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*

Supplementary Material

In this supplementary material, we give a discussion of some nested tests , the proofs of Theorems 4.1-4.2 and some related tables and further examinations of the real data examples in the original article.

S1 Discussion of some nested tests

It is interesting to see whether or not we can construct a nested test for TAR against STAR models or whether or not our test S_{2n} has local power for some transition speed $s_n \to \infty$. To make it simple, we consider the following LSTAR(1) model with $r_0 = 0$:

$$y_t = \frac{1}{1+\beta^{y_{t-1}}} + \varepsilon_t, \tag{S1.1}$$

where $\beta = e^{-s}$. If $\beta = 0$ (i.e. $s = \infty$), $1/(1+\beta^{y_{t-1}}) = I(y_{t-1} > 0)$, in which case model (S1.1) reduces to a TAR model. Consider the nested hypothesis:

$$\bar{H}_0 : \beta = 0 \quad vs. \quad \bar{H}_1 : \beta \in (0, 1]. \tag{S1.2}$$

We assume σ^2 is known. The difference of the log-likelihood functions is

$$LR_n(\beta) = \frac{1}{\sigma^2}\left[\sum_{t=1}^n (y_t - I(y_{t-1} > 0))^2 - \sum_{t=1}^n (y_t - \frac{1}{1+\beta^{y_{t-1}}})^2\right]. \tag{S1.3}$$

Under $\bar{H}_0$, we have

$$LR_n(\beta) = \frac{1}{\sigma^2}\left[-2\sum_{t=1}^n \varepsilon_t(I(y_{t-1} > 0) - \frac{1}{1+\beta^{y_{t-1}}}) - \sum_{t=1}^n (I(y_{t-1} > 0) - \frac{1}{1+\beta^{y_{t-1}}})^2\right]. \tag{S1.4}$$

Let $g(\beta) = E[I(y_{t-1} > 0) - \frac{1}{1+\beta^{y_{t-1}}}]^2$. It is not hard to show that

$$\frac{1}{n}\sum_{t=1}^n (I(y_{t-1} > 0) - \frac{1}{1+\beta^{y_{t-1}}})^2 \to_p g(\beta) \tag{S1.5}$$

uniformly for $\beta \in [0,1]$ and

$$S_n(\beta) := -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t (I(y_{t-1} > 0) - \frac{1}{1+\beta^{y_{t-1}}}) \overset{f.d.}{\Longrightarrow} \sigma B(g(\beta)) \qquad \text{(S1.6)}$$

in $C[0,1]$, where $f.d.$ denotes weak convergence in finite dimensions and $B(\cdot)$ is the standard Brownian motion. But we cannot prove the tightness of (S1.6). Under $\bar{H}_0$, we guess that the likelihood-ratio test statistic

$$LR_n := \max_{\beta\in[0,1]} LR_n(\beta) = \max_{\beta\in[0,1]} \left\{ 2\frac{\sqrt{n}}{\sigma^2} S_n(\beta) - \frac{n}{\sigma^2}[g(\beta) + o_p(1)] \right\}$$
$$\approx_d \max_{u\geq 0} \{2B(u) - u\}, \qquad \text{(S1.7)}$$

where we use the sample time transformation $u := \sigma^{-2} n g(\beta)$ and $\approx_d$ means approximately equal in distribution. We expect that the test (S1.7) should have local power when s_n satisfies $ng(e^{-s_n}) = O_p(1)$. However, we are not able to prove (S1.7) for the time being. We leave this issue as an open problem.

We conduct a small simulation for (S1.2) by using distribution (S1.7). The sample sizes (n) are 50, 100, 300, 500 and 1000 and the replication time is 1000. If we define $\xi = \max_{u\geq 0}\{W(u) - u\}$, then $P(\xi \leq x) = 1 - e^{-x/2}$; see Hansen (2000) (pp. 601). The critical values at $\alpha = 0.1$ and 0.05 are 4.62 and 6, respectively; see Table 1 in Hansen (1997). The results are reported in Table S1. From Table S1, we can see that if the true model is TAR (i.e., $\beta = 0$), the size tends to be 0 in all nominal levels for moderate sample sizes. This is because under $\bar{H}_0$, $LR_n(\beta) \approx -\sigma^{-2} n[g(\beta) + o_p(1)]$ for large n and $g(\beta)$ is a strictly increasing function with respect to β. It is very interesting to see that the test statistic (S1.7) is very powerful to detect a STAR model even for a small β and small sample sizes. This simulation shows that (S1.7) probably is a reasonable distribution and may provide a way to study the testing problem in (S1.2).

Table S1: The size and power using (S1.7).

