag{6.13}$$

and

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{t-1}'\hat{\theta}_{2n}I(q_{t-1}>\hat{r}_n)\varepsilon_t =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{t-1}'\theta_{20}I(q_{t-1}>\hat{r}_n)\varepsilon_t\\ &+ [\frac{1}{n}\sum_{t=1}^{n}X_{t-1}'I(q_{t-1}>\hat{r}_n)\varepsilon_t]\sqrt{n}(\hat{\theta}_{2n}-\theta_0)\\ =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{t-1}'\theta_{20}I(q_{t-1}>\hat{r}_n)\varepsilon_t + o_p(1). \end{aligned}\tag{6.14}$$

By Lemma 6.1(ii) and assumption 2.4, we can also show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{t-1}'\theta_{20}I(q_{t-1}>\hat{r}_n)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{t-1}'\theta_{20}I(q_{t-1}>r_0)\varepsilon_t + o_p(1). \tag{6.15}$$

By (6.10), (6.12)-(6.15), assumption 2.4 and Lemma 6.1(i), it follows that

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial L(0,\hat{\lambda}_n)}{\partial \delta} =& -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{t-1}'\theta_{20}D_t(r_0,s_0)\varepsilon_t\\ &+ [\frac{1}{n}\sum_{t=1}^{n}X_{t-1}'\theta_{20}D_t(r_0,s_0)\frac{\partial \varepsilon_t(\lambda_0)}{\partial \lambda'}]\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \varepsilon_t(\lambda_0)}{\partial \lambda}\varepsilon_t + o_p(1). \end{aligned}\tag{6.16}$$

By ergodic theorem and central limit theorem, we have

$$\frac{1}{\sqrt{n}}\frac{\partial L(0,\hat{\lambda}_n)}{\partial \delta} \longrightarrow_{\mathcal{L}} N(0,\sigma^2\omega_2), \tag{6.17}$$

Assumption 3.1 and the condition $E\|X_{t-1}\|^2(|q_{t-1}|^{2\kappa}+1) < \infty$ can guarantee the existence of ω_2. By (3.2), assumption 2.4, Lemma 6.1(i) and ergodic theorem,

$$-\frac{1}{n}\frac{\partial^2 L(0,\hat{\lambda}_n)}{\partial \delta^2} \to_p E\{\theta_{20}'X_{t-1}X_{t-1}'\theta_{20}D_t^2(r_0,s_0)\} = \omega_1. \tag{6.18}$$

By (3.3), (6.17), (6.18), $\hat{\sigma}_{0n}^2 \to_p \sigma^2$, $\hat{\omega}_{1n} \to_p \omega_1$, $\hat{\omega}_{2n} \to_p \omega_2$ and Slusky theorem, we have

$$\frac{T_{1n}}{\hat{\sigma}_{0n}^2}\frac{\hat{\omega}_{1n}}{\hat{\omega}_{2n}} \longrightarrow_{\mathcal{L}} \chi_1^2,$$

as $n \to \infty$. This completes the proof. $\square$

Proof of Theorem 3.2. By a similar argument as above, for a fixed $s \in [1/\bar{s}, \bar{s}]$, we replace $\varepsilon_t(\hat{\lambda}_n)$ with $\varepsilon_t(\hat{\theta}_n, \hat{r}_n)$ and take the derivatives with respect to θ in (6.9), $\partial\varepsilon_t(\theta, \hat{r}_n)/\partial\theta'$ does not depend on θ anymore. Denote $V_t(r) = \partial\varepsilon_t(\theta, r)/\partial\theta$. By assumption 2.5, $\hat{r}_n - r_0 = O_p(1/n)$, then, by (6.4) and the uniform boundedness of $D_t(r, s)$, it is not hard to show that,

$$\sup_{s\in[1/\bar{s},\bar{s}]} |\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n}D_t(\hat{r}_n, s)[\varepsilon_t(\theta_0, \hat{r}_n) - \varepsilon_t]| = o_p(1).$$

Then, for each $s \in [1/\bar{s}, \bar{s}]$, it follows that

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial L(1, \hat{\theta}_n, s, \hat{r}_n)}{\partial\delta} =& -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n}D_t(\hat{r}_n, s)\varepsilon_t \\ &- [\frac{1}{n}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n}D_t(\hat{r}_n, s)V_t(\hat{r}_n)']\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1), \qquad (6.19)\end{aligned}$$

where $o_p(1)$ holds uniformly in $s \in [1/\bar{s}, \bar{s}]$, as $n \to \infty$.

