Supplemental material for
"*An over-rejection puzzle of bootstrap average tests for the no-threshold-effect hypothesis*"

Kaiji Motegi* – Kobe University

John W. Dennis† – Institute for Defense Analyses

This draft: April 1, 2024

1 Introduction

In this supplemental material, we present technical details and extended Monte Carlo simulations that are omitted from the main paper. In Section 2, we describe Tong's (1978) *Threshold Autoregressive* (TAR) model and Hansen's (1996) wild-bootstrap tests for the null hypothesis of no threshold effects. In Section 3, a full version of Monte Carlo simulations is presented. We use the following notation throughout this document: $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers, $\lfloor a \rfloor$ is the largest integer not larger than $a \in \mathbb{R}$, $\mathbf{1}(A)$ is the indicator function which equals 1 if event A occurs and 0 otherwise, $\#A$ is the number of elements of set A, and $A \times B$ is the Cartesian product of sets A and B.

2 Threshold Autoregression (TAR)

Tong's (1978) TAR model with two regimes and lag length $p = 1$ is specified as follows.

$$y_t = \begin{cases} \alpha_1 + \phi_1 y_{t-1} + u_t & \text{if } \; x_{t-d} < \mu, \\ \alpha_2 + \phi_2 y_{t-1} + u_t & \text{if } \; x_{t-d} \geq \mu, \qquad t \in \{1, \ldots, n\}, \end{cases} \tag{1}$$

* *Corresponding author*. Graduate School of Economics, Kobe University. Address: 2-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501 Japan. E-mail: motegi@econ.kobe-u.ac.jp

† Institute for Defense Analyses (IDA). Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of IDA. E-mail: jay.dennis@alumni.unc.edu

where y is called a target variable; x is called a threshold variable; u is an error term; $\boldsymbol{\beta}_r = (\alpha_r, \phi_r)^\top$ is a vector of regression parameters in regime $r \in \{1, 2\}$; $d \in \mathbb{N}$ is the delay parameter; $\mu \in \mathbb{R}$ is the threshold parameter; and n is the sample size. To rewrite (1) in a matrix form, stack the parameters as

$$\underbrace{\boldsymbol{\beta}_1}_{2\times 1} = \begin{bmatrix} \alpha_1 \\ \phi_1 \end{bmatrix}, \quad \underbrace{\boldsymbol{\beta}_2}_{2\times 1} = \begin{bmatrix} \alpha_2 \\ \phi_2 \end{bmatrix}, \quad \underbrace{\boldsymbol{\beta}}_{4\times 1} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \underbrace{\boldsymbol{\gamma}}_{2\times 1} = \begin{bmatrix} d \\ \mu \end{bmatrix}.$$

Define binary variables which express the regime:

$$I_{1t}(\mu) = \mathbf{1}(x_t < \mu), \qquad I_{2t}(\mu) = \mathbf{1}(x_t \geq \mu).$$

Stack the regressors as

$$\underbrace{\boldsymbol{z}_{t-1}}_{2\times 1} = \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix}, \qquad \underbrace{\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})}_{4\times 1} = \begin{bmatrix} \boldsymbol{z}_{t-1} I_{1,t-d}(\mu) \\ \boldsymbol{z}_{t-1} I_{2,t-d}(\mu) \end{bmatrix}.$$

Then, model (1) can be rewritten as a single equation:

$$y_t = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top \boldsymbol{\beta} + u_t. \tag{2}$$

To perform estimation and inference, one needs to specify the choice space of $\boldsymbol{\gamma}$. The choice space of d is $D = \{1, \dots, \bar{d}\}$, where $\bar{d} \in \mathbb{N}$ signifies the upper bound. For the choice space of μ, let $x_{[1]} \leq \cdots \leq x_{[n]}$ be a sorted version of x. In general, the space of μ is specified as

$$\mathcal{X}_{\kappa,n} = \left\{ x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]} \right\}, \tag{3}$$

where $\kappa \in [0, 1)$ signifies the fraction of $\#\mathcal{X}_{\kappa,n}$ to n. Given κ, each of the two regimes accounts for at least $50(1-\kappa)\%$ of the whole sample. Choosing a too large value for κ (e.g., $\kappa = 0.9$) may cause an identification problem in small samples. A common choice following a suggestion of Andrews (1993) is $\kappa = 0.7$, in which case each regime accounts for at least 15% of the entire sample. The space of $\boldsymbol{\gamma}$ is denoted as $\Gamma_{\kappa,n} = D \times \mathcal{X}_{\kappa,n}$; when there is no risk of confusion, we use a short-hand notation Γ instead of $\Gamma_{\kappa,n}$ for expositional simplicity.

Consider testing the null hypothesis of no threshold effects $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ versus a fixed alternative hypothesis $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$. Under H_0, the nuisance parameter $\boldsymbol{\gamma}$ is not identified. Hansen (1996) provides a well-known solution to this issue via wild-bootstrap tests, which proceed as follows. Fixing $\boldsymbol{\gamma} \in \Gamma$, the regressor vector $\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})$ can be computed from data; hence, the least squares estimator for $\boldsymbol{\beta}$ conditional on $\boldsymbol{\gamma}$ can be computed by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left\{ \sum_{t=1}^{n} \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top \right\}^{-1} \left\{ \sum_{t=1}^{n} \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) y_t \right\}.$$

The resulting residual is $\hat{u}_t(\boldsymbol{\gamma}) = y_t - \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top \hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$. The regression score associated with (2) is $\boldsymbol{s}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) u_t$. The estimated regression score under H_1 is $\hat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) \hat{u}_t(\boldsymbol{\gamma})$.

The no-threshold-effect hypothesis can be rewritten as $H_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{0}$, where the selection matrix is given by

$$\underbrace{\boldsymbol{R}}_{2\times 4} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

The conditional Wald test statistic given $\boldsymbol{\gamma} \in \Gamma$ is formulated as

$$\mathcal{W}_n(\boldsymbol{\gamma}) = n \hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})^\top \boldsymbol{R}^\top \left\{ \boldsymbol{R} \hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) \boldsymbol{R}^\top \right\}^{-1} \boldsymbol{R} \hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}).$$

If the error term u is assumed to be homoscedastic, then the covariance matrix estimator is given by $\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \hat{\sigma}^2(\boldsymbol{\gamma}) \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}$, where $\hat{\sigma}^2(\boldsymbol{\gamma}) = n^{-1} \sum_{t=1}^{n} \hat{u}_t^2(\boldsymbol{\gamma})$ and $\boldsymbol{M}_n(\boldsymbol{\gamma}) = n^{-1} \sum_{t=1}^{n} \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top$. Alternatively, a heteroscedasticity-robust covariance matrix estimator is given by $\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}) \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}$, where $\hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}) = n^{-1} \sum_{t=1}^{n} \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma})^\top$.

Aggregation of $\{\mathcal{W}_n(\boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \Gamma\}$ can be achieved in several ways. Three common approaches are supremum (sup-), average (ave-), and exponential (exp-) transformations:

$$\text{sup}\mathcal{W}_n \equiv \sup_{\boldsymbol{\gamma} \in \Gamma} \mathcal{W}_n(\boldsymbol{\gamma}), \tag{4}$$

$$\text{ave}\mathcal{W}_n \equiv \frac{1}{\#\Gamma} \sum_{\boldsymbol{\gamma} \in \Gamma} \mathcal{W}_n(\boldsymbol{\gamma}), \tag{5}$$

$$\text{exp}\mathcal{W}_n \equiv \ln \left[\frac{1}{\#\Gamma} \sum_{\boldsymbol{\gamma} \in \Gamma} \exp \left\{ \frac{\mathcal{W}_n(\boldsymbol{\gamma})}{2} \right\} \right]. \tag{6}$$

These are called sup-Wald, ave-Wald, and exp-Wald test statistics, respectively. Since $\boldsymbol{\gamma}$

is unidentified under H_0, the asymptotic null distributions of the test statistics (4)-(6) are non-standard. Hence, we should approximate p-values via wild bootstrap. Let $g(\mathcal{W}_n)$ denote either $\sup \mathcal{W}_n$, ave$\mathcal{W}_n$, or exp$\mathcal{W}_n$, then proceed as follows:

Step 1 For each $b \in \{1, \ldots, B\}$, generate $\xi_t^{(b)} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ with $t \in \{1, \ldots, n\}$.

Step 2 Compute a bootstrap test statistic $g\{\mathcal{W}_n^{(b)}\}$, where

$$\mathcal{W}_n^{(b)}(\boldsymbol{\gamma}) = \hat{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma})^\top \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \boldsymbol{R}^\top \left\{ \boldsymbol{R} \hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) \boldsymbol{R}^\top \right\}^{-1} \boldsymbol{R} \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \hat{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}); \tag{7}$$

$$\hat{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) \xi_t^{(b)}. \tag{8}$$

Step 3 Repeat Steps 1-2 independently, resulting in $g\{\mathcal{W}_n^{(1)}\}, \ldots, g\{\mathcal{W}_n^{(B)}\}$.

Step 4 Compute the bootstrap p-value:

$$\hat{p}_n^B(H_0) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1} \left[g\left\{ \mathcal{W}_n^{(b)} \right\} \geq g(\mathcal{W}_n) \right].$$

Reject H_0 at the $100a\%$ level if $\hat{p}_n^B(H_0) < a$, where $a \in (0,1)$ is a nominal size.