	$\beta = 0$		$\beta = 0.01$		$\beta = 0.05$		$\beta = 0.1$		$\beta = 0.2$	
α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
n=50	0.020	0.008	0.134	0.081	0.219	0.134	0.283	0.164	0.409	0.285
n=100	0.031	0.016	0.290	0.185	0.461	0.318	0.557	0.420	0.684	0.577
n=300	0.017	0.009	0.667	0.573	0.849	0.787	0.912	0.880	0.966	0.953
n=500	0.009	0.005	0.854	0.806	0.995	0.939	0.993	0.985	0.999	0.999
n=1000	0.003	0.002	0.972	0.962	1.000	0.999	1.000	1.000	1.000	1.000

On the other hand, when testing STAR against TAR models, we can first fit a STAR model using the data and obtain a least squares estimator $\hat{s}_n$ of

s_0. Based on $\hat{s}_n$, we can formulate the hypothesis as

$$\mathcal{H}_0 : s_0 \in [s_1, s_2] \quad vs. \quad \mathcal{H}_1 : s_0 = \infty, \tag{S1.8}$$

where $[s_1, s_2]$ may be replaced by $(0, s_2]$ since we are more interested in s_2. Essentially, we can achieve this by testing whether $s_0 \le s_2$ or not, since a large s_0 indicates a STAR model is effectively a TAR model. To make it simple, we still consider model (S1.1). Let

$$l_t(s) = (y_t - \frac{1}{1 + e^{-sy_{t-1}}})^2.$$

Then, by a standard Taylor's expansion,

$$\begin{aligned}\sqrt{n}(\hat{s}_n - s_0) &= -[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(s_0)}{\partial^2 s}]^{-1}[\frac{1}{\sqrt{n}}\frac{\partial l_t(s_0)}{\partial s}] + o_p(1) \\ &= [\frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}^2 e^{-2s_0 y_{t-1}}}{(1+e^{-s_0 y_{t-1}})^4}]^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\frac{y_{t-1}e^{-s_0 y_{t-1}}}{(1+e^{-sy_{t-1}})^2} + o_p(1).\end{aligned} \tag{S1.9}$$

It is not hard to show that

$$\sqrt{n}(\hat{s}_n - s_0) \to_{\mathcal{L}} N(0, \sigma^2/E\frac{y_{t-1}^2 e^{-2s_0 y_{t-1}}}{(1+e^{-s_0 y_{t-1}})^4}). \tag{S1.10}$$

Then, a natural test for the null hypothesis $s_0 \le s_2$ would be

$$S_{3n} := \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}^2 e^{-2\hat{s}_n y_{t-1}}}{(1+e^{-\hat{s}_n y_{t-1}})^4}}\sqrt{n}(\hat{s}_n - s_2)}{\hat{\sigma}_n}, \tag{S1.11}$$

where $\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{t=1}^{n} l_t(\hat{s}_n)}$. We reject $\mathcal{H}_0$ at level α if $S_{3n} > N_{1-\alpha}$, where $N_{1-\alpha}$ is the $(1-\alpha)-$quantile of the standard normal distribution.

In Tables S2-S4, we present some simulation results for testing $\mathcal{H}_0$ using the test statistic S_{3n} in (S1.11). We choose the upper bound $s_2 = 1$, 2 and 5 respectively, and the $\bar{s}$ is set to be 20 when estimating $\hat{s}_n$ over the interval $[1/\bar{s}, \bar{s}]$. Meanwhile, we also report the results of the empirical power using our proposed test statistic S_{1n} to make a comparison in Tables S5-S7 with $\bar{s} = 10$, 20, and 40, respectively. For each s_2, the true switching parameter $s_0 = 1$, 2, 5, 10, 15 and ∞, where $s_2 = \infty$ corresponds to $\mathcal{H}_1$ in (S1.8). From Tables S2-S4, we can see that when s_2 is very small (e.g. $s_2 = 1$ in Table S2), we can obtain high power for a larger s_0 as well as the case with $s_0 = \infty$. When $s_0 > s_2$, the power will increase as the sample size increases, and we generally require $n = 1000$ for decent power. When s_2 increases a little bit (e.g. $s_2 = 5$ in Table S4), even for a sample size of 1000, the power is still not satisfactory. Overall, we conclude that the decent power is only achieved when s_2 is very small and the sample size is very large. In practice, s_2 is often very large, therefore, the test statistic (S1.11) is not a promising one compared to the test statistics in our original paper. While in Tables S5-S7, we can see that we always have non-trivial power for all the choices of $\bar{s}$. For each $\bar{s}$, the power will increase as the sample size increases and a small $\bar{s}$ often gives a higher power, since the estimator will be more stable when the searching interval $[1/\bar{s}, \bar{s}]$ is narrower.

Table S2: The size and power using (S1.11) with $s_2 = 1$.

