Statistica Sinica Preprint No: SS-2016-0497.R1

Title	TESTS FOR TAR MODELS VS. STAR MODELS–A SEPARATE FAMILY OF HYPOTHESES APPROACH
Manuscript ID	SS-2016-0497.R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0497
Complete List of Authors	Zhaoxing Gao Shiqing Ling and Howell Tong
Corresponding Author	Zhaoxing Gao
E-mail	Zhaoxing Gao

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

Zhaoxing Gao[1], Shiqing Ling[2] and Howell Tong[3]

[1]*Booth School of Business, University of Chicago*

[2]*Hong Kong University of Science and Technology*

[3]*University of Electronic Science and Technology of China and London School of Economics*

Abstract: The threshold autoregressive (TAR) model and the smooth threshold autoregressive (STAR) model have been popular parametric nonlinear time series models for the past three decades or so. 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 is smooth (except in the limit of infinitely fast switching for some cases). Following the approach initiated by Cox (1961, 1962), we treat the test problem as one of separate families of hypotheses. The test statistic under a STAR model is shown to follow asymptotically a chi-squared distribution, and the one under a TAR model can be 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 and phrases: Non-nested test, Separate family of hypotheses, STAR model, TAR model.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

1 Introduction

Regime switching models are a central area of research activities in time series analysis in both the statistical and the econometric literature. 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 by Tong (1978) and Tong and Lim (1980); see also Tong (2011). In the non-time series context, the idea of smooth regime switching was first introduced by Bacon and Watts (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 Luukkonen et al. (1988), Teräsvirta (1994), van Dijk et al. (2002), and Teräsvirta et al. (2010). They 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 the 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 vs. discontinuous hypotheses. As far as we are aware, our paper represents the first attempt at testing for separate families of

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

hypotheses in nonlinear time series analysis. However, there is an interesting challenge in this. Although the STAR model includes the TAR model as a special case for many smooth functions, 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 follows a complicated distribution, which is typically not a chi-squared distribution. To develop a test that has sufficient power and is simple to use in practice, we 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 develop non-nested tests for departure from a STAR/TAR model in the direction of a TAR/STAR model, within the context of 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 that the one under a TAR model can be 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 score tests and the related algorithm. Section 4 gives the asymptotic local power analysis. Section 5 presents a simulation study. Section 6 analyzes two empirical examples. Section 7 provides the proofs of the theorems. In the supplementary material, we give a discussion on some nested hypothesis testing approaches, and report some simulation results to make a comparison with

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

our proposed tests. The proofs of Theorems 4.1-4.2 and some related tables are also given in the supplementary material.

2 The Models and the Testing Procedure

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

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

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 ε_t is independent of $\mathcal{F}_{t-1}$. $G(q_{t-1}, s, r)$ is a smooth switching function, for example, the logistic smooth switching function

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

is a popular choice. Model (2.1) with logistic smooth switching function (2.2) is commonly called an LSTAR model. There are other smooth switching functions in the literature such as the normal distribution function in Chan and Tong (1986), the exponential STAR (ESTAR) models with

$$G(q_{t-1}, s, r) = 1 - e^{-s(q_{t-1}-r)^2},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

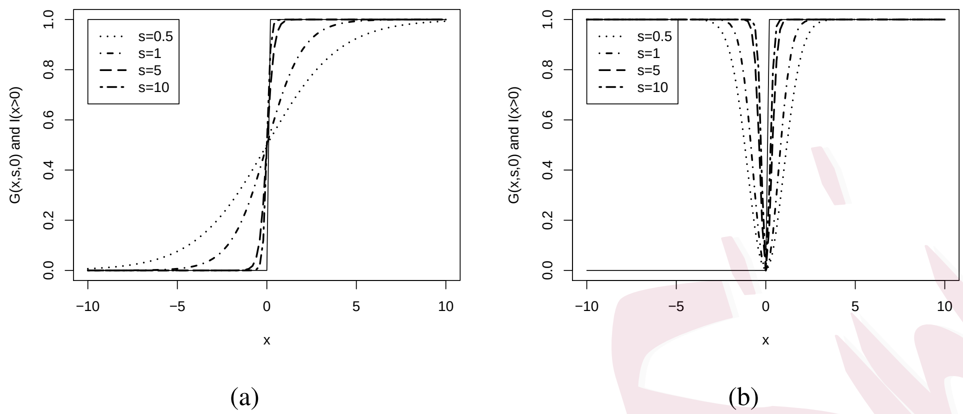

Figure 1: (a). $I(x > 0)$ and $G(x, s, 0) = 1/(1 + e^{-sx})$; (b). $I(x > 0)$ and $G(x, s, 0) = 1 - e^{-sx^2}$.

and the second order logistic smooth function; see van Dijk et al. (2002) for details.

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

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

where $I(\cdot)$ is the indicator function. Figure 1 plots $I(x > 0)$ and $G(x, s, 0)$ of the logistic and the exponential ones 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, especially for the logistic functions. Standard practice in STAR modeling restricts s to lie in a finite interval, namely $s \in [s_1, s_2]$ with $0 < s_1 < s_2 < \infty$. A similar restriction is assumed for s in the general $G(q_{t-1}, s, r)$. Note that a STAR model has one more parameter, s,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

than a TAR model of the same order.

Model (2.1) (under the restriction on s) and model (2.3) are two non-nested models. Testing for non-nested models has been studied in the literature, starting from Cox (1961, 1962). See also Cox (2013). In the econometric literature, Pesaran and Deaton (1978) proposed the Cox-Pesaran-Deaton (CPD) test, but the power of the CPD test is not clear in theory. Another approach to test non-nested models is to form a compound model as in Atkinson (1970), and treat the problem as one of testing model specification. This approach was further developed by Davidson and MacKinnon (1981) for the non-nested regression models; see also MacKinnon et al. (1983). From models (2.1) and (2.3), we can construct a compound model as

$$y_t = X'_{t-1}\theta_1 + (1-\delta)X'_{t-1}\theta_2 G(q_{t-1}, s, r) + \delta X'_{t-1}\theta_2 I(q_{t-1} > r) + \varepsilon_t. \tag{2.4}$$

Unlike Davidson and MacKinnon (1981), the 'slope' parameters (θ_2) in the smooth part and in the discontinuous part are the same. This is because their estimates tend to be very close whether we fit a TAR model or a STAR model to given data (-see Ekner and Nejstgaard (2013)). Based on model (2.4), we consider the following two hypotheses:

$$H_0 : \delta = 0 \text{ against } H_a : \delta \neq 0 \tag{2.5}$$

and

$$\tilde{H}_0 : \delta = 1 \text{ against } \tilde{H}_a : \delta \neq 1. \tag{2.6}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

Hypothesis (2.5) tests the departure of a STAR model in the direction of model (2.4) where $\delta \neq 0$, while hypothesis (2.6) tests the departure of a TAR model in the direction of model (2.4) where $\delta \neq 1$.