The testing procedure is analogous when the Wald test is replaced with the Lagrange Multiplier (LM) test. Let $\tilde{u}_t$ be the least squares residual from (2) with H_0 being imposed (i.e., a single-regime AR(1) model), then the estimated regression score under H_0 is $\tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})\tilde{u}_t$. If u is assumed to be homoscedastic, then we use $\tilde{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \tilde{\sigma}^2 \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}$, where $\tilde{\sigma}^2 = n^{-1}\sum_{t=1}^{n} \tilde{u}_t^2$. Alternatively, a heteroscedasticity-robust covariance matrix estimator is $\tilde{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \, \tilde{\boldsymbol{S}}_n(\boldsymbol{\gamma}) \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}$, where $\tilde{\boldsymbol{S}}_n(\boldsymbol{\gamma}) = n^{-1}\sum_{t=1}^{n} \tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma}) \tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma})^\top$. The conditional LM test statistic is given by

$$\mathcal{LM}_n(\boldsymbol{\gamma}) = n\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})^\top \boldsymbol{R}^\top \left\{ \boldsymbol{R} \tilde{\boldsymbol{V}}_n(\boldsymbol{\gamma}) \boldsymbol{R}^\top \right\}^{-1} \boldsymbol{R} \hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}).$$

Sup-LM, ave-LM, and exp-LM test statistics are obtained via (4)-(6), where $\mathcal{W}_n(\boldsymbol{\gamma})$ is replaced with $\mathcal{LM}_n(\boldsymbol{\gamma})$. Steps 1-4 are executed with (7)-(8) being replaced with

$$\mathcal{LM}_n^{(b)}(\boldsymbol{\gamma}) = \tilde{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma})^\top \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \boldsymbol{R}^\top \left\{ \boldsymbol{R} \tilde{\boldsymbol{V}}_n(\boldsymbol{\gamma}) \boldsymbol{R}^\top \right\}^{-1} \boldsymbol{R} \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \tilde{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}),$$

$$\tilde{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma}) \xi_t^{(b)}.$$

3 Monte Carlo simulation

In this section, we conduct a full version of Monte Carlo simulations. We explain our simulation design in Section 3.1, and discuss our main results in Section 3.2. Further simulation evidence is presented in Section 3.3.

3.1 Simulation design

Consider the following data generating process (DGP):

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \phi_0 & 0 \\ 0 & \psi_0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \nu_t \end{bmatrix}, \qquad \begin{bmatrix} \epsilon_t \\ \nu_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right). \tag{9}$$

The target variable y and the threshold variable x follow mutually independent single-regime AR(1) processes. Consider low, medium, or high persistence in each of y and x: $\phi_0, \psi_0 \in \{0.3, 0.6, 0.9\}$. The joint standard normality of the true errors is a conventional assumption which simplifies analysis. Generate $J = 1000$ Monte Carlo samples of size $n \in \{125, 250, 500, 1000\}$ from DGP (9).

For each sample generated from (9), fit a two-regime TAR model with lag length $p = 1$ as described in Section 2. The null hypothesis of no threshold effects is expressed as H_0 : $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, where $\boldsymbol{\beta}_r = (\alpha_r, \phi_r)^\top$ with $r \in \{1, 2\}$. Since the DGP is given by (9), H_0 is true in our experiment. We test H_0 via Hansen's (1996) wild-bootstrap tests with six test statistics: sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM statistics. The non-robust version of covariance matrix $\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma})$ is used, which is a correct choice as DGP (9) has homoscedastic errors.[1] Nominal size is determined from $a \in \{0.01, 0.05, 0.10\}$, and the number of bootstrap iterations is $B = 500$.

When implementing the bootstrap tests, the choice space of d is specified as $D = \{1, \dots, \bar{d}\}$ with the upper bound being $\bar{d} \in \{1, 2, 4, 8\}$. Specifically, D is either $\{1\}$, $\{1, 2\}$, $\{1, 2, 3, 4\}$, or $\{1, 2, 3, 4, 5, 6, 7, 8\}$. The choice space of μ is determined by choosing $\kappa \in \{0, 0.35, 0.7\}$ in (3). When $\kappa = 0$, we have that $\mathcal{X}_{\kappa,n} = \{x_{[0.5n]}\}$; that is to say, μ is fixed at (almost) the median value of x. When $\kappa = 0.35$, $\mathcal{X}_{\kappa,n}$ consists of observations above the 32.5 percentile and below the 67.5 percentile of x. When $\kappa = 0.7$, $\mathcal{X}_{\kappa,n}$ consists of observations above the 15 percentile and below the 85 percentile of x; this case matches the suggestion of Andrews (1993). In Table 1, we show specific values of $\mathcal{X}_{\kappa,n}$ and its cardinality for each pair

[1] We also used the heteroscedasticity-robust covariance matrix, getting similar results.

of $\kappa \in \{0, 0.35, 0.7\}$ and $n \in \{125, 250, 500, 1000\}$.

In finite samples, there is expected to be a bias-variance trade-off on the cardinality of $\Gamma_{\kappa,n} = D \times \mathcal{X}_{\kappa,n}$. Under H_0, the finite sample performance of the bootstrap tests should become worse due to larger variance as $\#\Gamma_{\kappa,n}$ increases. Under H_1, asymptotic bias arises if a true value of $\boldsymbol{\gamma}$ is not included in $\Gamma_{\kappa,n}$. In our current set-up, threshold effects are absent and H_0 is true given DGP (9). Hence, the empirical size of the bootstrap tests should be closest to the nominal size a for $(\bar{d}, \kappa) = (1, 0)$ and farthest for $(\bar{d}, \kappa) = (8, 0.7)$, particularly for small sample sizes. When the sample size is large enough (say $n = 1000$), the type-I error rate should be sufficiently close to a for any $\bar{d} \in \{1, 2, 4, 8\}$ and $\kappa \in \{0, 0.35, 0.7\}$. Keeping this conjecture in mind, we report rejection frequencies in the next section.

In summary, the simulation design of this supplemental material is more general than that of the main paper in four aspects. First, the degree of persistence in y is specified from $\phi_0 \in \{0.3, 0.6, 0.9\}$, while we focus on $\phi_0 = 0.6$ in the main paper. Second, we consider all of the sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests, whereas the supremum tests are omitted from the main paper. Third, the nominal size is determined from $a \in \{0.01, 0.05, 0.10\}$, while it is fixed at $a = 0.05$ in the main paper. Fourth, the fraction parameter for the space of μ is chosen from $\kappa \in \{0, 0.35, 0.7\}$, whereas we restrict our attention to $\kappa = 0.7$ in the main paper.

3.2 Simulation results

For $\phi_0 = \psi_0 = 0.3$, rejection frequencies associated with $\bar{d} \in \{1, 2, 4, 8\}$ are reported in Tables 2-5, respectively. The Wald tests tend to over-reject the correct null hypothesis of no threshold effects, $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, in small samples. When $(\bar{d}, \kappa, n, a) = (1, 0.7, 125, 0.05)$, the empirical size is $\{0.083, 0.076, 0.078\}$ for the sup-Wald, ave-Wald, and exp-Wald tests, respectively (Table 2). These over-rejections vanish as the sample size increases. The LM tests lead to more accurate empirical size in small samples. Continuing the same configuration, the empirical size is $\{0.045, 0.048, 0.048\}$ for the sup-LM, ave-LM, and exp-LM tests. When we raise the upper bound of D from $\bar{d} = 1$ to $\bar{d} = 8$, the LM tests become slightly conservative in small samples, but that is quickly corrected as n grows. The empirical size of the ave-LM test, for instance, is $\{0.038, 0.048, 0.039, 0.054\}$ for $n \in \{125, 250, 500, 1000\}$, respectively (Table 5). In what follows, we basically restrict our attention to the LM tests, and refer to the Wald tests whenever necessary.

Rejection frequencies under $(\phi_0, \psi_0) = (0.3, 0.6)$ are shown in Tables 6-9. As the persistence of x increases from $\psi_0 = 0.3$ to $\psi_0 = 0.6$, the ave-LM test starts to over-reject H_0 for large $\bar{d}$. Focusing on $(\bar{d}, \kappa, a) = (8, 0.7, 0.05)$, the empirical size of the ave-LM test is $\{0.063, 0.072, 0.092, 0.091\}$ for $n \in \{125, 250, 500, 1000\}$, respectively (Table 9). The tendency to over-reject H_0 becomes more salient when we further raise the persistence of x from $\psi_0 = 0.6$ to $\psi_0 = 0.9$; see Tables 10-13. Keeping $(\bar{d}, \kappa, a) = (8, 0.7, 0.05)$, the empirical size of the ave-LM test is $\{0.132, 0.141, 0.141, 0.159\}$ (Table 13). These results suggest that the ave-LM test loses control for the type-I error rate when x is persistent and the upper bound $\bar{d}$ is large. It is particularly surprising that the size distortion does not vanish in large samples; the rejection frequency exceeds the nominal size 5% by 10.9% for $n = 1000$. This indicates that the ave-LM test might converge to an unintended null distribution.