	$s_0 = 1$		$s_0 = 2$		$s_0 = 5$		$s_0 = 10$		$s_0 = 15$		$s_0 = \infty$	
α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
n=100	0	0	0	0	0	0	0	0	0	0	0	0
n=300	0.050	0.002	0.590	0.090	0.768	0.122	0.384	0.040	0.228	0.016	0.068	0.006
n=500	0.036	0.008	0.868	0.648	0.984	0.886	0.898	0.444	0.568	0.210	0.188	0.038
n=1000	0.056	0.016	0.982	0.948	1	1	1	0.998	1	0.806	1	0.576

Table S3: The size and power using (S1.11) with $s_2 = 2$.

	$s_0 = 1$		$s_0 = 2$		$s_0 = 5$		$s_0 = 10$		$s_0 = 15$		$s_0 = \infty$	
α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
n=100	0	0	0	0	0	0	0	0	0	0	0	0
n=300	0	0	0.002	0	0.018	0	0.004	0	0.01	0	0	0
n=500	0	0	0.034	0	0.720	0.006	0.622	0.002	0.390	0.004	0.100	0
n=1000	0	0	0.054	0.010	0.990	0.932	0.998	0.876	1	0.658	0.994	0.174

Table S4: The size and power using (S1.11) with $s_2 = 5$.

	$s_0 = 1$		$s_0 = 2$		$s_0 = 5$		$s_0 = 10$		$s_0 = 15$		$s_0 = \infty$	
α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
n=100	0	0	0	0	0	0	0	0	0	0	0	0
n=300	0	0	0	0	0	0	0	0	0	0	0	0
n=500	0	0	0	0	0	0	0	0	0	0	0	0
n=1000	0	0	0	0	0.022	0	0.268	0	0.484	0	0.526	0

Finally, it is also interesting to see the behavior of δ when using the

Table S5: The power using S_{1n} with $\bar{s} = 10$.

	α	$n = 100$	$n = 300$	$n = 500$	$n = 1000$
$s_0 = \infty$	0.1	0.426	0.740	0.872	0.986
	0.05	0.334	0.640	0.824	0.976

Table S6: The power using S_{1n} with $\bar{s} = 20$.

	α	$n = 100$	$n = 300$	$n = 500$	$n = 1000$
$s_0 = \infty$	0.1	0.338	0.546	0.642	0.886
	0.05	0.244	0.448	0.552	0.822

Table S7: The power using S_{1n} with $\bar{s} = 40$.

	α	$n = 100$	$n = 300$	$n = 500$	$n = 1000$
$s_0 = \infty$	0.1	0.298	0.370	0.492	0.668
	0.05	0.236	0.294	0.400	0.568

compound model (2.4) to fit the data. For ease of exposition, consider the model

$$y_t = (1-\delta)\frac{1}{1+e^{-sy_{t-1}}} + \delta I(y_{t-1} > 0) + \varepsilon_t. \tag{S1.12}$$

We assume that $\delta \in [0,1]$ and $s \in [s_1, s_2]$. Under these conditions, δ and s are identifiable except the case when $\delta = 1$. We first consider the hypothesis in (2.5). Under H_0 (e.g. $\delta = 0$), we use model (S1.12) to fit the data. Let $\theta = (\delta, s)'$ and

$$\tilde{l}_t(\theta) = [y_t - (1-\delta)\frac{1}{1+e^{-sy_{t-1}}} - \delta I(y_{t-1} > 0)]^2.$$

If the least squares estimator $\hat{\theta}_n$ of θ_0 is consistent, we can further show that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial^2\theta})^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} + o_p(1), \tag{S1.13}$$

where $\theta_0 = (0, s_0)$. Similar to (S1.10), $\sqrt{n}\hat{\delta}_n$ is asymptotically normal and a Wald-type statistic can be constructed to test H_0 in (2.5). Unfortunately, the Hessian matrix $E\frac{\partial^2 \tilde{l}_t(\theta)}{\partial^2\theta}$ is not positive definite at the true parameter θ_0, which means that $\hat{\theta}_n$ is not a consistent estimator and (S1.13) does not hold in this case. Simulation results also suggest that there is no power when using a Wald-type test statistic based on (S1.13) . We do not report the details here.