We will study the score tests for (2.5) and (2.6) in the next section. Let $\theta = (\theta_1', \theta_2')'$ and $\lambda = (\theta', s, r)'$, and 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 some assumptions.

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

For this, see the discussions in Chan and Tong (1986) for the STAR model and Chan (1993) for the TAR model.

Assumption 2.2. (i) ε_t and q_t have absolutely continuous distributions with uniformly continuous and positive densities on R and $E\varepsilon_t^4 < \infty$; (ii) The conditional density of X_t given $q_t = r$, $f_{X|q}(x|r)$, is 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 $E\varepsilon_t^4 < \infty$ 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 (S2.2) in the supplementary material.

Assumption 2.3. (i) $E(\|X_t\|^2 | q_t = r) \leq K < \infty$ for all $r \in \Gamma$; (ii) $E(\|X_t\|^2 I(r_1 < q_t \leq r_2) | \mathcal{F}_{t-p}) \leq K \wp_{t-p} |r_2 - r_1|$, 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), but we only require a 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. For most smooth transition functions, a second moment is enough to satisfy Lemma 7.1 and the functions of interest in the proofs of Theorems 3.1 and 3.2, including the LSTAR and ESTAR models. 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

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

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 inequality above is due to Taylor's expansion and the boundedness of the density function of ε_t by Assumption 2.2. Let

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

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

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),

$$\hat{\lambda}_n = \arg\min_{\lambda \in \Lambda} \sum_{t=1}^{n} \varepsilon_t^2(\lambda), \tag{2.7}$$

$$(\hat{\theta}_n, \hat{r}_n) = \arg\min_{(\theta,r) \in \Theta \times \Gamma} \sum_{t=1}^{n} \varepsilon_t^2(\theta, r). \tag{2.8}$$

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 the STAR model. For general conditions, see 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 $\Sigma_2 = E[\partial \varepsilon_t(\theta_0, r_0)/\partial \theta \partial \varepsilon_t(\theta_0, r_0)/\partial \theta']$.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

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.

3 Asymptotic Properties of Score Tests

Consider the (conditional) quasi-log-likelihood function of model (2.4),

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

Let $D_t(r, s) = G(q_{t-1}, s, r) - I(q_{t-1} > r)$. We first consider the hypothesis (2.5), under H_0 (i.e. $\delta = 0$), we obtain the score function and information matrix as follows.

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

and

$$\frac{\partial^2 L(0,\lambda)}{\partial^2\delta} = -\sum_{t=1}^{n}\theta'_2 X_{t-1}X'_{t-1}\theta_2 D_t^2(r, s). \qquad (3.2)$$

The score based test statistic for testing H_0 is defined as

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

where $\hat{\lambda}_n$ is defined in (2.7).

Assumption 3.1.

$(i).\ |G(q_{t-1}, s, r)| \leq 1;$

$(ii).\ |\dfrac{\partial G(q_{t-1}, s, r)}{\partial s}| \leq K(|q_{t-1}|^{\alpha_1} + 1)$ and $|\dfrac{\partial G(q_{t-1}, s, r)}{\partial r}| \leq K(|q_{t-1}|^{\alpha_2} + 1);$

$(iii).\ |\dfrac{\partial^2 G(q_{t-1}, s, r)}{\partial^2 s}| \leq K(|q_{t-1}|^{\alpha_3} + 1)$ and $|\dfrac{\partial^2 G(q_{t-1}, s, r)}{\partial^2 r}| \leq K(|q_{t-1}|^{\alpha_4} + 1);$

$(iv).\ |\dfrac{\partial^2 G(q_{t-1}, s, r)}{\partial r \partial s}| \leq K(|q_{t-1}|^{\alpha} + 1),$

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). 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 (7.6). 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 $\alpha = 1$.

Let

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

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

with their estimators

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

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

respectively, where $\hat{\Sigma}_{1n} = \sum_{t=1}^{n}[\partial\varepsilon_t(\hat{\lambda}_n)/\partial\lambda\partial\varepsilon_t(\hat{\lambda}_n)/\partial\lambda']/n$. 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.

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.

Under H_0, we need to specify the interval $[s_1, s_2]$ for grid search to give an estimator $\hat{s}_n$. van Dijk et al. (2002) (pp. 21) also discussed this issue without giving a recommended interval. In the absence of theoretical results, we can either follow the suggestion of Di Narzo et al. (2013), adopting a default interval $s \in [1, 40]$, or choose other intervals according to simulation experience, when using `lstar` function in R to fit an LSTAR model.

Next, we consider the hypothesis (2.6). We fix $s > 0$ as a constant in (2.1). Under $\tilde{H}_0$, we

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

obtain the score function and information matrix as follows.

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

and

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

For a given $s>0$, the score-based test statistic for testing $\tilde{H}_0$ against $\tilde{H}_a$ is

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

where $\hat{\theta}_n$ and $\hat{r}_n$ are defined in (2.8). In (3.6), we have a nuisance parameter s, which is not identified under $\tilde{H}_0$. In the spirit of Francq et al. (2010), 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 $\Longrightarrow$ denote the weak convergence.

Theorem 3.2. *Under $\tilde{H}_0$, 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,*

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

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 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.

Under $\tilde{H}_0$, we also need to specify the form of the smooth function G and different G's may give different power.