We can evaluate the roles of the persistence of y by comparing the results with $\phi_0 = 0.3$ (Tables 2-13), results with $\phi_0 = 0.6$ (Tables 14-25), and results with $\phi_0 = 0.9$ (Tables 26-37). The Wald tests suffer from over-rejections in small samples when both y and x are persistent. Focusing on $(\phi_0, \psi_0, \bar{d}, \kappa, a) = (0.9, 0.9, 1, 0.7, 0.05)$, the empirical size with $n \in \{125, 250, 500, 1000\}$ is $\{0.135, 0.074, 0.065, 0.060\}$ for the sup-Wald test, while it is $\{0.057, 0.048, 0.053, 0.053\}$ for the sup-LM test (Table 34). For each $\phi_0 \in \{0.3, 0.6, 0.9\}$, we keep observing the puzzling size distortion of the ave-LM test under large ψ_0 and large $\bar{d}$. Given $(\psi_0, \bar{d}, \kappa, n, a) = (0.9, 8, 0.7, 1000, 0.05)$, the empirical size of the ave-LM test is $\{0.159, 0.163, 0.153\}$ for $\phi_0 \in \{0.3, 0.6, 0.9\}$, respectively (Tables 13, 25, and 37). In what follows, we focus on the intermediate case (i.e., $\phi_0 = 0.6$) and draw further implications on the size distortion of the ave-LM test.

In view of Tables 14-25, our key findings are summarized as follows. First, the ave-LM test over-rejects the true no-threshold-effect hypothesis H_0 when the threshold variable x is sufficiently persistent and the upper bound of the choice space for the delay parameter, $\bar{d}$, is sufficiently large. Second, the ave-Wald test also produces size distortions when both ψ_0 and $\bar{d}$ are large. In previous simulation studies such as Hansen (1996, Table II) and Ahmad and Donayre (2016, Table 1), the average tests with persistent x and large $\bar{d}$ are not covered. Our paper is likely the first work that inspects the relevant case.

Third, the size distortion of the average tests exists for any nominal sizes considered. Taking $(\psi_0, \bar{d}, \kappa, n) = (0.9, 8, 0.7, 1000)$ as an example, the empirical size of the ave-LM test is $\{0.088, 0.163, 0.200\}$ for $a \in \{0.01, 0.05, 0.10\}$, respectively (Table 25). The size distortion is substantially large for each significance level compared with the fact that our DGP and model are quite simple. Fourth, narrowing down the choice space of μ alleviates the size distortion only modestly. Let $(\psi_0, \bar{d}, n, a) = (0.9, 8, 1000, 0.05)$, then the empirical size of the ave-LM test is $\{0.133, 0.150, 0.163\}$ for $\kappa \in \{0, 0.35, 0.7\}$, respectively (Table 25). It is reasonable that the smaller value of κ leads to the sharper inference under H_0 due to decreased variance. Nevertheless, it is curious that the ave-LM test produces the excess rejection rate of 8.3% even when we fix the threshold at $\mu = x_{[500]}$ (i.e., $\kappa = 0$).

Fifth and importantly, the sup-LM and exp-LM tests achieve accurate size. When $(\psi_0, \bar{d}, \kappa, a) = (0.9, 8, 0.7, 0.05)$, the empirical size corresponding to $n \in \{125, 250, 500, 1000\}$ is $\{0.018, 0.026, 0.034, 0.036\}$ for the sup-LM test and $\{0.025, 0.033, 0.044, 0.050\}$ for the exp-LM test (Table 25). While we do not know what distorts the size of the average tests only, a practical solution is to use the sup-LM or exp-LM test, since their size is sufficiently close to the nominal size in both small and large samples. The sup-Wald and exp-Wald tests perform well in large samples, but their small sample performance is not satisfactory when both ϕ_0 and ψ_0 take large values; recall that the Wald tests suffer from over-rejections when $\phi_0 = \psi_0 = 0.9$ and $n = 125$ (Tables 34-37). Only the ave-Wald and ave-LM tests produce size distortions in large samples, and it is an interesting future task to explain why.

3.3 Further simulation evidence

To further characterize the over-rejection puzzle of the bootstrap average tests, we present additional simulation evidence. We restrict our attention to the ave-Wald and ave-LM tests with $(\phi_0, \psi_0, \kappa, n) = (0.6, 0.9, 0, 1000)$, a main scenario in which the over-rejection puzzle emerges. Recall that D signifies the choice space of the delay parameter d. In the previous sections, the lower bound of D was always fixed at 1. Consequently, an exact role of D remains unclear. Which distorts the size of the average tests: a large cardinality of D or a large candidate value for d? To answer this question, we consider some additional specifications for D with various lower and upper bounds: $\{1\}$, $\{4\}$, $\{8\}$, $\{1, 2\}$, $\{4, 5\}$, $\{7, 8\}$, $\{1, 2, 3, 4\}$, $\{2, 4, 6, 8\}$, $\{5, 6, 7, 8\}$, and $\{1, 2, 3, 4, 5, 6, 7, 8\}$. In the previous sections, we considered $\{1\}$, $\{1, 2\}$, $\{1, 2, 3, 4\}$, and $\{1, 2, 3, 4, 5, 6, 7, 8\}$, but did not consider the other six cases. Other configurations are the same as in the previous sections.

The resulting rejection frequencies are reported in Table 38. It is the cardinality of D, *not* a specific value for candidate d, that plays a key role in the over-rejection puzzle. When $\#D = 1$, the empirical size of the average tests is almost identical to the nominal size. The empirical size of the ave-LM test with $D = \{8\}$, for instance, is $\{0.014, 0.054, 0.112\}$ for $a \in \{0.01, 0.05, 0.10\}$, respectively. As the dimension of D increases, the average tests over-reject correct H_0 more frequently. The excess rejection rates are around $\{2\%, 5\%, 8\%\}$ for $\#D \in \{2, 4, 8\}$, respectively. Note that, conditional on the value of $\#D$, we observe similar rejection rates irrespective of the specific candidate values contained in D.

References

Ahmad, Y., and L. Donayre (2016): "Outliers and persistence in threshold autoregressive processes," *Studies in Nonlinear Dynamics & Econometrics*, 20, 37–56.

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Hansen, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64(2), 413–430.

Tong, H. (1978): "On a threshold model," in *Pattern Recognition and Signal Processing*, ed. by C. H. Chen. Sijthoff and Noordhoff, Amsterdam.

Table 1: Specification of the choice space of threshold parameter μ

κ	n	$\mathcal{X}_{\kappa,n}$	$\#\mathcal{X}_{\kappa,n}$
0	125	$\{x_{[62]}\}$	1
0	250	$\{x_{[125]}\}$	1
0	500	$\{x_{[250]}\}$	1
0	1000	$\{x_{[500]}\}$	1
0.35	125	$\{x_{[40]}, \ldots, x_{[84]}\}$	45
0.35	250	$\{x_{[81]}, \ldots, x_{[168]}\}$	88
0.35	500	$\{x_{[162]}, \ldots, x_{[337]}\}$	176
0.35	1000	$\{x_{[325]}, \ldots, x_{[675]}\}$	351
0.7	125	$\{x_{[18]}, \ldots, x_{[106]}\}$	89
0.7	250	$\{x_{[37]}, \ldots, x_{[212]}\}$	176
0.7	500	$\{x_{[75]}, \ldots, x_{[425]}\}$	351
0.7	1000	$\{x_{[150]}, \ldots, x_{[850]}\}$	701

The choice space of threshold parameter μ is specified as $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x; κ signifies the fraction of $\#\mathcal{X}_{\kappa,n}$ to the sample size n. This table reports specific values of $\mathcal{X}_{\kappa,n}$ and its cardinality for each pair of $\kappa \in \{0, 0.35, 0.7\}$ and $n \in \{125, 250, 500, 1000\}$.

Table 2: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.3$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.013, 0.060, 0.123	0.011, 0.053, 0.106	0.012, 0.045, 0.101	0.004, 0.050, 0.099
0.00	ave-W	0.013, 0.060, 0.123	0.011, 0.053, 0.106	0.012, 0.045, 0.101	0.004, 0.050, 0.099
0.00	exp-W	0.013, 0.060, 0.123	0.011, 0.053, 0.106	0.012, 0.045, 0.101	0.004, 0.050, 0.099
0.00	sup-LM	0.006, 0.053, 0.103	0.012, 0.045, 0.098	0.011, 0.043, 0.093	0.006, 0.050, 0.097
0.00	ave-LM	0.006, 0.053, 0.103	0.012, 0.045, 0.098	0.011, 0.043, 0.093	0.006, 0.050, 0.097
0.00	exp-LM	0.006, 0.053, 0.103	0.012, 0.045, 0.098	0.011, 0.043, 0.093	0.006, 0.050, 0.097
0.35	sup-W	0.013, 0.061, 0.121	0.016, 0.068, 0.123	0.009, 0.048, 0.098	0.008, 0.051, 0.114
0.35	ave-W	0.011, 0.048, 0.112	0.011, 0.058, 0.118	0.011, 0.050, 0.097	0.007, 0.046, 0.101
0.35	exp-W	0.013, 0.048, 0.113	0.012, 0.056, 0.123	0.010, 0.049, 0.094	0.008, 0.046, 0.100
0.35	sup-LM	0.006, 0.040, 0.088	0.009, 0.055, 0.113	0.009, 0.046, 0.094	0.009, 0.044, 0.111
0.35	ave-LM	0.007, 0.038, 0.093	0.009, 0.052, 0.107	0.010, 0.047, 0.096	0.006, 0.046, 0.101
0.35	exp-LM	0.006, 0.037, 0.088	0.008, 0.049, 0.111	0.010, 0.046, 0.089	0.008, 0.045, 0.100
0.70	sup-W	0.023, 0.083, 0.150	0.021, 0.076, 0.140	0.011, 0.062, 0.119	0.006, 0.042, 0.096
0.70	ave-W	0.021, 0.076, 0.151	0.016, 0.076, 0.123	0.009, 0.051, 0.102	0.005, 0.037, 0.095
0.70	exp-W	0.021, 0.078, 0.141	0.021, 0.065, 0.132	0.013, 0.056, 0.111	0.005, 0.039, 0.097
0.70	sup-LM	0.004, 0.045, 0.099	0.014, 0.059, 0.108	0.008, 0.055, 0.110	0.004, 0.038, 0.089
0.70	ave-LM	0.010, 0.048, 0.111	0.014, 0.065, 0.115	0.003, 0.043, 0.096	0.005, 0.035, 0.091
0.70	exp-LM	0.003, 0.048, 0.100	0.013, 0.053, 0.111	0.009, 0.053, 0.102	0.005, 0.033, 0.089