Now, we consider the hypothesis in (2.6). When $\delta = 1$, s is not identifiable and we can take s as a nuisance parameter. For each $s \in [s_1, s_2]$, the profile least squares estimator of δ satisfies

$$\hat{\delta}_n(s) - 1 = -\frac{\sum_{t=1}^n \varepsilon_t D_t(s)}{\sum_{t=1}^n D_t(s)^2}, \tag{S1.14}$$

where $D_t(s) = (\frac{1}{1+e^{-sy_{t-1}}} - I(y_{t-1} > 0))$. Following the techniques in the proofs of Theorem 3.2, we can show that $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^n \tilde{l}(\hat{\delta}_n(s), s) \to_p \sigma^2$, $\hat{\delta}_n(s) \to_p 1$ uniformly in $D[s_1, s_2]$, and

$$\sqrt{n}(\hat{\delta}_n(s) - 1) \Longrightarrow \frac{\sigma G(s)}{w(s)} \quad \text{in} \quad D[s_1, s_2],$$

where $G(s)$ is a Gaussian process with mean zero and covariance $EG(s)G(\tau) = ED_t(s)D_t(\tau)$, and $w(s) = ED_t(s)^2$. By continuous mapping theorem,

$$S_{4n} := \sup_{s\in[s_1,s_2]} \frac{n(\hat{\delta}_n(s) - 1)^2}{\hat{\sigma}_n^2(s)} \Longrightarrow T := \sup_{s\in[s_1,s_2]} \frac{G(s)^2}{w(s)^2}. \tag{S1.15}$$

We reject $\tilde{H}_0$ at level α if $S_{4n} > T_{1-\alpha}$, where $T_{1-\alpha}$ is the $(1-\alpha)-$quantile of T in $(S1.15)$. In practice, $[s_1, s_2]$ can be replaced by $[1/\bar{s}, \bar{s}]$, the distribution of T can be approximated by a bootstrap method.

We report some simulation results in Tables S5-S6 based on $(S1.15)$. Similarly, we also report the results of the empirical power using our proposed test statistic S_{2n} in the original paper to make a comparison. We choose $\bar{s} = 10$ and 20, respectively, and $s_0 = 2$, 5 and 10 for each case. From Table S5, we can see that the size is very accurate for all sample sizes and the power increases as the sample size becomes larger and we can achieve decent power for $n \geq 500$. When $\bar{s}$ increases, the power will decrease a little bit. While in Tables S10-S11, we can see that our proposed test gives a higher power for all cases and when $\bar{s}$ decreases, the power does not necessarily decrease.

Table S8: The size and power using (S1.15) with $\bar{s} = 10$.

	n	$n=100$		$n=300$		$n=500$		$n=1000$	
	α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
$\delta=1$		0.036	0.022	0.064	0.028	0.082	0.024	0.086	0.050
$\delta=0$	s_0=2	0.002	0	0.778	0.030	1	0.986	1	1
	s_0=5	0.008	0.004	0.404	0	0.882	0.790	0.996	0.988
	s_0=10	0.016	0.010	0.244	0	0.718	0.572	0.948	0.890

Table S9: The size and power using (S1.15) with $\bar{s} = 20$.

	n	$n = 100$		$n = 300$		$n = 500$		$n = 1000$	
	α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
$\delta = 1$		0.040	0.022	0.056	0.034	0.090	0.028	0.080	0.028
	s_0=2	0.002	0.002	0	0	0.820	0.016	1	0.998
$\delta = 0$	s_0=5	0.004	0	0.054	0	0.712	0.044	0.982	0.942
	s_0=10	0.004	0	0	0	0.394	0	0.898	0.792

Table S10: The power using S_{2n} with $\bar{s} = 10$.

	n	$n = 100$		$n = 300$		$n = 500$		$n = 1000$	
	α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
	s_0=2	0.798	0.700	0.998	0.990	1	1	1	1
$\delta = 0$	s_0=5	0.532	0.428	0.870	0.794	0.964	0.930	0.998	0.996
	s_0=10	0.376	0.248	0.618	0.474	0.772	0.672	0.940	0.888

Table S11: The power using S_{2n} with $\bar{s} = 20$.

	n	$n = 100$		$n = 300$		$n = 500$		$n = 1000$	
	α	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
	s_0=2	0.800	0.680	0.994	0.980	1	1	1	1
$\delta = 0$	s_0=5	0.586	0.432	0.878	0.788	0.960	0.916	1	1
	s_0=10	0.440	0.294	0.676	0.554	0.818	0.696	0.980	0.952

S2 Proofs of Lemma 7.1

Proof. (i). We only consider the case when $r \geq r_0$. We note that

$$P\left(\sup_{\substack{\|\theta-\theta_0\|\leq\eta \\ |r-r_0|\leq\eta}} \frac{1}{n} \left| \sum_{t=1}^{n} [f(X_t,\theta)I(q_t \leq r) - f(X_t,\theta_0)I(q_t \leq r_0)] \right| \geq \epsilon \right)$$