Following Hansen (1996) and Francq et al. (2010), among others, we use the supremum statistic $\sup_{s \in [1/\bar{s}, \bar{s}]} T_{2n}(s)/\hat{\sigma}_{1n}^2$ as our test statistic, where $\hat{\sigma}_{1n}^2 = -2L(1, \hat{\theta}_n, s, \hat{r}_n)/n$, 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 $\tilde{H}_0$. By Theorem 3.2 and the Continuous Mapping Theorem, it follows that

$$S_{2n} := \sup_{s \in [1/\bar{s}, \bar{s}]} \frac{T_{2n}(s)}{\hat{\sigma}_{1n}^2} \longrightarrow_{\mathcal{L}} \sup_{s \in [1/\bar{s}, \bar{s}]} \frac{Z^2(s)}{\omega(s)}, \tag{3.7}$$

which is the limiting distribution of our test statistic. Following Hansen (1996), Francq et al. (2010), and using (7.4), (7.23) 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 an i.i.d. $N(0, 1)$ sample $\varepsilon_1^{(i)}, ..., \varepsilon_n^{(i)}$;

(ii) set

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

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

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

Here $\hat{\Sigma}_{2n} = \sum_{t=p+1}^{n}[\partial\varepsilon_t(\hat{\theta}_n,\hat{r}_n)/\partial\theta\partial\varepsilon_t(\hat{\theta}_n,\hat{r}_n)/\partial\theta']/n$. Conditional on $\{y_1,...,y_n\}$, the sequence $\{S^{(i)}, i=1,...,N\}$ constitutes an independent and identically distributed sample of the random variable $\sup_{s\in[1/\bar{s},\bar{s}]}\frac{T_{2n}(s)}{\hat{\sigma}_{1n}^2}$. 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

$$\left\{\sup_{s\in[1/\bar{s},\bar{s}]}\frac{T_{2n}(s)}{\hat{\sigma}_{1n}^2} > c_\alpha\right\}.$$

The limiting distribution in (3.7) depends on the data and the simulated distribution by Algorithm 1 only converges exactly to the limiting one in (3.7) under $\tilde{H}_0$, e.g. the data come from a TAR model. When the data come from an LSTAR one, we are not clear about the limiting behavior of the estimators and hence Algorithm 1 does not necessarily converge to the exact one in (3.7). However, when the data come from LSTAR models, the power is still satisfactory, as can be seen

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

from our empirical results in Section 5.

As for the choice of $\bar{s}$, we are not aware of any definitive guidance in the statistical literature concerning this issue in a general context. See, e.g., Chan (1990) for threshold problems and Davis et al. (1995) for change-point problems, among others. Following the reference manual of `lstar` function in Di Narzo et al. (2013), we also recommend a default $\bar{s} = 40$ in practice according to our simulation experience.

4 Asymptotic Power Under Local Alternatives

This section investigates the asymptotic local power of S_{1n} and S_{2n} defined in Theorem 3.1 and (3.7), respectively. We consider the hypotheses,

$$H_0 : \delta = 0 \quad \text{against} \quad H_{a_n} : \delta = \frac{\gamma}{\sqrt{n}} \quad \text{for some fixed} \quad \gamma \neq 0, \tag{4.1}$$

$$\tilde{H}_0 : \delta = 1 \quad \text{against} \quad \tilde{H}_{a_n} : \delta = 1 + \frac{\gamma}{\sqrt{n}} \quad \text{for some fixed} \quad \gamma \neq 0, \tag{4.2}$$

where (4.1) and (4.2) correspond to (2.5) and (2.6), respectively. For model (2.4), we define

$$\varepsilon(\lambda, \delta) = y_t - X'_{t-1}\theta_1 - (1-\delta)X'_{t-1}\theta_2 G(q_{t-1}, s, r) - \delta X'_{t-1}\theta_2 I(q_{t-1} > r). \tag{4.3}$$

Let $\mathcal{F}^Z$ be the Borel $\sigma-$field on R^Z with $Z = \{0, \pm 1, \pm 2, ...\}$ and P a probability measure on $(R^Z, \mathcal{F}^Z)$. Let $P^n_{\lambda,\delta}$ be the restriction of P on $\mathcal{F}_n$, the $\sigma-$field generated by $\{Y_0, y_1, ..., y_n\}$ with

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

$Y_0 = \{y_0, y_{-1}, ..., y_{1-p}\}$. Suppose the errors $\{\varepsilon_1(\lambda,\delta), \varepsilon_2(\lambda,\delta), ...\}$ under $P^n_{\lambda,\delta}$ are i.i.d. with density g, and are independent of Y_0. The log-likelihood ratio $\Lambda_{n,\lambda}(\delta_1,\delta_2)$ of P^n_{λ,δ_2} to P^n_{λ,δ_1} is then

$$\Lambda_{n,\lambda}(\delta_1,\delta_2) = \sum_{t=1}^{n}[\log g(\varepsilon_t(\lambda,\delta_2)) - \log g(\varepsilon_t(\lambda,\delta_1))].$$

For simplicity, we assume $\varepsilon_t(\lambda_0,\delta_0) \sim N(0,\sigma^2)$ in the rest of this section. This can be generalized to the non-normal case without difficulty; see, for example, Jeganathan (1995). Thus, the density g of ε_t is absolutely continuous with derivatives and finite Fisher information $0 < I(g) = \int_{-\infty}^{+\infty}[g'(x)/g(x)]^2 g(x)dx < \infty$. In this section, all the expectations are taken under H_0 or $\tilde{H}_0$ according to the context.

Theorem 4.1. *If Assumptions 2.1-2.5 and 3.1 hold, then $P^n_{\lambda_0,\delta_0+\gamma/\sqrt{n}}$ is contiguous to $P^n_{\lambda_0,\delta_0}$, where $\delta_0 = 0$ or 1.*

Theorem 4.2. *Suppose that Assumptions 2.1-2.5 and 3.1 hold.*

(i) Under H_{a_n}, if the conditions in Theorem 3.1 are satisfied, we have

$$S_{1n} \to_{\mathcal{L}} \chi_1^2\left(\frac{\gamma\sqrt{\omega_2}}{\sigma}\right); \tag{4.4}$$

(ii) Under $\tilde{H}_{a_n}$, if the conditions in Theorem 3.2 are satisfied, we have

$$S_{2n} \longrightarrow_{\mathcal{L}} \sup_{s\in[1/\bar{s},\bar{s}]} \frac{[Z(s)+\sigma^{-1}\mu(s)]^2}{\omega(s)}, \tag{4.5}$$

as $n \to \infty$, where S_{1n} and ω_2 are defined as in Theorem 3.1, $\chi_1^2(\frac{\gamma\sqrt{\omega_2}}{\sigma})$ is a non-central chi-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

squared distribution with mean $1+\frac{\gamma\sqrt{\omega_2}}{\sigma}$*;* S_{2n}*,* $\omega(s)$ *and* $Z(s)$ *are defined as in (3.7) and Theorem 3.2, and* $\mu(s) = \gamma E[Z(s)Z(s_0)]$ *for some* s_0 *which is specified under* $\tilde{H}_{a_n}$*.*