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 3: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.3$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.022, 0.073, 0.124	0.008, 0.045, 0.106	0.010, 0.040, 0.080	0.011, 0.045, 0.093
0.00	ave-W	0.019, 0.076, 0.128	0.007, 0.050, 0.114	0.008, 0.048, 0.099	0.012, 0.056, 0.101
0.00	exp-W	0.022, 0.072, 0.123	0.007, 0.044, 0.108	0.010, 0.042, 0.081	0.012, 0.048, 0.100
0.00	sup-LM	0.013, 0.057, 0.111	0.005, 0.042, 0.097	0.008, 0.038, 0.077	0.010, 0.044, 0.091
0.00	ave-LM	0.014, 0.060, 0.113	0.004, 0.043, 0.105	0.006, 0.041, 0.090	0.011, 0.054, 0.100
0.00	exp-LM	0.014, 0.056, 0.108	0.005, 0.042, 0.098	0.009, 0.042, 0.078	0.010, 0.044, 0.098
0.35	sup-W	0.016, 0.072, 0.139	0.010, 0.057, 0.120	0.016, 0.062, 0.117	0.012, 0.059, 0.103
0.35	ave-W	0.018, 0.071, 0.133	0.011, 0.050, 0.105	0.017, 0.056, 0.100	0.010, 0.041, 0.093
0.35	exp-W	0.017, 0.069, 0.132	0.009, 0.053, 0.110	0.012, 0.059, 0.097	0.010, 0.055, 0.093
0.35	sup-LM	0.009, 0.045, 0.102	0.007, 0.043, 0.100	0.011, 0.056, 0.104	0.011, 0.054, 0.099
0.35	ave-LM	0.012, 0.050, 0.106	0.009, 0.046, 0.095	0.016, 0.056, 0.095	0.009, 0.039, 0.094
0.35	exp-LM	0.010, 0.043, 0.101	0.006, 0.040, 0.089	0.009, 0.055, 0.091	0.010, 0.051, 0.092
0.70	sup-W	0.018, 0.082, 0.143	0.013, 0.058, 0.114	0.007, 0.051, 0.109	0.006, 0.049, 0.098
0.70	ave-W	0.021, 0.084, 0.128	0.016, 0.061, 0.120	0.013, 0.057, 0.110	0.005, 0.040, 0.087
0.70	exp-W	0.019, 0.077, 0.139	0.008, 0.061, 0.107	0.007, 0.046, 0.099	0.006, 0.035, 0.098
0.70	sup-LM	0.002, 0.036, 0.084	0.005, 0.044, 0.082	0.004, 0.040, 0.097	0.005, 0.041, 0.093
0.70	ave-LM	0.009, 0.057, 0.095	0.009, 0.049, 0.102	0.008, 0.051, 0.101	0.004, 0.037, 0.081
0.70	exp-LM	0.004, 0.037, 0.095	0.004, 0.042, 0.086	0.005, 0.039, 0.092	0.006, 0.029, 0.094

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 4: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.3$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.019, 0.076, 0.137	0.010, 0.056, 0.120	0.012, 0.049, 0.094	0.011, 0.052, 0.091
0.00	ave-W	0.018, 0.068, 0.133	0.013, 0.059, 0.124	0.016, 0.055, 0.097	0.012, 0.057, 0.112
0.00	exp-W	0.019, 0.075, 0.141	0.010, 0.054, 0.120	0.012, 0.050, 0.099	0.011, 0.053, 0.091
0.00	sup-LM	0.008, 0.049, 0.109	0.005, 0.046, 0.102	0.010, 0.048, 0.091	0.008, 0.051, 0.085
0.00	ave-LM	0.008, 0.053, 0.098	0.011, 0.045, 0.109	0.013, 0.052, 0.094	0.012, 0.055, 0.109
0.00	exp-LM	0.009, 0.049, 0.110	0.005, 0.043, 0.108	0.010, 0.050, 0.089	0.008, 0.050, 0.088
0.35	sup-W	0.018, 0.070, 0.135	0.020, 0.067, 0.137	0.013, 0.057, 0.115	0.009, 0.047, 0.103
0.35	ave-W	0.021, 0.071, 0.127	0.016, 0.071, 0.128	0.012, 0.056, 0.101	0.014, 0.054, 0.103
0.35	exp-W	0.018, 0.068, 0.124	0.016, 0.072, 0.121	0.012, 0.055, 0.102	0.008, 0.048, 0.099
0.35	sup-LM	0.007, 0.029, 0.081	0.011, 0.050, 0.103	0.010, 0.044, 0.105	0.006, 0.048, 0.092
0.35	ave-LM	0.015, 0.052, 0.097	0.013, 0.060, 0.113	0.009, 0.051, 0.092	0.013, 0.054, 0.097
0.35	exp-LM	0.009, 0.028, 0.080	0.008, 0.054, 0.101	0.010, 0.047, 0.094	0.007, 0.046, 0.097
0.70	sup-W	0.007, 0.054, 0.128	0.013, 0.066, 0.132	0.008, 0.054, 0.126	0.017, 0.059, 0.116
0.70	ave-W	0.019, 0.085, 0.149	0.016, 0.077, 0.144	0.017, 0.068, 0.123	0.017, 0.072, 0.128
0.70	exp-W	0.005, 0.053, 0.137	0.012, 0.062, 0.128	0.010, 0.058, 0.122	0.015, 0.058, 0.117
0.70	sup-LM	0.001, 0.013, 0.049	0.006, 0.039, 0.093	0.004, 0.038, 0.099	0.015, 0.052, 0.108
0.70	ave-LM	0.004, 0.044, 0.104	0.010, 0.050, 0.118	0.013, 0.059, 0.119	0.014, 0.062, 0.119
0.70	exp-LM	0.001, 0.014, 0.061	0.005, 0.036, 0.094	0.009, 0.039, 0.099	0.014, 0.051, 0.110

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 5: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.3$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.014, 0.071, 0.130	0.009, 0.051, 0.108	0.005, 0.055, 0.114	0.007, 0.052, 0.110
0.00	ave-W	0.017, 0.087, 0.157	0.013, 0.067, 0.117	0.012, 0.066, 0.127	0.009, 0.054, 0.105
0.00	exp-W	0.015, 0.068, 0.141	0.008, 0.051, 0.114	0.006, 0.054, 0.117	0.007, 0.052, 0.102
0.00	sup-LM	0.009, 0.050, 0.094	0.005, 0.035, 0.090	0.005, 0.048, 0.104	0.007, 0.051, 0.101
0.00	ave-LM	0.008, 0.059, 0.129	0.005, 0.047, 0.103	0.011, 0.058, 0.116	0.008, 0.048, 0.102
0.00	exp-LM	0.008, 0.047, 0.100	0.004, 0.039, 0.093	0.005, 0.048, 0.110	0.007, 0.049, 0.098
0.35	sup-W	0.014, 0.076, 0.144	0.016, 0.063, 0.119	0.015, 0.070, 0.126	0.013, 0.062, 0.101
0.35	ave-W	0.007, 0.065, 0.139	0.017, 0.067, 0.123	0.017, 0.068, 0.135	0.011, 0.057, 0.114
0.35	exp-W	0.013, 0.057, 0.133	0.016, 0.057, 0.107	0.014, 0.064, 0.120	0.011, 0.057, 0.103
0.35	sup-LM	0.002, 0.028, 0.079	0.009, 0.046, 0.087	0.009, 0.058, 0.111	0.012, 0.057, 0.096
0.35	ave-LM	0.002, 0.035, 0.103	0.011, 0.056, 0.108	0.015, 0.058, 0.120	0.008, 0.053, 0.111
0.35	exp-LM	0.002, 0.026, 0.065	0.011, 0.047, 0.081	0.013, 0.059, 0.107	0.008, 0.051, 0.096
0.70	sup-W	0.014, 0.068, 0.143	0.018, 0.068, 0.123	0.007, 0.046, 0.093	0.011, 0.057, 0.131
0.70	ave-W	0.019, 0.076, 0.147	0.014, 0.066, 0.122	0.008, 0.049, 0.109	0.013, 0.058, 0.135
0.70	exp-W	0.012, 0.061, 0.131	0.017, 0.059, 0.120	0.008, 0.044, 0.094	0.009, 0.053, 0.117
0.70	sup-LM	0.000, 0.016, 0.054	0.005, 0.034, 0.087	0.004, 0.038, 0.070	0.011, 0.043, 0.114
0.70	ave-LM	0.005, 0.038, 0.082	0.008, 0.048, 0.094	0.006, 0.039, 0.092	0.013, 0.054, 0.126
0.70	exp-LM	0.000, 0.020, 0.046	0.008, 0.037, 0.078	0.006, 0.037, 0.068	0.009, 0.047, 0.107