$$\begin{aligned}
\leq & P\left(\sup_{\substack{\|\theta-\theta_0\|\leq\eta \\ |r-r_0|\leq\eta}} \frac{1}{n}\left|\sum_{t=1}^{n}[f(X_t,\theta)-f(X_t,\theta_0)]I(q_t\leq r)\right| \geq \frac{\epsilon}{2}\right) \\
& + P\left(\sup_{|r-r_0|\leq\eta} \frac{1}{n}\left|\sum_{t=1}^{n}[f(X_t,\theta_0)I(r_0<q_t\leq r)]\right| \geq \frac{\epsilon}{2}\right) \\
\leq & P\left(\sup_{\|\theta-\theta_0\|\leq\eta} \frac{1}{n}\sum_{t=1}^{n}|f(X_t,\theta)-f(X_t,\theta_0)| \geq \frac{\epsilon}{2}\right) \\
& + P\left(\frac{1}{n}\sum_{t=1}^{n}|f(X_t,\theta_0)|\, I(r_0<q_t\leq r_0+\eta) \geq \frac{\epsilon}{2}\right) \\
\triangleq & P_{1n}+P_{2n}. \qquad \text{(S2.1)}
\end{aligned}$$

Let

$$H_{1t}(\eta) = \sup_{\|\theta-\theta_0\|\leq\eta} |f(X_t,\theta)-f(X_t,\theta_0)|.$$

Since $E\sup_{\theta\in\Theta}|f(X_t,\theta)|<\infty$, $f(X_t,\theta)$ is continuous in θ and Θ is a compact set, by dominated convergence theorem, for any $\epsilon>0$, there exists an $\eta_1>0$ small enough, such that $EH_{1t}(\eta_1)<\epsilon/4$.

First we note that 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,

$$|E[ZI(r_1<q_r\leq r_2)]| = |E[Z|q_t=r^*]f_q(r^*)|\,|r_2-r_1|, \qquad \text{(S2.2)}$$

where r^* lies between r_1 and r_2, and f_q is the density of q_t.

Let

$$H_{2t}(\eta) = |f(X_t,\theta_0)|I(r_0<q_t\leq r_0+\eta),$$

by assumptions 2.2-2.3 and (S2.2), we have

$$EH_{2t}(\eta)\leq K\eta.$$

Then we choose $\eta_2<\epsilon/(4K)$ so that $EH_{2t}(\eta_2)<\epsilon/4$. Now, we choose $\eta=\min\{\eta_1,\eta_2\}$ and hence $H_{1t}(\eta)$ and $H_{2t}(\eta)$ are both strictly stationary and ergodic. By ergodic theorem, we have

$$P_{1n}\leq P\left(\frac{1}{n}\sum_{t=1}^{n}[H_{1t}(\eta)-EH_{1t}(\eta)]\geq\frac{\epsilon}{4}\right)+P\left(EH_{1t}(\eta)\geq\frac{\epsilon}{4}\right)\to 0, \qquad \text{(S2.3)}$$

and

$$P_{2n} \leq P\left(\frac{1}{n}\sum_{t=1}^{n}[H_{2t}(\eta) - EH_{2t}(\eta)] \geq \frac{\epsilon}{4}\right) + P\left(EH_{2t}(\eta) \geq \frac{\epsilon}{4}\right) \to 0 \quad \text{(S2.4)}$$

as $n \to \infty$. By (S2.1), (S2.3) and (S2.4), we conclude that (7.1) 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,

$$\begin{aligned}
&P(\sup_{0\leq r\leq \frac{M}{\sqrt{n}}} \frac{1}{\sqrt{n}}|\sum_{t=1}^{n} f(X_t, \theta_0)I(0 < q_t \leq r)\varepsilon_t| \geq \epsilon) \\
&\leq P(\sup_{1\leq i\leq N} \sup_{\frac{M_{i-1}}{\sqrt{n}}\leq r\leq \frac{M_i}{\sqrt{n}}} \frac{1}{\sqrt{n}}|\sum_{t=1}^{n} f(X_t, \theta_0)I(0 < q_t \leq r)\varepsilon_t| \geq \epsilon) \\
&\leq P(\sup_{1\leq i\leq N} \sup_{\frac{M_{i-1}}{\sqrt{n}}\leq r\leq \frac{M_i}{\sqrt{n}}} \frac{1}{\sqrt{n}}|\sum_{t=1}^{n} f(X_t, \theta_0)I(\frac{M_{i-1}}{\sqrt{n}} < q_t \leq r)\varepsilon_t| \geq \epsilon/2) \\
&\quad + \sum_{i=1}^{N} P(\frac{1}{\sqrt{n}}|\sum_{t=1}^{n} f(X_t, \theta_0)I(0 < q_t \leq \frac{M_{i-1}}{\sqrt{n}})\varepsilon_t| \geq \epsilon/2) \\
&\leq \left\{\sum_{i=1}^{N} P\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[|f(X_t, \theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}} < q_t \leq \frac{M_i}{\sqrt{n}})\right.\right.\right. \\
&\quad \left.\left. -E(|f(X_t, \theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}} < q_t \leq \frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-1})\right] \geq \frac{\epsilon}{2(p+1)}\right) + \cdots \\
&\quad + \sum_{i=1}^{N} P\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[E(|f(X_t, \theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}} < q_t \leq \frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p+1})\right.\right. \\
&\quad \left.\left.\left. -E(|f(X_t, \theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}} < q_t \leq \frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p})\right] \geq \frac{\epsilon}{2(p+1)}\right)\right\} \\
&\quad + P(\sup_{1\leq i\leq N} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} E(|f(X_t, \theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}} < q_t \leq \frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p}) \geq \frac{\epsilon}{2(p+1)}) \\
&\quad + \sum_{i=1}^{N} P(\frac{1}{\sqrt{n}}|\sum_{t=1}^{n} f(X_t, \theta_0)I(0 < q_t \leq \frac{M_{i-1}}{\sqrt{n}})\varepsilon_t| \geq \epsilon/2) \\
&\triangleq \Pi_{1n} + \Pi_{2n} + \Pi_{3n}. \qquad \text{(S2.5)}
\end{aligned}$$