5 Simulation Studies

We examined the performance of the statistic S_{1n} and S_{2n} in finite samples through Monte Carlo experiments. In the experiments, we used the logistic smooth functions in (2.2). Similar results can be obtained from others. The sample sizes (n) were 400, 800, 1500, 3000, and 5000, and the number of replications was 500 for each case. The null hypothesis H_0 was the LSTAR(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 was given by (2.2) with $q_{t-1} = y_{t-1}$. The null hypothesis $\tilde{H}_0$ was 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 used the package `tsDyn` in `R` software and `lstar` function to fit the logistic STAR model when testing H_0. From Table 1, 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, one requires a sample size in excess of 1500 for decent power. The results are summarized in Table 1.

When testing $\tilde{H}_0$, we set $\bar{s} = 15$, 30, and 45 in (3.7). We first simulated the critical values by Algorithm 1 with $N = 10000$ and they are reported in Tables S12-S13 in the supplementary material. Based on the critical values in Table S13, we used 500 replications in this experiment

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

Table 1: Empirical size and power for testing H_0.

			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

for each case and Tables 2–4 report the sizes and powers when testing $\tilde{H}_0$ for $\bar{s} = 15$, 30, and 45, respectively. From Tables 2–4, the sizes are very close to their nominal levels, and the power increases with the sample size. We plot the power against different values of s_0 in Figure 2 for $\bar{s} = 15$. Similar patterns can be found for other $\bar{s}$. For each $\bar{s}$, the power is initially lower when $s_0 = 1$, 2 than 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 $n \geq 3000$. When $\bar{s}$ becomes larger, the power seems to decrease slightly at each corresponding slot. Moreover, Tables 2–4 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 5 shows the relation when $n = 400$, providing the mean of 500 estimators for each s_0. In view of Figure 1, a larger estimator $\tilde{s}_n$ gives rise to less difference between the smooth function and the indicator function and hence a lower power, and a smaller

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

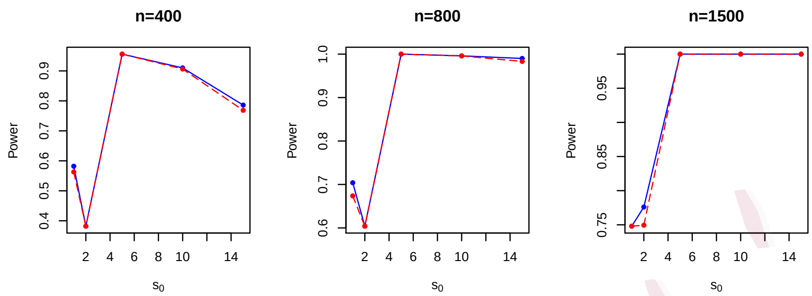

Figure 2: Power for testing $\tilde{H}_0$ with $\bar{s} = 15$ for different values of s_0. The solid line denotes the power at level $\alpha = 0.1$ and the dotted line at level $\alpha = 0.05$.

one gives higher power. The result in Table 5 conforms to the ones we obtained in Tables 2–4.

Table 2: Empirical size and power for testing $\tilde{H}_0$ when $\bar{s} = 15$.

				n				
	data	s_0	α	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

We provide some additional simulation results when we increased the number of parameters. The null hypothesis H_0 was the LSTAR(5) model with $(\theta_0', r_0) = (-1, -0.4, -0.8, -0.1,$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

Table 3: Empirical size and power for testing $\tilde{H}_0$ 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

$0.2, 0.2, 2, 0.9, 0.4, 0.3, -0.2, -0.2, 0.8)$ and the null hypothesis $\tilde{H}_0$ is a TAR(5) with the same parameters. We only report the results of the empirical power in Tables 6-7 for testing H_0 and $\tilde{H}_0$, respectively, since the size is not of interest. From Table 6, the power is already quite satisfactory when the sample size is small (e.g. $n = 300$). For $n = 400$, 800, and 1500, the power is higher than the corresponding one in Table 1. In Table 7, we only report the results with $\bar{s} = 15$ since it is similar for the other cases. The power is also higher than the corresponding one in Table 4 for each sample size. The high power of a small sample size in both Tables 6 and 7 suggests that our results in Section 6 are convincing when we have many parameters in the fitted models.

Table 4: Empirical size and power for testing $\tilde{H}_0$ when $\bar{s} = 45$.

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

Table 5: The realized estimator $\tilde{s}_n$ for different true value s_0 under $\tilde{H}_0$ 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

Table 6: Empirical power for testing H_0 with $p = 5$.

	data	α	n 300	400	800	1500
power	TAR(5)	0.1	0.718	0.754	0.834	0.872
		0.05	0.674	0.694	0.782	0.806
		0.01	0.586	0.602	0.684	0.720

Table 7: Empirical size and power for testing $\tilde{H}_0$ when $\bar{s} = 15$ with $p = 5$.

	data	s_0	α	n 300	400	800	1500
power	LSTAR(5)	$s_0 = 1$	0.1	0.65	0.74	0.93	1
			0.05	0.54	0.56	0.87	1
			0.01	0.15	0.28	0.69	1
power	LSTAR(5)	$s_0 = 2$	0.1	0.90	0.99	1	1
			0.05	0.82	0.98	1	1
			0.01	0.56	0.88	1	1

6 Data Examples

We re-visited two real data sets to illustrate our tests. 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 these examples, after an appropriate reparametrization.