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 6: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.6$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.016, 0.064, 0.124	0.010, 0.065, 0.115	0.009, 0.049, 0.106	0.010, 0.051, 0.107
0.00	ave-W	0.016, 0.064, 0.124	0.010, 0.065, 0.115	0.009, 0.049, 0.106	0.010, 0.051, 0.107
0.00	exp-W	0.016, 0.064, 0.124	0.010, 0.065, 0.115	0.009, 0.049, 0.106	0.010, 0.051, 0.107
0.00	sup-LM	0.010, 0.054, 0.112	0.008, 0.057, 0.109	0.007, 0.044, 0.100	0.011, 0.051, 0.105
0.00	ave-LM	0.010, 0.054, 0.112	0.008, 0.057, 0.109	0.007, 0.044, 0.100	0.011, 0.051, 0.105
0.00	exp-LM	0.010, 0.054, 0.112	0.008, 0.057, 0.109	0.007, 0.044, 0.100	0.011, 0.051, 0.105
0.35	sup-W	0.023, 0.081, 0.147	0.022, 0.072, 0.140	0.010, 0.050, 0.097	0.013, 0.049, 0.091
0.35	ave-W	0.022, 0.072, 0.126	0.017, 0.067, 0.125	0.006, 0.046, 0.092	0.009, 0.048, 0.096
0.35	exp-W	0.024, 0.072, 0.137	0.022, 0.067, 0.126	0.008, 0.047, 0.093	0.012, 0.049, 0.099
0.35	sup-LM	0.009, 0.053, 0.119	0.015, 0.060, 0.123	0.008, 0.047, 0.092	0.011, 0.048, 0.089
0.35	ave-LM	0.010, 0.054, 0.110	0.014, 0.060, 0.117	0.005, 0.043, 0.088	0.009, 0.046, 0.098
0.35	exp-LM	0.012, 0.053, 0.112	0.017, 0.056, 0.118	0.005, 0.043, 0.089	0.012, 0.048, 0.097
0.70	sup-W	0.011, 0.067, 0.148	0.010, 0.057, 0.128	0.007, 0.051, 0.111	0.009, 0.053, 0.109
0.70	ave-W	0.015, 0.069, 0.142	0.009, 0.047, 0.101	0.005, 0.051, 0.108	0.015, 0.049, 0.098
0.70	exp-W	0.010, 0.066, 0.139	0.007, 0.050, 0.106	0.009, 0.051, 0.103	0.008, 0.050, 0.100
0.70	sup-LM	0.002, 0.028, 0.087	0.008, 0.037, 0.090	0.006, 0.044, 0.102	0.008, 0.048, 0.099
0.70	ave-LM	0.008, 0.045, 0.110	0.004, 0.036, 0.085	0.005, 0.047, 0.099	0.014, 0.046, 0.093
0.70	exp-LM	0.003, 0.031, 0.092	0.006, 0.035, 0.084	0.007, 0.047, 0.094	0.007, 0.047, 0.097

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 7: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.6$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.013, 0.061, 0.112	0.013, 0.058, 0.109	0.011, 0.049, 0.107	0.009, 0.046, 0.094
0.00	ave-W	0.023, 0.066, 0.122	0.022, 0.066, 0.129	0.013, 0.058, 0.103	0.013, 0.054, 0.108
0.00	exp-W	0.014, 0.062, 0.115	0.014, 0.062, 0.110	0.011, 0.054, 0.110	0.009, 0.044, 0.099
0.00	sup-LM	0.007, 0.044, 0.093	0.011, 0.053, 0.101	0.011, 0.043, 0.106	0.009, 0.043, 0.092
0.00	ave-LM	0.013, 0.052, 0.106	0.017, 0.062, 0.120	0.013, 0.055, 0.101	0.013, 0.052, 0.106
0.00	exp-LM	0.007, 0.046, 0.097	0.012, 0.052, 0.099	0.010, 0.048, 0.104	0.009, 0.041, 0.096
0.35	sup-W	0.015, 0.072, 0.127	0.021, 0.078, 0.125	0.010, 0.054, 0.108	0.010, 0.046, 0.096
0.35	ave-W	0.014, 0.080, 0.140	0.024, 0.075, 0.130	0.016, 0.071, 0.103	0.016, 0.056, 0.097
0.35	exp-W	0.013, 0.065, 0.126	0.022, 0.066, 0.120	0.011, 0.049, 0.101	0.008, 0.039, 0.090
0.35	sup-LM	0.004, 0.041, 0.092	0.013, 0.059, 0.106	0.007, 0.046, 0.095	0.009, 0.040, 0.087
0.35	ave-LM	0.007, 0.055, 0.120	0.021, 0.065, 0.117	0.013, 0.064, 0.101	0.015, 0.055, 0.094
0.35	exp-LM	0.003, 0.038, 0.102	0.012, 0.058, 0.109	0.006, 0.046, 0.093	0.008, 0.037, 0.086
0.70	sup-W	0.019, 0.083, 0.143	0.008, 0.065, 0.126	0.013, 0.059, 0.112	0.014, 0.056, 0.112
0.70	ave-W	0.024, 0.096, 0.161	0.032, 0.081, 0.136	0.030, 0.088, 0.136	0.027, 0.070, 0.122
0.70	exp-W	0.019, 0.081, 0.137	0.014, 0.062, 0.121	0.014, 0.062, 0.111	0.017, 0.064, 0.099
0.70	sup-LM	0.006, 0.032, 0.085	0.004, 0.046, 0.096	0.010, 0.048, 0.098	0.010, 0.056, 0.108
0.70	ave-LM	0.015, 0.058, 0.120	0.018, 0.066, 0.121	0.027, 0.081, 0.126	0.027, 0.065, 0.114
0.70	exp-LM	0.007, 0.041, 0.087	0.004, 0.046, 0.096	0.012, 0.054, 0.096	0.012, 0.056, 0.095

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 8: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.6$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.013, 0.071, 0.131	0.014, 0.062, 0.112	0.011, 0.061, 0.104	0.009, 0.054, 0.107
0.00	ave-W	0.035, 0.087, 0.152	0.023, 0.076, 0.129	0.024, 0.075, 0.119	0.023, 0.080, 0.131
0.00	exp-W	0.014, 0.082, 0.133	0.014, 0.066, 0.114	0.012, 0.061, 0.117	0.009, 0.060, 0.112
0.00	sup-LM	0.006, 0.053, 0.107	0.012, 0.052, 0.101	0.010, 0.055, 0.103	0.009, 0.054, 0.105
0.00	ave-LM	0.025, 0.077, 0.131	0.017, 0.065, 0.120	0.024, 0.072, 0.113	0.025, 0.078, 0.129
0.00	exp-LM	0.006, 0.056, 0.115	0.011, 0.054, 0.095	0.012, 0.054, 0.111	0.010, 0.056, 0.108
0.35	sup-W	0.012, 0.060, 0.125	0.015, 0.065, 0.117	0.010, 0.059, 0.124	0.011, 0.070, 0.129
0.35	ave-W	0.024, 0.085, 0.144	0.025, 0.084, 0.136	0.023, 0.090, 0.143	0.031, 0.086, 0.140
0.35	exp-W	0.014, 0.051, 0.115	0.013, 0.061, 0.110	0.010, 0.065, 0.112	0.009, 0.073, 0.134
0.35	sup-LM	0.007, 0.031, 0.073	0.009, 0.044, 0.092	0.011, 0.048, 0.105	0.008, 0.062, 0.121
0.35	ave-LM	0.015, 0.056, 0.113	0.021, 0.072, 0.123	0.022, 0.078, 0.140	0.028, 0.087, 0.136
0.35	exp-LM	0.003, 0.027, 0.075	0.009, 0.048, 0.092	0.007, 0.053, 0.106	0.007, 0.068, 0.129
0.70	sup-W	0.018, 0.073, 0.151	0.013, 0.068, 0.129	0.008, 0.045, 0.100	0.013, 0.056, 0.101
0.70	ave-W	0.033, 0.130, 0.194	0.033, 0.115, 0.173	0.024, 0.072, 0.126	0.030, 0.088, 0.134
0.70	exp-W	0.017, 0.070, 0.154	0.009, 0.068, 0.134	0.008, 0.049, 0.101	0.008, 0.055, 0.106
0.70	sup-LM	0.001, 0.027, 0.064	0.005, 0.037, 0.087	0.004, 0.036, 0.078	0.009, 0.046, 0.088
0.70	ave-LM	0.015, 0.080, 0.144	0.023, 0.087, 0.148	0.018, 0.065, 0.123	0.029, 0.085, 0.130
0.70	exp-LM	0.001, 0.030, 0.061	0.003, 0.042, 0.103	0.006, 0.044, 0.083	0.009, 0.051, 0.095