By (S2.2) and assumptions 2.2-2.3, we have the following three inequalities in order,

$$
\begin{aligned}
&E\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-j})\right.\right.\\
&\left.\left.-E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-j-1})\right]\right)^2\\
\le&2E|f(X_t,\theta_0)\varepsilon_t|^2 I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})\\
\le&K\frac{\delta}{\sqrt{n}},
\end{aligned}
\tag{S2.6}
$$

$$
\begin{aligned}
&E[\sup_{1\le i\le N}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E(|f(X_t,\theta_0)\varepsilon_t|I(\frac{M_{i-1}}{\sqrt{n}}<q_t\le\frac{M_i}{\sqrt{n}})|\mathcal{F}_{t-p})]\\
\le&E[\sup_{1\le i\le N}\frac{K}{\sqrt{n}}\sum_{t=1}^{n}\wp_{t-p}\frac{M_i-M_{i-1}}{\sqrt{n}}]E|\varepsilon_t|\\
\le&\delta\{\frac{K}{n}\sum_{t=1}^{n}E[\wp_{t-p}]\}E|\varepsilon_t|\\
\le&K\delta,
\end{aligned}
\tag{S2.7}
$$

and

$$
\begin{aligned}
&E(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}f(X_t,\theta_0)I(0<q_t\le\frac{M_{i-1}}{\sqrt{n}})\varepsilon_t)^2\\
=&Ef(X_t,\theta_0)^2\varepsilon_t^2 I(0<q_t\le\frac{M_{i-1}}{\sqrt{n}})\\
\le&K\frac{M_{i-1}}{\sqrt{n}},
\end{aligned}
\tag{S2.8}
$$

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 (S2.6)-(S2.8), we have

$$
\Pi_{1n}+\Pi_{2n}+\Pi_{3n}\le\sum_{i=1}^{N}\frac{Kp\delta}{\sqrt{n}[\epsilon/(2(p+1))]^2}+\frac{K\delta}{[\epsilon/(2(p+1))]}+\sum_{i=1}^{N}\frac{M_{i-1}}{\sqrt{n}(\epsilon/2)^2}\to 0,
\tag{S2.9}
$$

as $n\to\infty$ and $\delta\to 0$. Then, (7.2) follows from (S2.5) and (S2.9). $\square$

S3 Proofs of Theorems 4.1-4.2

Proof of Theorem 4.1. We only consider the case when $\delta_0 = 0$ since it is similar for $\delta_0 = 1$. Under H_0, $\varepsilon_t = \varepsilon_t(\lambda_0, \delta_0)$ and $g(x) = 1/(\sqrt{2\pi}\sigma)e^{-x^2/(2\sigma^2)}$. Thus,

$$
\begin{aligned}
\Lambda_{n,\lambda}(0, \frac{\gamma}{\sqrt{n}}) =& \frac{1}{2\sigma^2}\sum_{t=1}^{n}[\varepsilon_t^2(\lambda_0, 0) - \varepsilon_t^2(\lambda_0, \frac{\gamma}{\sqrt{n}})] \\
=& \frac{1}{2\sigma^2}\sum_{t=1}^{n}\{\varepsilon_t^2 - [\varepsilon_t + \frac{\gamma}{\sqrt{n}}X'_{t-1}\theta_{20}D_t(r_0, s_0)]^2\} \\
=& -\frac{\gamma}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}X'_{t-1}\theta_{20}D_t(r_0, s_0)\varepsilon_t \\
& -\frac{\gamma^2}{2n\sigma^2}\sum_{t=1}^{n}[X'_{t-1}\theta_{20}D_t(r_0, s_0)]^2, \qquad \text{(S3.1)}
\end{aligned}
$$