The first data set consists of the Wolf's annual sunspot numbers, which are available at `http://www.sidc.oma.be/sunspot-data/`. 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, $y_t = 2\{(1 + z_t)^{1/2} - 1\}$, where z_t is the

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

original sunspot number. Ekner and Nejstgaard (2013) reproduced the LSTAR model, as well as fitted a TAR model, as follows (standard deviations in parentheses):[1]

$$\begin{aligned}
H_0: \quad y_t \quad = \quad & 1.46y_{t-1} - 0.76y_{t-2} + 0.17y_{t-7} + 0.11y_{t-9} \\
& (0.08) \qquad (0.13) \qquad (0.05) \qquad (0.04) \qquad\qquad (6.1) \\
& +(2.65 - 0.54y_{t-1} + 0.75y_{t-2} - 0.47y_{t-3} \\
& \quad (0,85) \ (0.13) \qquad (0.18) \qquad (0.11) \\
& +0.32y_{t-4} - 0.26y_{t-5} - 0.24y_{t-8} + 0.17y_{t-10})\hat{G}(y_{t-2}, 5.46/\hat{\sigma}_{y_{t-2}}, 7.88), \\
& \quad (0.11) \qquad (0.07) \qquad (0.05) \qquad (0.06)
\end{aligned}$$

$$\begin{aligned}
\tilde{H}_0: \quad y_t \quad = \quad & 1.43y_{t-1} - 0.77y_{t-2} + 0.17y_{t-7} + 0.12y_{t-9} \\
& (0.08) \qquad (0.14) \qquad (0.05) \qquad (0.05) \\
& +(2.69 - 0.45y_{t-1} + 0.69y_{t-2} - 0.48y_{t-3} \\
& \quad (0,70) \ (0.11) \qquad (0.18) \qquad (0.11) \qquad\qquad (6.2) \\
& +0.36y_{t-4} - 0.27y_{t-5} - 0.21y_{t-8} + 0.14y_{t-10})I(y_{t-2} > 6.39), \\
& \quad (0.11) \qquad (0.07) \qquad (0.05) \qquad (0.05)
\end{aligned}$$

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., (6.1)), the results are summarized in Table 8. From Table 8, we do not reject (6.1) at each of the three levels and the $p-$value is 0.764. Then we tested under $\tilde{H}_0$ with $\bar{s} = 15$, 30, and 45, respectively. The

[1]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).

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

results are summarized in Table 9. From Table 9, we again do not reject (6.2) 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 (6.1) and 13 for (6.2)), 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), who find 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.

Table 8: Testing (6.1). (NR=not rejected, R=rejected).

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

Table 9: Testing (6.2). (NR=not rejected, R=rejected).

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

As a second example, we re-examined 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. These two series are published together with Gauss programs used to estimate their model at `http://swopec.hhs.se/hastef/abs/hastef0380.htm`. Ekner and Nejstgaard (2013) re-examined this LSTAR model as well as fitted a TAR model

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

as follows (standard deviations in parentheses).

$$
\begin{aligned}
H_0: \quad \Delta y_t \quad = \quad & \underset{(0.07)}{0.479} + \underset{(0.07)}{0.645} D_{1,t} - \underset{(0.10)}{0.342} D_{2,t} - \underset{(0.09)}{0.68} D_{3,t} - \underset{(0.11)}{0.725} D_{4,t} - \underset{(0.10)}{0.649} D_{5,t} \\
& - \underset{(0.09)}{0.317} D_{6,t} - \underset{(0.09)}{0.410} D_{7,t} - \underset{(0.09)}{0.501} D_{8,t} - \underset{(0.09)}{0.554} D_{9,t} - \underset{(0.07)}{0.306} D_{10,t} \\
& + [-\underset{(0.01)}{0.040} y_{t-1} - \underset{(0.08)}{0.146} \Delta y_{t-1} - \underset{(0.06)}{0.101} \Delta y_{t-6} + \underset{(0.06)}{0.097} \Delta y_{t-8} - \underset{(0.06)}{0.123} \Delta y_{t-10} \\
& + \underset{(0.07)}{0.129} \Delta y_{t-13} - \underset{(0.06)}{0.103} \Delta y_{t-15}] \times [1 - \hat{G}(\Delta_{12} y_{t-1}, 23.15/\hat{\sigma}_{\Delta_{12} y_{t-1}}, 0.274)] \\
& + [-\underset{(0.01)}{0.011} y_{t-1} + \underset{(0.08)}{0.225} \Delta y_{t-1} + \underset{(0.08)}{0.307} \Delta y_{t-2} - \underset{(0.07)}{0.119} \Delta y_{t-7} - \underset{(0.09)}{0.155} \Delta y_{t-13} \\
& - \underset{(0.09)}{0.215} \Delta y_{t-14} - \underset{(0.09)}{0.235} \Delta y_{t-15}] \times \hat{G}(\Delta_{12} y_{t-1}, 23.15/\hat{\sigma}_{\Delta_{12} y_{t-1}}, 0.274), \qquad (6.3)
\end{aligned}
$$

$$
\begin{aligned}
\tilde{H}_0: \quad \Delta y_t \;=\; & \underset{(0.07)}{0.473} + \underset{(0.07)}{0.644} D_{1,t} - \underset{(0.10)}{0.343} D_{2,t} - \underset{(0.09)}{0.675} D_{3,t} - \underset{(0.11)}{0.721} D_{4,t} - \underset{(0.10)}{0.641} D_{5,t} \\
& - \underset{(0.09)}{0.308} D_{6,t} - \underset{(0.09)}{0.410} D_{7,t} - \underset{(0.08)}{0.505} D_{8,t} - \underset{(0.09)}{0.546} D_{9,t} - \underset{(0.07)}{0.295} D_{10,t} \\
& + [-\underset{(0.01)}{0.040} y_{t-1} - \underset{(0.08)}{0.14} \Delta y_{t-1} - \underset{(0.06)}{0.094} \Delta y_{t-6} + \underset{(0.06)}{0.092} \Delta y_{t-8} - \underset{(0.06)}{0.116} \Delta y_{t-10} \\
& + \underset{(0.07)}{0.136} \Delta y_{t-13} - \underset{(0.06)}{0.106} \Delta y_{t-15}] \times I(\Delta_{12} y_{t-1} \leq 0.268) \\
& + [-\underset{(0.01)}{0.012} y_{t-1} + \underset{(0.08)}{0.227} \Delta y_{t-1} + \underset{(0.08)}{0.307} \Delta y_{t-2} - \underset{(0.07)}{0.094} \Delta y_{t-7} - \underset{(0.09)}{0.146} \Delta y_{t-13} \\
& - \underset{(0.09)}{0.211} \Delta y_{t-14} - \underset{(0.09)}{0.216} \Delta y_{t-15}] \times I(\Delta_{12} y_{t-1} > 0.268)
\end{aligned} \tag{6.4}
$$

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 obtained $\hat{\sigma}_{\Delta_{12} y_{t-1}} = 1.35$, giving $\hat{s}_n = 17.15$. The results of testing H_0 (i.e., (6.3)) are summarized in Table 10. From Table 10, we reject (6.3) 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 tested under $\tilde{H}_0$ and chose $\bar{s} = 15$, 30 and 45, respectively. The results are summarized in Table 11. From Table 11, we do not reject (6.4) at any of the three levels for each $\bar{s}$, with the $p-$value of 0.99 for each $\bar{s}$. The rejection of the STAR model at 0.1 significance level and no rejection of the TAR model at any of the significance lever could suggest that a TAR model is more plausible, in line with the conclusion by Ekner and Nejstgaard (2013). They found that, for

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

Table 10: Testing (6.3). (NR=not rejected, R=rejected).