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor\{1-0.5(1-\kappa)\}n\rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 9: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.6$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.022, 0.065, 0.121	0.010, 0.066, 0.115	0.013, 0.050, 0.115	0.009, 0.052, 0.105
0.00	ave-W	0.030, 0.094, 0.150	0.030, 0.091, 0.148	0.023, 0.094, 0.138	0.025, 0.085, 0.142
0.00	exp-W	0.022, 0.071, 0.140	0.010, 0.066, 0.127	0.013, 0.053, 0.109	0.009, 0.057, 0.108
0.00	sup-LM	0.012, 0.048, 0.088	0.007, 0.051, 0.104	0.012, 0.044, 0.103	0.009, 0.051, 0.098
0.00	ave-LM	0.016, 0.074, 0.120	0.025, 0.084, 0.137	0.021, 0.086, 0.137	0.025, 0.079, 0.138
0.00	exp-LM	0.011, 0.049, 0.089	0.007, 0.059, 0.109	0.013, 0.051, 0.105	0.007, 0.052, 0.104
0.35	sup-W	0.019, 0.083, 0.161	0.012, 0.044, 0.117	0.010, 0.065, 0.117	0.012, 0.049, 0.099
0.35	ave-W	0.030, 0.095, 0.153	0.029, 0.081, 0.143	0.029, 0.095, 0.142	0.027, 0.090, 0.141
0.35	exp-W	0.015, 0.074, 0.141	0.008, 0.046, 0.112	0.008, 0.068, 0.120	0.010, 0.054, 0.099
0.35	sup-LM	0.004, 0.039, 0.082	0.006, 0.031, 0.080	0.005, 0.056, 0.109	0.009, 0.042, 0.092
0.35	ave-LM	0.019, 0.068, 0.116	0.022, 0.067, 0.124	0.022, 0.090, 0.138	0.027, 0.086, 0.136
0.35	exp-LM	0.006, 0.035, 0.078	0.005, 0.027, 0.083	0.006, 0.053, 0.107	0.008, 0.049, 0.092
0.70	sup-W	0.008, 0.075, 0.146	0.012, 0.062, 0.106	0.009, 0.055, 0.113	0.013, 0.045, 0.091
0.70	ave-W	0.041, 0.124, 0.211	0.034, 0.096, 0.158	0.038, 0.102, 0.155	0.043, 0.099, 0.151
0.70	exp-W	0.009, 0.072, 0.138	0.008, 0.060, 0.109	0.010, 0.057, 0.123	0.013, 0.046, 0.096
0.70	sup-LM	0.001, 0.015, 0.048	0.002, 0.036, 0.071	0.005, 0.039, 0.095	0.011, 0.038, 0.082
0.70	ave-LM	0.014, 0.063, 0.130	0.017, 0.072, 0.133	0.032, 0.092, 0.139	0.038, 0.091, 0.148
0.70	exp-LM	0.001, 0.018, 0.049	0.003, 0.038, 0.074	0.009, 0.045, 0.095	0.012, 0.045, 0.084

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 10: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.9$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.017, 0.068, 0.137	0.015, 0.069, 0.125	0.007, 0.035, 0.093	0.011, 0.051, 0.111
0.00	ave-W	0.017, 0.068, 0.137	0.015, 0.069, 0.125	0.007, 0.035, 0.093	0.011, 0.051, 0.111
0.00	exp-W	0.017, 0.068, 0.137	0.015, 0.069, 0.125	0.007, 0.035, 0.093	0.011, 0.051, 0.111
0.00	sup-LM	0.013, 0.052, 0.118	0.010, 0.062, 0.121	0.005, 0.035, 0.091	0.011, 0.055, 0.110
0.00	ave-LM	0.013, 0.052, 0.118	0.010, 0.062, 0.121	0.005, 0.035, 0.091	0.011, 0.055, 0.110
0.00	exp-LM	0.013, 0.052, 0.118	0.010, 0.062, 0.121	0.005, 0.035, 0.091	0.011, 0.055, 0.110
0.35	sup-W	0.016, 0.088, 0.139	0.019, 0.059, 0.112	0.017, 0.061, 0.114	0.011, 0.049, 0.096
0.35	ave-W	0.019, 0.069, 0.131	0.012, 0.052, 0.093	0.016, 0.057, 0.104	0.006, 0.050, 0.098
0.35	exp-W	0.019, 0.064, 0.139	0.014, 0.055, 0.095	0.015, 0.063, 0.102	0.010, 0.049, 0.094
0.35	sup-LM	0.013, 0.057, 0.113	0.010, 0.050, 0.102	0.015, 0.057, 0.107	0.010, 0.049, 0.093
0.35	ave-LM	0.012, 0.054, 0.105	0.007, 0.048, 0.085	0.012, 0.054, 0.097	0.006, 0.051, 0.093
0.35	exp-LM	0.016, 0.052, 0.109	0.009, 0.049, 0.086	0.011, 0.055, 0.098	0.006, 0.045, 0.093
0.70	sup-W	0.022, 0.098, 0.171	0.006, 0.057, 0.123	0.010, 0.052, 0.103	0.022, 0.069, 0.128
0.70	ave-W	0.035, 0.088, 0.157	0.008, 0.059, 0.120	0.006, 0.054, 0.102	0.017, 0.074, 0.121
0.70	exp-W	0.023, 0.097, 0.152	0.004, 0.052, 0.108	0.007, 0.052, 0.103	0.022, 0.071, 0.120
0.70	sup-LM	0.008, 0.053, 0.108	0.001, 0.040, 0.088	0.007, 0.050, 0.092	0.020, 0.065, 0.125
0.70	ave-LM	0.015, 0.064, 0.121	0.004, 0.046, 0.104	0.005, 0.048, 0.095	0.016, 0.072, 0.118
0.70	exp-LM	0.013, 0.049, 0.112	0.001, 0.035, 0.090	0.007, 0.044, 0.091	0.017, 0.066, 0.115

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 11: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.9$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.013, 0.051, 0.108	0.009, 0.046, 0.098	0.011, 0.052, 0.103	0.005, 0.032, 0.068
0.00	ave-W	0.029, 0.076, 0.130	0.024, 0.082, 0.145	0.032, 0.079, 0.125	0.024, 0.059, 0.100
0.00	exp-W	0.015, 0.057, 0.110	0.010, 0.057, 0.105	0.013, 0.057, 0.103	0.006, 0.038, 0.074
0.00	sup-LM	0.010, 0.045, 0.089	0.008, 0.037, 0.089	0.009, 0.048, 0.095	0.006, 0.033, 0.064
0.00	ave-LM	0.022, 0.061, 0.114	0.021, 0.078, 0.135	0.029, 0.078, 0.124	0.024, 0.060, 0.099
0.00	exp-LM	0.010, 0.050, 0.093	0.008, 0.047, 0.100	0.011, 0.054, 0.102	0.007, 0.035, 0.067
0.35	sup-W	0.011, 0.057, 0.115	0.018, 0.059, 0.119	0.014, 0.053, 0.101	0.007, 0.048, 0.086
0.35	ave-W	0.022, 0.086, 0.152	0.035, 0.099, 0.143	0.033, 0.093, 0.134	0.033, 0.091, 0.132
0.35	exp-W	0.009, 0.057, 0.115	0.020, 0.064, 0.121	0.013, 0.051, 0.099	0.011, 0.057, 0.106
0.35	sup-LM	0.005, 0.032, 0.083	0.014, 0.054, 0.099	0.010, 0.043, 0.090	0.008, 0.047, 0.084
0.35	ave-LM	0.016, 0.070, 0.124	0.028, 0.091, 0.136	0.027, 0.082, 0.131	0.030, 0.089, 0.132
0.35	exp-LM	0.006, 0.036, 0.078	0.014, 0.056, 0.110	0.010, 0.046, 0.094	0.011, 0.056, 0.102
0.70	sup-W	0.020, 0.077, 0.129	0.012, 0.045, 0.102	0.014, 0.051, 0.090	0.003, 0.047, 0.093
0.70	ave-W	0.047, 0.098, 0.148	0.034, 0.088, 0.144	0.037, 0.089, 0.135	0.031, 0.094, 0.140
0.70	exp-W	0.021, 0.073, 0.132	0.012, 0.054, 0.103	0.014, 0.058, 0.096	0.009, 0.050, 0.097
0.70	sup-LM	0.005, 0.033, 0.079	0.007, 0.034, 0.068	0.011, 0.040, 0.080	0.004, 0.038, 0.087
0.70	ave-LM	0.027, 0.073, 0.120	0.027, 0.074, 0.125	0.030, 0.084, 0.126	0.030, 0.089, 0.138
0.70	exp-LM	0.006, 0.043, 0.081	0.009, 0.039, 0.087	0.010, 0.049, 0.090	0.009, 0.043, 0.091

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 12: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.9$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.009, 0.057, 0.103	0.011, 0.050, 0.088	0.009, 0.035, 0.080	0.008, 0.027, 0.065
0.00	ave-W	0.054, 0.120, 0.182	0.043, 0.103, 0.148	0.044, 0.114, 0.162	0.034, 0.087, 0.147
0.00	exp-W	0.011, 0.066, 0.125	0.013, 0.058, 0.101	0.012, 0.044, 0.104	0.009, 0.040, 0.079
0.00	sup-LM	0.006, 0.040, 0.084	0.009, 0.046, 0.079	0.008, 0.033, 0.074	0.006, 0.022, 0.063
0.00	ave-LM	0.042, 0.100, 0.155	0.040, 0.094, 0.142	0.041, 0.110, 0.157	0.035, 0.085, 0.145
0.00	exp-LM	0.007, 0.047, 0.105	0.009, 0.052, 0.089	0.010, 0.044, 0.091	0.007, 0.036, 0.077
0.35	sup-W	0.018, 0.066, 0.118	0.017, 0.056, 0.093	0.011, 0.039, 0.094	0.008, 0.033, 0.072
0.35	ave-W	0.077, 0.150, 0.214	0.054, 0.120, 0.160	0.051, 0.118, 0.158	0.050, 0.107, 0.171
0.35	exp-W	0.021, 0.074, 0.145	0.021, 0.061, 0.100	0.010, 0.049, 0.101	0.009, 0.044, 0.084
0.35	sup-LM	0.006, 0.042, 0.078	0.008, 0.041, 0.072	0.009, 0.033, 0.080	0.008, 0.032, 0.066
0.35	ave-LM	0.051, 0.134, 0.189	0.048, 0.109, 0.152	0.049, 0.107, 0.156	0.049, 0.105, 0.166
0.35	exp-LM	0.010, 0.049, 0.104	0.012, 0.045, 0.089	0.010, 0.044, 0.091	0.010, 0.039, 0.081
0.70	sup-W	0.014, 0.057, 0.111	0.012, 0.048, 0.086	0.008, 0.029, 0.078	0.008, 0.035, 0.071
0.70	ave-W	0.067, 0.130, 0.174	0.071, 0.146, 0.185	0.060, 0.121, 0.165	0.043, 0.107, 0.156
0.70	exp-W	0.018, 0.064, 0.122	0.015, 0.056, 0.117	0.011, 0.044, 0.082	0.012, 0.036, 0.072
0.70	sup-LM	0.002, 0.022, 0.049	0.006, 0.031, 0.064	0.008, 0.022, 0.054	0.006, 0.034, 0.064
0.70	ave-LM	0.038, 0.110, 0.141	0.060, 0.127, 0.170	0.058, 0.112, 0.154	0.041, 0.102, 0.149
0.70	exp-LM	0.004, 0.026, 0.063	0.007, 0.045, 0.084	0.009, 0.035, 0.076	0.008, 0.031, 0.068