Now, we look at the first term on the right-hand side of (6.19). Let $\zeta = (\theta_2', r)'$ and $g_t(\zeta, s) = X_{t-1}'\theta_2 G_t(q_{t-1}, s, r)$. Then,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n}G(q_{t-1}, s, \hat{r}_n)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_t(\zeta_0, s)\varepsilon_t + [\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t(\zeta_n^*, s)}{\partial\zeta'}\varepsilon_t]\sqrt{n}(\hat{\zeta}_n - \zeta_0) \qquad (6.20)$$

where ζ_n^* lies between $\hat{\zeta}_n$ and ζ_0, and

$$\frac{\partial g_t(\zeta_n^*, s)}{\partial\zeta'} = (X_{t-1}'G(q_{t-1}, s, r_n^*), X_{t-1}'\theta_{2n}^*\frac{\partial G(q_{t-1}, s, r_n^*)}{\partial r}).$$

By assumption 3.1, we can show that for any s, $\tau \in [1/\bar{s}, \bar{s}]$,

$$\begin{aligned}|\frac{\partial g_t(\zeta_n^*, s)}{\partial\zeta'} - \frac{\partial g_t(\zeta_n^*, \tau)}{\partial\zeta'}| \leq& K(|X_{t-1}'|(|q_{t-1}|^{\alpha_1} + 1), \|X_{t-1}\|(|q_{t-1}|^{\alpha_4} + 1))|s - \tau| \\ :=& J_t|s - \tau|, \qquad (6.21)\end{aligned}$$

where J_t is strictly stationary and ergodic. Denote $\Delta(\eta) = \{(\theta_2, r) : \|\theta_2 - \theta_0\| + |r - r_0| \leq \eta\}$. By (6.21), a standard piecewise argument on $s \in [1/\bar{s}, \bar{s}]$ and Lemma 6.1(i), we can show that

$$\sup_{s\in[1/\bar{s},\bar{s}]}\sup_{\Delta(\eta)} |\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t(\zeta, s)}{\partial\zeta'}\varepsilon_t - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t(\zeta_0, s)}{\partial\zeta'}\varepsilon_t| = o_p(1), \qquad (6.22)$$

as η small enough. By ergodic theorem, (6.21) and a standard piecewise argument as Lemma A.1 in Francq et al. (2010)

$$\sup_{s\in[1/\bar{s},\bar{s}]} |\frac{1}{n}\sum_{t=1}^{n} \frac{\partial g_t(\zeta_0, s)}{\partial \zeta'}\varepsilon_t| = o_p(1). \tag{6.23}$$

By assumption 2.5, (6.22) and (6.23), it follows that

$$\sup_{s\in[1/\bar{s},\bar{s}]} |\frac{1}{n}\sum_{t=1}^{n} \frac{\partial g_t(\zeta_n^*, s)}{\partial \zeta'}\varepsilon_t| = o_p(1). \tag{6.24}$$

By assumption 2.5, (6.4) and a similar argument as (6.14), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n} I(q_{t-1} > \hat{r}_n)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20} I(q_{t-1} > r_0)\varepsilon_t + o_p(1). \tag{6.25}$$

By (6.20) and (6.24)-(6.25), it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n} D_t(\hat{r}_n, s)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20} D_t(r_0, s)\varepsilon_t + o_p(1), \tag{6.26}$$

where $o_p(1)$ holds uniformly in $s \in [1/\bar{s}, \bar{s}]$.

We then consider the second term on the right-hand side of (6.19). Let $B_t(\theta_2, r, s) = X_{t-1}'\theta_2 D_t(r, s)V(r)'$. By assumption 3.1, for any s, $\tau \in [1/\bar{s}, \bar{s}]$, and each θ_2 and r, by Taylor's expansion, we have

$$|B_t(\theta_2, r, s) - B_t(\theta_2, r, \tau)|^2 \leq K|X_{t-1}'\theta_2 V_t(r)'|(|q_{t-1}|^{\alpha_1} + 1)|s - \tau| = Q_t|s - \tau|. \tag{6.27}$$

where Q_t is strictly stationary and ergodic.

By Lemma 6.1(i), a standard piecewise argument on $s \in [1/\bar{s}, \bar{s}]$ and (6.27), we can show that for any $\epsilon > 0$, there exits an $\eta > 0$ such that

$$\lim_{n\to\infty} P(\sup_{s\in[1/\bar{s},\bar{s}]} \sup_{\Delta(\eta)} \frac{1}{n}|\sum_{t=1}^{n}[B_t(\theta_2, r, s) - B_t(\theta_{20}, r_0, s)]| \geq \epsilon) = 0. \tag{6.28}$$

By assumption 2.5, (6.26) and (6.28), (6.19) reduces to

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial L(1, \hat{\theta}_n, s, \hat{r}_n)}{\partial \delta} =& -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20} D_t(r_0, s)\varepsilon_t \\
&+ [\frac{1}{n}\sum_{t=1}^{n} X_{t-1}'\theta_{20} D_t(r_0, s)V_t(r_0)']\Sigma_2^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} V_t(r_0)\varepsilon_t + o_p(1) \\
\triangleq& u_{1n}(s) + u_{2n}(s) + o_p(1).
\end{aligned} \tag{6.29}$$

where $o_p(1)$ holds uniformly in $s \in [1/\bar{s}, \bar{s}]$.