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

Table 11: Testing (6.4). (NR=not rejected, R=rejected).

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

the STAR model, the profile likelihood of the s parameter is rather flat and the maximum occurs at a rather large value of s, and the concluded that 'a large and imprecise estimate of s implies that the LSTAR model is effectively a TAR model.'

7 Proofs of Theorems 3.1-3.2

To prove Theorems 3.1 and 3.2, we need a lemma. Its proofs can be found in the supplementary material.

Lemma 7.1. *Let $\{X_t\}$ be a strictly stationary and ergodic process, $f(X_t, \theta)$ be a measurable function with respect to X_t, and $\theta \in \Theta$, a compact set in R^d for some integer $d > 0$.*

(i) If $E \sup_{\theta \in \Theta} |f(X_t, \theta)| < \infty$, $f(X_t, \theta)$ is continuous in θ and satisfies Assumption 2.3, with replacing $\|X_t\|^2$ by $|f(X_t, \theta)|$, then for any $\epsilon > 0$, there exists an $\eta > 0$ such that

$$\lim_{n \to \infty} P\left(\sup_{\substack{\|\theta - \theta_0\| \le \eta \\ |r - r_0| \le \eta}} \frac{1}{n} \Big| \sum_{t=1}^{n} [f(X_t, \theta) I(q_t \le r) - f(X_t, \theta_0) I(q_t \le r_0)] \Big| \ge \epsilon \right) = 0; \qquad (7.1)$$

(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 $q_t \in \mathcal{F}_t^p$, has a bounded, continuous and positive density $f_q(x)$ on R, then for any $\epsilon > 0$ and $\theta_0 \in \Theta$,

$$\lim_{n\to\infty} P(\sup_{0\le r\le \frac{M}{\sqrt{n}}} \frac{1}{\sqrt{n}} |\sum_{t=1}^{n} f(X_t, \theta_0) I(0 < q_t \le r)\varepsilon_t| \ge \epsilon) = 0, \tag{7.2}$$

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

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

$$\varepsilon_t(\hat{\lambda}_n) = \varepsilon_t(\lambda_0) + \frac{\partial \varepsilon_t(\lambda_{nt})}{\partial \lambda'}(\hat{\lambda}_n - \lambda_0) = \varepsilon_t + \frac{1}{\sqrt{n}} \frac{\partial \varepsilon_t(\lambda_{nt})}{\partial \lambda'} \sqrt{n}(\hat{\lambda}_n - \lambda_0), \tag{7.3}$$

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

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{\partial L(0, \hat{\lambda}_n)}{\partial \delta} &= -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1}' \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n)\varepsilon_t - \frac{1}{n} \sum_{t=1}^{n} X_{t-1}' \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \frac{\partial \varepsilon_t(\lambda_{nt})}{\partial \lambda'} \sqrt{n}(\hat{\lambda}_n - \lambda_0) \\ &= -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1}' \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n)\varepsilon_t \\ &\quad - \frac{1}{n} \sum_{t=1}^{n} X_{t-1}' \hat{\theta}_{2n} D_t(\hat{r}_n, \hat{s}_n) \frac{\partial \varepsilon_t(\hat{\lambda}_n)}{\partial \lambda'} \sqrt{n}(\hat{\lambda}_n - \lambda_0) + R_n, \end{aligned} \tag{7.4}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

where

$$\begin{aligned} R_n =& \frac{1}{n}\sum_{t=1}^{n} X'_{t-1}\hat{\theta}_{2n}D_t(\hat{r}_n,\hat{s}_n)\left(\frac{\partial\varepsilon_t(\hat{\lambda}_n)}{\partial\lambda'} - \frac{\partial\varepsilon_t(\lambda_{nt})}{\partial\lambda'}\right)\sqrt{n}(\hat{\lambda}_n-\lambda_0) \\ =& \frac{1}{n^{3/2}}\sum_{t=1}^{n} X'_{t-1}\hat{\theta}_{2n}D_t(\hat{r}_n,\hat{s}_n)\sqrt{n}(\hat{\lambda}_n-\lambda_{nt})'\frac{\partial^2\varepsilon_t(\lambda^*_{nt})}{\partial\lambda\partial\lambda'}\sqrt{n}(\hat{\lambda}_n-\lambda_0), \end{aligned} \tag{7.5}$$

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 (7.3), $\sqrt{n}(\hat{\lambda}_n-\lambda_0) = O_p(1)$, $\sup_{t\le n}\sqrt{n}|\hat{\lambda}_n-\lambda_{nt}| \le \sqrt{n}|\hat{\lambda}_n-\lambda_0| = O_p(1)$. For any matrix or vector $A=(a_{ij})$, let $|A| = (|a_{ij}|)$. 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

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

with

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

By Assumption 2.4 and Lemma 7.1(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,

$$R_n = o_p(1). \tag{7.6}$$

Consider the first term on the right-hand side of (7.4). 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 7.1(i), we can show that, for some ξ_n^* lying between $\hat{\xi}_n$ and ξ_0,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_t(\hat{\xi}_n)\varepsilon_t =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_t(\xi_0)\varepsilon_t + [\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t(\xi_n^*)}{\partial \xi'}\varepsilon_t]\sqrt{n}(\hat{\xi}_n - \xi_0)\\ =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_t(\xi_0)\varepsilon_t + o_p(1),\end{aligned} \tag{7.7}$$

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X'_{t-1}\hat{\theta}_{2n}I(q_{t-1}>\hat{r}_n)\varepsilon_t =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}X'_{t-1}\theta_{20}I(q_{t-1}>\hat{r}_n)\varepsilon_t\\ &+ [\frac{1}{n}\sum_{t=1}^{n}X'_{t-1}I(q_{t-1}>\hat{r}_n)\varepsilon_t]\sqrt{n}(\hat{\theta}_{2n}-\theta_0)\\ =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}X'_{t-1}\theta_{20}I(q_{t-1}>\hat{r}_n)\varepsilon_t + o_p(1).\end{aligned} \tag{7.8}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