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 13: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.3$, $\psi_0 = 0.9$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.014, 0.055, 0.109	0.011, 0.046, 0.088	0.009, 0.042, 0.084	0.006, 0.042, 0.081
0.00	ave-W	0.086, 0.153, 0.205	0.066, 0.142, 0.188	0.073, 0.135, 0.188	0.074, 0.132, 0.184
0.00	exp-W	0.017, 0.072, 0.137	0.011, 0.059, 0.101	0.010, 0.062, 0.113	0.010, 0.058, 0.102
0.00	sup-LM	0.007, 0.036, 0.078	0.008, 0.031, 0.078	0.007, 0.038, 0.081	0.007, 0.039, 0.081
0.00	ave-LM	0.071, 0.138, 0.184	0.059, 0.130, 0.180	0.072, 0.125, 0.184	0.072, 0.127, 0.181
0.00	exp-LM	0.007, 0.048, 0.104	0.009, 0.053, 0.094	0.009, 0.059, 0.106	0.013, 0.056, 0.097
0.35	sup-W	0.014, 0.064, 0.112	0.012, 0.048, 0.089	0.007, 0.036, 0.074	0.006, 0.031, 0.069
0.35	ave-W	0.084, 0.155, 0.205	0.090, 0.144, 0.187	0.079, 0.135, 0.188	0.077, 0.150, 0.190
0.35	exp-W	0.015, 0.074, 0.133	0.013, 0.062, 0.102	0.008, 0.047, 0.091	0.010, 0.047, 0.086
0.35	sup-LM	0.005, 0.032, 0.063	0.007, 0.035, 0.065	0.004, 0.028, 0.061	0.006, 0.029, 0.065
0.35	ave-LM	0.064, 0.135, 0.179	0.082, 0.130, 0.173	0.076, 0.130, 0.182	0.072, 0.147, 0.187
0.35	exp-LM	0.006, 0.036, 0.090	0.009, 0.049, 0.082	0.004, 0.041, 0.085	0.008, 0.044, 0.084
0.70	sup-W	0.014, 0.050, 0.120	0.005, 0.039, 0.094	0.004, 0.037, 0.072	0.006, 0.034, 0.069
0.70	ave-W	0.102, 0.183, 0.235	0.092, 0.163, 0.208	0.087, 0.152, 0.194	0.097, 0.162, 0.213
0.70	exp-W	0.017, 0.063, 0.122	0.008, 0.057, 0.109	0.007, 0.043, 0.089	0.008, 0.047, 0.079
0.70	sup-LM	0.000, 0.014, 0.037	0.002, 0.017, 0.055	0.002, 0.022, 0.055	0.007, 0.029, 0.058
0.70	ave-LM	0.064, 0.132, 0.183	0.077, 0.141, 0.190	0.078, 0.141, 0.188	0.091, 0.159, 0.207
0.70	exp-LM	0.004, 0.025, 0.050	0.004, 0.034, 0.081	0.005, 0.037, 0.078	0.008, 0.042, 0.077

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.3$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \dots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 14: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.3$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.013, 0.070, 0.129	0.016, 0.060, 0.113	0.013, 0.060, 0.107	0.008, 0.053, 0.091
0.00	ave-W	0.013, 0.070, 0.129	0.016, 0.060, 0.113	0.013, 0.060, 0.107	0.008, 0.053, 0.091
0.00	exp-W	0.013, 0.070, 0.129	0.016, 0.060, 0.113	0.013, 0.060, 0.107	0.008, 0.053, 0.091
0.00	sup-LM	0.005, 0.060, 0.114	0.012, 0.055, 0.105	0.013, 0.057, 0.098	0.009, 0.050, 0.092
0.00	ave-LM	0.005, 0.060, 0.114	0.012, 0.055, 0.105	0.013, 0.057, 0.098	0.009, 0.050, 0.092
0.00	exp-LM	0.005, 0.060, 0.114	0.012, 0.055, 0.105	0.013, 0.057, 0.098	0.009, 0.050, 0.092
0.35	sup-W	0.018, 0.084, 0.149	0.015, 0.057, 0.116	0.012, 0.052, 0.109	0.012, 0.063, 0.129
0.35	ave-W	0.014, 0.062, 0.140	0.012, 0.061, 0.118	0.011, 0.052, 0.102	0.013, 0.066, 0.115
0.35	exp-W	0.014, 0.069, 0.141	0.014, 0.058, 0.115	0.008, 0.050, 0.108	0.011, 0.065, 0.119
0.35	sup-LM	0.006, 0.059, 0.125	0.011, 0.050, 0.108	0.007, 0.052, 0.098	0.011, 0.062, 0.123
0.35	ave-LM	0.009, 0.047, 0.114	0.008, 0.055, 0.106	0.008, 0.046, 0.103	0.013, 0.064, 0.111
0.35	exp-LM	0.006, 0.048, 0.111	0.011, 0.050, 0.104	0.009, 0.047, 0.102	0.011, 0.062, 0.118
0.70	sup-W	0.011, 0.071, 0.129	0.013, 0.075, 0.137	0.010, 0.058, 0.116	0.008, 0.057, 0.118
0.70	ave-W	0.016, 0.057, 0.113	0.015, 0.085, 0.141	0.010, 0.055, 0.103	0.011, 0.058, 0.109
0.70	exp-W	0.011, 0.062, 0.116	0.012, 0.073, 0.144	0.010, 0.059, 0.108	0.005, 0.055, 0.118
0.70	sup-LM	0.003, 0.034, 0.075	0.009, 0.051, 0.114	0.006, 0.050, 0.102	0.004, 0.048, 0.110
0.70	ave-LM	0.007, 0.041, 0.083	0.009, 0.071, 0.126	0.006, 0.041, 0.096	0.010, 0.055, 0.103
0.70	exp-LM	0.004, 0.033, 0.071	0.007, 0.048, 0.121	0.007, 0.049, 0.094	0.005, 0.053, 0.113

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 15: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.3$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.023, 0.078, 0.132	0.008, 0.056, 0.101	0.006, 0.051, 0.094	0.011, 0.058, 0.103
0.00	ave-W	0.024, 0.077, 0.134	0.006, 0.055, 0.103	0.012, 0.054, 0.117	0.014, 0.057, 0.111
0.00	exp-W	0.023, 0.081, 0.135	0.007, 0.055, 0.103	0.006, 0.052, 0.102	0.011, 0.061, 0.105
0.00	sup-LM	0.014, 0.065, 0.117	0.006, 0.047, 0.086	0.005, 0.048, 0.090	0.008, 0.056, 0.103
0.00	ave-LM	0.017, 0.066, 0.113	0.004, 0.048, 0.092	0.009, 0.052, 0.113	0.013, 0.053, 0.109
0.00	exp-LM	0.014, 0.065, 0.117	0.006, 0.050, 0.095	0.005, 0.048, 0.099	0.009, 0.060, 0.100
0.35	sup-W	0.020, 0.080, 0.145	0.014, 0.059, 0.126	0.012, 0.064, 0.106	0.010, 0.058, 0.111
0.35	ave-W	0.019, 0.072, 0.141	0.014, 0.061, 0.114	0.009, 0.049, 0.100	0.011, 0.057, 0.096
0.35	exp-W	0.017, 0.067, 0.138	0.011, 0.056, 0.114	0.013, 0.057, 0.102	0.007, 0.053, 0.104
0.35	sup-LM	0.005, 0.048, 0.097	0.009, 0.045, 0.110	0.010, 0.056, 0.097	0.007, 0.057, 0.106
0.35	ave-LM	0.010, 0.053, 0.107	0.011, 0.049, 0.102	0.008, 0.048, 0.094	0.011, 0.052, 0.092
0.35	exp-LM	0.007, 0.046, 0.105	0.007, 0.047, 0.092	0.009, 0.051, 0.095	0.006, 0.052, 0.097
0.70	sup-W	0.014, 0.071, 0.135	0.013, 0.065, 0.118	0.011, 0.059, 0.112	0.011, 0.062, 0.118
0.70	ave-W	0.022, 0.074, 0.136	0.014, 0.063, 0.119	0.016, 0.055, 0.103	0.019, 0.070, 0.107
0.70	exp-W	0.016, 0.068, 0.133	0.013, 0.063, 0.110	0.008, 0.054, 0.107	0.012, 0.067, 0.108
0.70	sup-LM	0.003, 0.029, 0.073	0.008, 0.044, 0.096	0.007, 0.044, 0.097	0.010, 0.058, 0.110
0.70	ave-LM	0.008, 0.046, 0.096	0.011, 0.046, 0.095	0.011, 0.050, 0.095	0.020, 0.068, 0.103
0.70	exp-LM	0.006, 0.032, 0.076	0.007, 0.040, 0.089	0.006, 0.044, 0.092	0.012, 0.055, 0.104