To prove (a), first, we prove the convergence of the finite-dimensional distributions. Note that the sequence in (6.29) are square-integrable stationary martingale difference. The conclusion follows from the central limit theorem of Billingsley (1961),

Then, we show that the sequence is tight. By the independence between ε_t and X_{t-1}, and assumption 3.1, for some $\tilde{s}_1$, $\tilde{s}_2$ between s and τ in $[1/\bar{s}, \bar{s}]$, we have,

$$\begin{aligned} E[u_{1n}(s) - u_{1n}(\tau)]^2 =& E(X_{t-1}\theta_{20})^2(\frac{\partial G(q_{t-1}, \tilde{s}_1, r_0)}{\partial s})^2(s-\tau)^2\sigma^2 \\ &\leq K^2 E(X_{t-1}\theta_{20})^2(|q_{t-1}|^{\alpha_1}+1)^2(s-\tau)^2\sigma^2 \\ &\leq K(s-\tau)^2 \end{aligned} \tag{6.30}$$

and

$$\begin{aligned} E[u_{2n}(s) - u_{2n}(\tau)]^2 =& E\left\{[\frac{1}{n}\sum_{t=1}^{n} X'_{t-1}\theta_{20}\frac{\partial G(q_{t-1}, \tilde{s}_2, r_0)}{\partial s}V_t(r_0)']\Sigma_2^{-1}[\frac{1}{n}\sum_{t=1}^{n} X'_{t-1}\theta_{20}\right. \\ &\left.\times\frac{\partial G(q_{t-1}, \tilde{s}_2, r_0)}{\partial s}V_t(r_0)]\right\}(s-\tau)^2\sigma^2. \\ \leq& K(s-\tau)^2\sigma^2, \end{aligned} \tag{6.31}$$

where (6.31) holds by assumption 3.1(ii) and ergodic theorem. The existence of the expectations can be guaranteed by $E\|X_{t-1}\|^2(|q_{t-1}|^{2\alpha_1}+1) < \infty$.

By (6.30) and (6.31), the tightness follows from Theorem 12.3 of Billingsley (1968). By central limit theorem and ergodic theorem, the form of the limiting Gaussian process follows immediately from (6.29). Thus, (a) holds.

To prove (b), by (3.5), let

$$Z_t(\theta_2, r, s) = \theta_2' X_{t-1} X_{t-1} \theta_2' D_t^2(r, s).$$

Then, by Taylor's expansion and for some $\tilde{s}_3 \in [\tau, s]$,

$$\begin{aligned} |Z_t(\theta_2, r, s) - Z_t(\theta_2, r, \tau)| =& 2|\theta_2' X_{t-1} X'_{t-1}\theta_2 D_t(r, s)||\frac{\partial G(q_{t-1}, \tilde{s}_3, r)}{\partial s}||s-\tau| \\ \leq& 2K|\theta_2' X_{t-1} X'_{t-1}\theta_2|(|q_{t-1}|^{\alpha_1}+1)|s-\tau| \\ \triangleq& A_t(\theta_2)|s-\tau|, \end{aligned} \tag{6.32}$$

where $A_t(\theta_2)$ is strictly stationary and ergodic. Then, by (6.32), Lemma 6.1(i) and a standard piecewise argument on $s \in [1/\bar{s}, \bar{s}]$, it is not hard to show that, for any $\epsilon > 0$, there exists an $\eta > 0$ such that

$$\lim_{n\to\infty} P(\sup_{s\in[1/\bar{s},\bar{s}]} \sup_{\Delta(\eta)} \frac{1}{n}|\sum_{t=1}^{n}[Z_t(\theta_2, r, s) - Z_t(\theta_{20}, r_0, s)]| \geq \epsilon) = 0. \tag{6.33}$$

By (6.32), ergodic theorem and a similar standard piecewise argument again on $s \in [1/\bar{s}, \bar{s}]$ or Lemma A.1 in Francq et al. (2010), we can show that

$$\sup_{s\in[1/\bar{s},\bar{s}]} |\frac{1}{n}\sum_{t=1}^{n} Z_t(\theta_{20}, r_0, s) - \omega(s)| = o_p(1), \tag{6.34}$$

where $\omega(s)$ is defined in Theorem 3.2. By assumption 2.5, (b) follows from (6.33) and (6.34). This completes the proof. □

References

Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, **58**, 525–534.