By Lemma 7.1(ii) and Assumption 2.4, we can also show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20}I(q_{t-1}>\hat{r}_n)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20}I(q_{t-1}>r_0)\varepsilon_t + o_p(1). \tag{7.9}$$

By (7.4), (7.6)-(7.9), Assumption 2.4, and Lemma 7.1(i), it follows that

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial L(0,\hat{\lambda}_n)}{\partial\delta} =& -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20}D_t(r_0,s_0)\varepsilon_t \\ &+ [\frac{1}{n}\sum_{t=1}^{n} X_{t-1}'\theta_{20}D_t(r_0,s_0)\frac{\partial\varepsilon_t(\lambda_0)}{\partial\lambda'}]\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\varepsilon_t(\lambda_0)}{\partial\lambda}\varepsilon_t + o_p(1).\end{aligned} \tag{7.10}$$

By the Ergodic Theorem and Central Limit Theorem, we have

$$\frac{1}{\sqrt{n}}\frac{\partial L(0,\hat{\lambda}_n)}{\partial\delta} \longrightarrow_{\mathcal{L}} N(0,\sigma^2\omega_2), \tag{7.11}$$

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 7.1(i), and the Ergodic Theorem,

$$-\frac{1}{n}\frac{\partial^2 L(0,\hat{\lambda}_n)}{\partial^2\delta} \to_p E\{\theta_{20}'X_{t-1}X_{t-1}'\theta_{20}D_t^2(r_0,s_0)\} = \omega_1. \tag{7.12}$$

By (3.3), (7.11)-(7.12), $\hat{\sigma}_{0n}^2 \to_p \sigma^2$, $\hat{\omega}_{1n}\to_p \omega_1$, $\hat{\omega}_{2n}\to_p\omega_2$, and Slutsky's theorem, we have

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

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

Proof of Theorem 3.2. With a similar argument, 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 (7.3), $\partial\varepsilon_t(\theta, \hat{r}_n)/\partial\theta'$ does not depend on θ anymore. Write $V_t(r) = \partial\varepsilon_t(\theta, r)/\partial\theta$. By Assumption 2.5, $\hat{r}_n - r_0 = O_p(1/n)$, and, by (S2.2) and the uniform boundedness of $D_t(r, s)$, it is not hard to show that

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

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

$$\begin{aligned} \frac{1}{\sqrt{n}}\frac{\partial L(1, \hat{\theta}_n, s, \hat{r}_n)}{\partial\delta} =& -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X'_{t-1}\hat{\theta}_{2n}D_t(\hat{r}_n, s)\varepsilon_t \\ & -[\frac{1}{n}\sum_{t=1}^{n} X'_{t-1}\hat{\theta}_{2n}D_t(\hat{r}_n, s)V_t(\hat{r}_n)']\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1), \end{aligned} \tag{7.13}$$

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

Consider the first term on the right-hand side of (7.13). Let $\zeta = (\theta'_2, r)'$ and $g_t(\zeta, s) = X'_{t-1}\theta_2 G_t(q_{t-1}, s, r)$. Then,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X'_{t-1}\hat{\theta}_{2n}G(q_{t-1}, s, \hat{r}_n)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_t(\zeta_0, s)\varepsilon_t + [\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t(\zeta_n^*, s)}{\partial\zeta'}\varepsilon_t]\sqrt{n}(\hat{\zeta}_n - \zeta_0) \tag{7.14}$$

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, $\tau \in [1/\bar{s}, \bar{s}]$,

$$\begin{aligned} |\frac{\partial g_t(\zeta_n^*, s)}{\partial \zeta'} - \frac{\partial g_t(\zeta_n^*, \tau)}{\partial \zeta'}| \leq & K(|X_{t-1}'|(|q_{t-1}|^{\alpha_1} + 1), \|X_{t-1}\|(|q_{t-1}|^{\alpha_4} + 1))|s - \tau| \\ \triangleq & J_t|s - \tau|, \end{aligned} \quad (7.15)$$

where J_t is strictly stationary and ergodic. Let $\Delta(\eta) = \{(\theta_2, r) : \|\theta_2 - \theta_0\| + |r - r_0| \leq \eta\}$. By (7.15), a standard piecewise argument on $s \in [1/\bar{s}, \bar{s}]$ and Lemma 7.1(i), we can show that

$$\sup_{s\in[1/\bar{s},\bar{s}]} \sup_{\Delta(\eta)} |\frac{1}{n}\sum_{t=1}^{n} \frac{\partial g_t(\zeta, s)}{\partial \zeta'}\varepsilon_t - \frac{1}{n}\sum_{t=1}^{n} \frac{\partial g_t(\zeta_0, s)}{\partial \zeta'}\varepsilon_t| = o_p(1), \quad (7.16)$$

for η small enough. By the Ergodic Theorem, (7.15) and a standard piecewise argument as Lemma A.1 in Francq et al. (2010),

$$\sup_{s\in[1/\bar{s},\bar{s}]} |\frac{1}{n}\sum_{t=1}^{n} \frac{\partial g_t(\zeta_0, s)}{\partial \zeta'}\varepsilon_t| = o_p(1). \quad (7.17)$$

By Assumption 2.5, (7.16) and (7.17), it follows that

$$\sup_{s\in[1/\bar{s},\bar{s}]} |\frac{1}{n}\sum_{t=1}^{n} \frac{\partial g_t(\zeta_n^*, s)}{\partial \zeta'}\varepsilon_t| = o_p(1). \quad (7.18)$$

By Assumption 2.5, (S2.2), and a similar argument as (7.9), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n} I(q_{t-1} > \hat{r}_n)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20} I(q_{t-1} > r_0)\varepsilon_t + o_p(1). \tag{7.19}$$

By (7.14) and (7.18)-(7.19), it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\hat{\theta}_{2n} D_t(\hat{r}_n, s)\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}'\theta_{20} D_t(r_0, s)\varepsilon_t + o_p(1), \tag{7.20}$$

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

Consider the second term on the right-hand side of (7.13). 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

$$|B_t(\theta_2, r, s) - B_t(\theta_2, r, \tau)|^2 \leq K|X_{t-1}'\theta_2 V_t(r)'|(|q_{t-1}|^{\alpha_1} + 1)|s - \tau| = Q_t|s - \tau|. \tag{7.21}$$

where Q_t is strictly stationary and ergodic.