where $D_t(r, s)$ is defined as (3.1). By ergodic theorem and central limit theorem under $P^n_{\lambda_0,\delta_0}$, we have

$$
\begin{aligned}
\Lambda_{n,\lambda}(0, \frac{\gamma}{\sqrt{n}}) &\to_{\mathcal{L}} N\left(-\frac{\gamma^2}{2\sigma^2}E[X'_{t-1}\theta_{20}D_t(r_0, s_0)]^2, \frac{\gamma^2}{\sigma^2}E[X'_{t-1}\theta_{20}D_t(r_0, s_0)]^2\right) \\
&=_d N(-\frac{\mu^2}{2}, \mu^2),
\end{aligned}
$$

where $\mu^2 = \frac{\gamma^2}{\sigma^2}E[X'_{t-1}\theta_{20}D_t(r_0, s_0)]^2 = \frac{\gamma^2}{\sigma^2}\omega_1$ and ω_1 is defined as Theorem 3.1. It follows from Corollary 1 in Hájek et al. (1999) (pp. 253) that $P^n_{\lambda_0,\delta_0+\gamma/\sqrt{n}}$ is contiguous to $P^n_{\lambda_0,\delta_0}$. This completes the proof of Theorem 4.1. □

Proof of Theorem 4.2. (i). By (7.10) and (S3.1), we can show that under $P^n_{\lambda,0}$,

$$
[\frac{1}{\sqrt{n}}\frac{\partial L(0, \hat{\lambda}_n)}{\partial \delta}, \Lambda_{n,\lambda}(0, \frac{\gamma}{\sqrt{n}})]' \to_{\mathcal{L}} N(\tilde{\mu}_0, \tilde{\Sigma}_0)
$$

with

$$
\tilde{\mu}_0 = \begin{pmatrix} 0 \\ -\frac{\gamma^2}{2\sigma^2}\omega_1 \end{pmatrix} \quad \text{and} \quad \tilde{\Sigma}_0 = \begin{pmatrix} \sigma^2\omega_2 & \gamma\omega_2 \\ \gamma\omega_2 & \frac{\gamma^2}{\sigma^2}\omega_1 \end{pmatrix}.
$$

By (7.20), Theorem 4.1 and Le Cam's third Lemma in Hájek et al. (1999), under $P^n_{\lambda_0,\gamma/\sqrt{n}}$, we have

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

as $n \to \infty$. Under $P^n_{\lambda_0,0}$, by Lemma 7.1(i) and ergodic theorem, we have $\hat{\sigma}^2_{0n} \to_p \sigma^2$, $\hat{\omega}_{1n} \to_p \omega_1$ and $\hat{\omega}_{2n} \to_p \omega_2$ which are not random, then the covariance between $\Lambda_{n,\lambda}(0, \frac{\gamma}{\sqrt{n}})$ and each of $\hat{\sigma}^2_{0n}$, $\hat{\omega}_{1n}$ and $\hat{\omega}_{2n}$ will converge to 0 under $P^n_{\lambda_0,0}$. Then, by Le Cam's third Lemma, we also have $\hat{\sigma}^2_{0n} \to_p \sigma^2$, $\hat{\omega}_{1n} \to_p \omega_1$ and $\hat{\omega}_{2n} \to_p \omega_2$ under $P^n_{\lambda_0,\gamma/\sqrt{n}}$. Thus, (4.4) follows from continuous mapping and Slutsky theorems.
(ii) We fist prove that

$$\frac{1}{\sqrt{n}}\frac{\partial L(1,\hat{\theta}_n,s,\hat{r}_n)}{\partial \delta} \Longrightarrow \sigma Z(s)+\mu(s) \quad in \quad D[1/\bar{s},\bar{s}], \tag{S3.2}$$

where $Z(s)$ is defined in Theorem 3.2 and $\mu(s)$ will be given later.

Since we have shown the tightness of $\frac{1}{\sqrt{n}}\frac{\partial L(1,\hat{\theta}_n,s,\hat{r}_n)}{\partial \delta}$ in Theorem 3.2 under $P^n_{\lambda_0,1}$, by the contiguity and Lemma 4 in Hájek et al. (1999) (pp. 260), we also have the tightness of $\frac{1}{\sqrt{n}}\frac{\partial L(1,\hat{\theta}_n,s,\hat{r}_n)}{\partial \delta}$ under $P^n_{\lambda_0,1+\gamma/\sqrt{n}}$. By ergodic theorem and central limit theorem under $P^n_{\lambda_0,1}$, we can show that the finite dimensional distributions of $[\frac{1}{\sqrt{n}}\frac{\partial L(1,\hat{\theta}_n,s,\hat{r}_n)}{\partial \delta}, \Lambda_{n,\lambda}(1, 1+\frac{\gamma}{\sqrt{n}})]'$ converge weakly to a Gaussian process $\tilde{G}(s)$ with the mean and the covariance kernel