Bera, A. K. and McAleer, M. (1989). Nested and non-nested procedures for testing linear and log-linear regression models. *Sankhyā, Series B*, **50**, 212–224.

Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proc. Amer. Math. Soc.*, **12(5)**, 788–792.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

Chan, K. S. (1990). Testing for Threshold Autoregression. *Ann. Stat.*, **18**, 1886–1894.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Stat.*, **21**, 520–533.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time Series Anal.*, **7**, 179–190.

Chan, K. S. and Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *J. Roy. Stat. Soc.* **B52**, 469–476.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, **1**, 105–123.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Stat. Soc.*, **B24**, 406–424.

Cox, D. R. (2013). A return to an old paper: tests of separate families of hypotheses. *J. Roy. Stat. Soc.*, **B75**, 207–215.

Davidson, R. and MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, **49**, 781–793.

Ekner, L. E. and Nejstgaard, E. (2013). Parameter identification in the logistic star model. *Social Science Research Network, SSRN 2330263.*

Francq, C. , Horváth, L. and Zakoïan, J-M. Sup-tests for linearity in a general nonlinear AR(1) model. *Econometric Th.*, **26**, 965–993

Ghaddar, D. K. and Tong, H. (1981). Data transformation and self-exciting threshold autoregression. *J. Roy. Stat. Soc.*, **C30**, 238–248.

Hansen B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, **64**, 413–430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, **68**, 575–603.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Stat.*, **6**, 629–642.

Ling, S. and McAleer, M. (2010). A general asymptotic theory for time-series models. *Stat. Neerl.*, **64(1)**, 97–111.

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

MacKinnon, J. G., White, H., and Davidson, R. (1983). Tests for model specification in the presence of alternative hypotheses: Some further results. *J. Econometrics*, **21**, 53–70.

McAleer, M. (1995). The significance of testing empirical non-nested models. *J. Econometrics*, **67**, 149–171.

Ozaki, T. (1980). Non-linear time series models for non-linear random vibrations. *J. App. Prob.*, **17**, 84–93.

Pesaran, M. H. and Weeks, M. (2001). Non-nested hypothesis testing: an overview. *A companion to theoretical econometrics*, ed. B. H. Baltagi, Blackwell Pub., 279–309.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *J. Amer. Stat. Ass.*, **89**, 208–218.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. (2010). *Modelling nonlinear economic time series*. Oxford: Oxford Univ. Press.

Tong, H. (1978). On a threshold model. In *Patterm recognition and signal processing,* ed. C.H. Chen, 575–586. The Netherlands: Sijthoff & Noordhoff.

Tong, H. (2011). Threshold models in time series analysis–30 years on. *Stat. & its Interface*, *4*, 107–118.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Stat. Soc.*, **B42**, 245–292.

van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002). Smooth transition autoregressive models-a survey of recent developments. *Econometric Rev.*, **21**, 1–47.

Current Departmental Research Reports

Serial No.	Date	Research Report Title	Author(s)
503	Apr-13	A geometric process credibility model	Yeh Lam
504	Jun-13	On mixture double autoregressive time series models	Zhao Liu and Guodong Li
505	Jun-13	A simple formula for mixing estimators with different convergence rates	Stephen M.S. Lee and Mehdi Soleymani
506	Nov-13	On singular value distribution of large-dimensional autocovariance matrices	Zeng Li, Guangming Pan and Jianfeng Yao
507	Nov-13	Threshold models in time series analysis-some reflections	Howell Tong
508	Nov-13	An essay on statistics and dynamical phenomena to high school children	Howell Tong
509	Dec-13	Exact distribution of spurious correlations, Minimax Sample Size and direct selection of predictors	Kai Wang Ng
510	Dec-13	Nested sub-sample search algorithm for estimation of threshold models	Dong Li and Howell Tong
511	Jan-14	New procedure for high dimensional classification of general populations	Zhaoyuan Li and Jianfeng Yao
512	Mar-14	On buffered threshold GARCH models	Pak Hang Lo, Wai Keung Li, Philip L.H. Yu and Guodong Li
513	Nov-14	Optimal proportional reinsurance with common shock dependence	Kam Chuen Yuen, Zhibin Liang and Ming Zhou
514	May-15	Resampling-based post-model-selection inference for linear regression models	Stephen M.S. Lee and Yilei Wu
515	Sep-15	Tests for TAR models vs. STAR models - a separate family of hypotheses approach	Zhaoxing Gao, Shiqing Ling and Howell Tong

❖❖❖ *The complete listing can be found at* ❖❖❖
http://www.saasweb.hku.hk/research/staff_research_report.php
Requests for off prints may be sent to saas@hku.hk by e-mail