By Lemma 7.1(i), a standard piecewise argument on $s \in [1/\bar{s}, \bar{s}]$ and (7.21), we can show that for any $\epsilon > 0$, there exits an $\eta > 0$ such that

$$\lim_{n\to\infty} P(\sup_{s\in[1/\bar{s},\bar{s}]} \sup_{\Delta(\eta)} \frac{1}{n}|\sum_{t=1}^{n}[B_t(\theta_2, r, s) - B_t(\theta_{20}, r_0, s)]| \geq \epsilon) = 0. \tag{7.22}$$

By Assumption 2.5, (7.20), and (7.22), (7.13) reduces to

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial L(1,\hat{\theta}_n,s,\hat{r}_n)}{\partial\delta} =& -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{t-1}'\theta_{20}D_t(r_0,s)\varepsilon_t \\ &+[\frac{1}{n}\sum_{t=1}^{n}X_{t-1}'\theta_{20}D_t(r_0,s)V_t(r_0)']\Sigma_2^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}V_t(r_0)\varepsilon_t+o_p(1) \\ &\triangleq u_{1n}(s)+u_{2n}(s)+o_p(1). \qquad (7.23)\end{aligned}$$

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 (7.23) are square-integrable stationary martingale differences. 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,

$$\begin{aligned}E[u_{1n}(s)-u_{1n}(\tau)]^2 =& E(X_{t-1}\theta_{20})^2(\frac{\partial G(q_{t-1},\tilde{s}_1,r_0)}{\partial s})^2(s-\tau)^2\sigma^2 \\ &\leq K^2E(X_{t-1}\theta_{20})^2(|q_{t-1}|^{\alpha_1}+1)^2(s-\tau)^2\sigma^2 \\ &\leq K(s-\tau)^2, \qquad (7.24)\end{aligned}$$

$$\begin{aligned} E[u_{2n}(s)-u_{2n}(\tau)]^2 =& E\left\{[\frac{1}{n}\sum_{t=1}^{n}X_{t-1}'\theta_{20}\frac{\partial G(q_{t-1},\tilde{s}_2,r_0)}{\partial s}V_t(r_0)']\Sigma_2^{-1}[\frac{1}{n}\sum_{t=1}^{n}X_{t-1}'\theta_{20}\right. \\ &\left.\times\frac{\partial G(q_{t-1},\tilde{s}_2,r_0)}{\partial s}V_t(r_0)]\right\}(s-\tau)^2\sigma^2. \\ \leq& K(s-\tau)^2\sigma^2, \end{aligned} \tag{7.25}$$

where (7.25) holds by Assumption 3.1(ii) and the 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 (7.24) and (7.25), the tightness follows from Theorem 12.3 of Billingsley (1968). By the Central Limit Theorem and the Ergodic Theorem, the form of the limiting Gaussian process follows immediately from (7.32). 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 $\tilde{s}_3\in[\tau,s]$,

$$\begin{aligned} |Z_t(\theta_2,r,s)-Z_t(\theta_2,r,\tau)| =& 2|\theta_2'X_{t-1}X_{t-1}'\theta_2D_t(r,s)||\frac{\partial G(q_{t-1},\tilde{s}_3,r)}{\partial s}||s-\tau| \\ \leq& 2K|\theta_2'X_{t-1}X_{t-1}'\theta_2|(|q_{t-1}|^{\alpha_1}+1)|s-\tau| \\ \triangleq& A_t(\theta_2)|s-\tau|, \end{aligned} \tag{7.26}$$

where $A_t(\theta_2)$ is strictly stationary and ergodic. By (7.26), Lemma 7.1(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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

$$\lim_{n\to\infty} P(\sup_{s\in[1/\bar{s},\bar{s}]} \sup_{\Delta(\eta)} \frac{1}{n}|\sum_{t=1}^{n}[Z_t(\theta_2, r, s) - Z_t(\theta_{20}, r_0, s)]| \geq \epsilon) = 0. \tag{7.27}$$

By (7.26), the Ergodic Theorem and a similar standard piecewise argument on $s \in [1/\bar{s}, \bar{s}]$ or Lemma A.1 in Francq et al. (2010), we can show that

$$\sup_{s\in[1/\bar{s},\bar{s}]} |\frac{1}{n}\sum_{t=1}^{n} Z_t(\theta_{20}, r_0, s) - \omega(s)| = o_p(1), \tag{7.28}$$

where $\omega(s)$ is defined in Theorem 3.2. By Assumption 2.5, (b) follows from (7.27) and (7.28). This completes the proof. □

Supplementary Materials

Owing to space constraint, a discussion of some nested tests, the proofs of Theorems 4.1-4.2 and some related tables are provided in the supplementary material.

Acknowledgments

We are grateful to the Editor, an Associate Editor and the referees for their insightful comments and suggestions that have substantially improved the presentation and the content of this paper. This work was supported in part by Hong Kong Research Grants Commission Grants HKUST 603413, GRF 16500117, GRF 16500915 and GRF 16307516.

References

Atkinson, A. C. (1970). A method for discriminating between models. *J. Roy. Stat. Soc.* **B32**, 323–353.

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.

Davis R. A, Huang D, Yao Y. C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Ann. Stat.* **23(1)**, 282–304.

Di Narzo, A. F., Aznare, J. J., and Stigler, M. (2013). The R package `tsDyn`. *Available on-line at* `http://cran.r-project.org/`.

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.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. Trending multiple time series. *Econometric Th* **11**, 818–887.

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.

Ling, S. and Tong, H. (2011). Score based goodness-of-fit tests for time series. *Stat. Sinica* **21(4)**, 1807–1829.

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 Deaton A. S. (1978) Testing non-nested nonlinear regression models. *Econometrica* **46**, 677–694.

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.

Booth School of Business, University of Chicago

E-mail: mazxgao@gmail.com

Department of Mathematics, Hong Kong University of Science and Technology

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0497

E-mail: maling@ust.hk

School of Mathematical Sciences, University of Electronic Science and Technology of China

Department of Statistics, London School of Economics

E-mail: howell.tong@gmail.com