$$\tilde{\mu}_1 = \begin{pmatrix} 0 \\ -\frac{\gamma^2}{2\sigma^2}\omega_1 \end{pmatrix} \quad \text{and} \quad \tilde{\Sigma}_1(s,\tau) = \begin{pmatrix} \sigma^2 EZ(s)Z(\tau) & \gamma EZ(s)Z(s_0) \\ \gamma EZ(\tau)Z(s_0) & \frac{\gamma^2}{\sigma^2}\omega_1 \end{pmatrix},$$

for some $s, \tau \in [1/\bar{s}, \bar{s}]$ and a specified s_0 under the local alternative model with $\delta = 1 + \frac{\gamma}{\sqrt{n}}$. By Le Cam's third Lemma, we have (S3.2) with $\mu(s) = \gamma EZ(s)Z(s_0)$. By a similar argument as (i), we have $\hat{\sigma}^2_{1n} \to_p \sigma^2$ under both $P^n_{\lambda_0,1}$ and $P^n_{\lambda_0,1+\gamma/\sqrt{n}}$. Then, it follows from (S3.2) and continuous mapping theorem that (4.5) holds. This completes the proof of Theorem 4.2. □

S4 Simulated critical values c_α when testing $\tilde{H}_0$

For each $\bar{s}$, under $\tilde{H}_0$, 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, using one data set we simulate the critical values c_α with $\alpha = 0.1$, 0.05 and 0.01. Table 12 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 S12, 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 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 S13 summarizes the simulated critical values with $\bar{s} = 15$, 30 and 45, respectively. In practice, we should obtain the critical values for a fixed sample size according the the length of the data. For each $\bar{s}$, we choose $s_0 = 1$, 2, 5, 10 and 15 respectively in the LSTAR model.

Table S12: Simulated critical values c_α when testing $\tilde{H}_0$ 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

Table S13: Simulated critical values c_α when testing $\tilde{H}_0$.

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

S5 Further examinations of the real data examples

We further examine our tests by taking (6.1) and (6.2) in turn as a true model and using 1000 replications to verify the results obtained in Tables 8-9. In each replication, we generate a time series of sample sizes of 280 (equal to the original sample size) from models (6.1) and (6.2), respectively. The initial values $\{y_1, ..., y_{10}\}$ are taken from the original data. Table S14 summarizes the percentage of rejected times when testing H_0 and $\tilde{H}_0$, respectively. From Table S14, we can see substantial size distortion under H_0 while the size distortion under $\tilde{H}_0$ is minimal.

The above analysis suggests that the conclusion of Ekner and Nejstgaard (2013) is credible and we may conclude that of the two models, the TAR model is the more plausible.

Table S14: Percentage of rejection when testing H_0 (6.1) and $\tilde{H}_0$ (6.2). (POR=percentage of rejection).

	Testing H_0			Testing $\tilde{H}_0$			
Data	Model (6.1)			Model (6.2)			
α	0.1	0.05	0.01		0.1	0.05	0.01
				$\bar{s} = 15$	0.150	0.082	0.021
POR	0.230	0.159	0.060	$\bar{s} = 30$	0.127	0.064	0.011
				$\bar{s} = 45$	0.210	0.065	0.008

Let us examine our tests further by taking (6.3) and (6.4) in turn as a true model and using 1000 replications. In each replication, we generate a time series of sample sizes of 259 (equal to the original sample size) from models (6.3) and (6.4), respectively. The initial values $\{y_1, ..., y_{15}\}$ are taken from the original data. Table S15 summarizes the percentage of rejection when testing H_0 and $\tilde{H}_0$, respectively. From Table S15, we can see substantial size distortion in both cases, suggesting caution in interpreting the test results. As a matter of fact, the very large number of parameters for both models tends to suggest serious model over-parametrization.

Table S15: Percentage of rejection when testing H_0 (6.3) and $\tilde{H}_0$ (6.4). (POR=percentage of rejection).

	Testing H_0			Testing $\tilde{H}_0$			
Data	Model (6.3)			Model (6.4)			
α	0.1	0.05	0.01		0.1	0.05	0.01
				$\bar{s} = 15$	0.200	0.120	0.055
POR	0.210	0.140	0.040	$\bar{s} = 30$	0.235	0.130	0.045
				$\bar{s} = 45$	0.248	0.145	0.038

References

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

Hájek, J., Šidák, Z. and Sen, P. K. (1999). *Theory of rank tests.* Academic Press.

Hansen B. E. (1997) Inference in TAR models. *Studies in Nonlinear Dynamics & Econometrics*, **2(1)**, 1–14.