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 16: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.3$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.015, 0.074, 0.134	0.006, 0.061, 0.108	0.008, 0.061, 0.106	0.006, 0.048, 0.097
0.00	ave-W	0.023, 0.066, 0.128	0.009, 0.054, 0.099	0.010, 0.049, 0.097	0.013, 0.051, 0.102
0.00	exp-W	0.015, 0.075, 0.133	0.006, 0.056, 0.106	0.007, 0.058, 0.108	0.007, 0.051, 0.097
0.00	sup-LM	0.006, 0.052, 0.110	0.004, 0.046, 0.096	0.006, 0.052, 0.102	0.005, 0.047, 0.095
0.00	ave-LM	0.011, 0.050, 0.090	0.006, 0.046, 0.091	0.009, 0.044, 0.093	0.009, 0.050, 0.100
0.00	exp-LM	0.006, 0.051, 0.111	0.004, 0.044, 0.095	0.006, 0.052, 0.099	0.006, 0.050, 0.095
0.35	sup-W	0.016, 0.065, 0.152	0.007, 0.048, 0.114	0.014, 0.057, 0.108	0.011, 0.056, 0.124
0.35	ave-W	0.018, 0.065, 0.126	0.009, 0.059, 0.124	0.016, 0.068, 0.120	0.018, 0.070, 0.116
0.35	exp-W	0.015, 0.061, 0.127	0.006, 0.046, 0.106	0.007, 0.059, 0.102	0.010, 0.058, 0.118
0.35	sup-LM	0.005, 0.032, 0.083	0.005, 0.034, 0.084	0.006, 0.055, 0.100	0.009, 0.054, 0.118
0.35	ave-LM	0.011, 0.040, 0.099	0.005, 0.053, 0.107	0.013, 0.064, 0.112	0.015, 0.069, 0.115
0.35	exp-LM	0.005, 0.035, 0.077	0.005, 0.031, 0.082	0.006, 0.056, 0.094	0.008, 0.055, 0.110
0.70	sup-W	0.014, 0.068, 0.124	0.016, 0.064, 0.134	0.016, 0.058, 0.113	0.009, 0.058, 0.135
0.70	ave-W	0.019, 0.073, 0.137	0.020, 0.081, 0.145	0.012, 0.072, 0.134	0.019, 0.070, 0.125
0.70	exp-W	0.010, 0.068, 0.109	0.014, 0.064, 0.120	0.016, 0.059, 0.111	0.011, 0.058, 0.127
0.70	sup-LM	0.001, 0.028, 0.062	0.004, 0.040, 0.084	0.012, 0.045, 0.088	0.007, 0.053, 0.120
0.70	ave-LM	0.006, 0.043, 0.096	0.014, 0.049, 0.122	0.009, 0.058, 0.123	0.017, 0.062, 0.121
0.70	exp-LM	0.001, 0.023, 0.062	0.008, 0.044, 0.089	0.014, 0.049, 0.093	0.008, 0.056, 0.119

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 17: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.3$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.010, 0.073, 0.138	0.012, 0.067, 0.123	0.010, 0.052, 0.097	0.016, 0.055, 0.106
0.00	ave-W	0.015, 0.056, 0.128	0.016, 0.074, 0.138	0.013, 0.059, 0.115	0.015, 0.056, 0.103
0.00	exp-W	0.010, 0.071, 0.134	0.012, 0.068, 0.130	0.010, 0.055, 0.100	0.017, 0.052, 0.100
0.00	sup-LM	0.003, 0.051, 0.096	0.006, 0.049, 0.101	0.010, 0.048, 0.086	0.016, 0.054, 0.101
0.00	ave-LM	0.010, 0.045, 0.090	0.011, 0.062, 0.126	0.011, 0.051, 0.110	0.014, 0.053, 0.097
0.00	exp-LM	0.003, 0.048, 0.097	0.007, 0.049, 0.111	0.010, 0.045, 0.089	0.015, 0.052, 0.098
0.35	sup-W	0.017, 0.080, 0.158	0.013, 0.057, 0.117	0.014, 0.069, 0.120	0.012, 0.059, 0.107
0.35	ave-W	0.016, 0.082, 0.142	0.011, 0.064, 0.121	0.018, 0.063, 0.110	0.016, 0.067, 0.111
0.35	exp-W	0.013, 0.069, 0.144	0.011, 0.052, 0.115	0.014, 0.067, 0.106	0.012, 0.053, 0.104
0.35	sup-LM	0.007, 0.037, 0.085	0.007, 0.042, 0.092	0.010, 0.055, 0.111	0.010, 0.053, 0.098
0.35	ave-LM	0.005, 0.049, 0.109	0.007, 0.052, 0.101	0.016, 0.051, 0.097	0.014, 0.066, 0.108
0.35	exp-LM	0.006, 0.036, 0.080	0.007, 0.035, 0.086	0.007, 0.058, 0.092	0.009, 0.049, 0.094
0.70	sup-W	0.014, 0.082, 0.162	0.010, 0.061, 0.118	0.007, 0.045, 0.098	0.016, 0.064, 0.109
0.70	ave-W	0.019, 0.080, 0.163	0.012, 0.061, 0.128	0.008, 0.067, 0.113	0.020, 0.057, 0.111
0.70	exp-W	0.012, 0.080, 0.152	0.010, 0.055, 0.111	0.006, 0.047, 0.100	0.014, 0.059, 0.105
0.70	sup-LM	0.001, 0.019, 0.059	0.003, 0.031, 0.075	0.006, 0.032, 0.076	0.013, 0.057, 0.100
0.70	ave-LM	0.006, 0.037, 0.088	0.008, 0.050, 0.095	0.007, 0.059, 0.105	0.019, 0.055, 0.102
0.70	exp-LM	0.001, 0.019, 0.062	0.001, 0.032, 0.073	0.005, 0.034, 0.086	0.011, 0.052, 0.096

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 18: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.6$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.013, 0.083, 0.143	0.015, 0.061, 0.112	0.016, 0.057, 0.099	0.007, 0.047, 0.105
0.00	ave-W	0.013, 0.083, 0.143	0.015, 0.061, 0.112	0.016, 0.057, 0.099	0.007, 0.047, 0.105
0.00	exp-W	0.013, 0.083, 0.143	0.015, 0.061, 0.112	0.016, 0.057, 0.099	0.007, 0.047, 0.105
0.00	sup-LM	0.007, 0.060, 0.130	0.011, 0.050, 0.109	0.012, 0.054, 0.092	0.006, 0.050, 0.103
0.00	ave-LM	0.007, 0.060, 0.130	0.011, 0.050, 0.109	0.012, 0.054, 0.092	0.006, 0.050, 0.103
0.00	exp-LM	0.007, 0.060, 0.130	0.011, 0.050, 0.109	0.012, 0.054, 0.092	0.006, 0.050, 0.103
0.35	sup-W	0.024, 0.085, 0.167	0.013, 0.058, 0.111	0.015, 0.056, 0.108	0.012, 0.055, 0.111
0.35	ave-W	0.021, 0.075, 0.143	0.009, 0.060, 0.099	0.014, 0.055, 0.103	0.011, 0.049, 0.096
0.35	exp-W	0.025, 0.075, 0.157	0.011, 0.057, 0.102	0.016, 0.055, 0.101	0.011, 0.048, 0.106
0.35	sup-LM	0.012, 0.060, 0.128	0.010, 0.047, 0.099	0.014, 0.050, 0.104	0.011, 0.052, 0.110
0.35	ave-LM	0.013, 0.057, 0.129	0.006, 0.057, 0.092	0.012, 0.053, 0.099	0.008, 0.045, 0.097
0.35	exp-LM	0.013, 0.053, 0.123	0.010, 0.051, 0.093	0.014, 0.052, 0.098	0.011, 0.043, 0.101
0.70	sup-W	0.014, 0.076, 0.146	0.008, 0.058, 0.115	0.015, 0.065, 0.113	0.010, 0.049, 0.113
0.70	ave-W	0.018, 0.075, 0.141	0.004, 0.047, 0.117	0.014, 0.050, 0.110	0.012, 0.049, 0.104
0.70	exp-W	0.013, 0.079, 0.148	0.006, 0.051, 0.116	0.013, 0.057, 0.109	0.010, 0.049, 0.104
0.70	sup-LM	0.002, 0.037, 0.097	0.005, 0.041, 0.091	0.009, 0.053, 0.105	0.010, 0.046, 0.103
0.70	ave-LM	0.008, 0.045, 0.116	0.003, 0.038, 0.098	0.011, 0.045, 0.100	0.008, 0.049, 0.101
0.70	exp-LM	0.004, 0.036, 0.097	0.003, 0.037, 0.100	0.011, 0.050, 0.095	0.010, 0.046, 0.102

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 19: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.6$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.015, 0.064, 0.113	0.017, 0.058, 0.105	0.019, 0.053, 0.100	0.006, 0.046, 0.094
0.00	ave-W	0.021, 0.072, 0.127	0.021, 0.070, 0.125	0.019, 0.072, 0.116	0.014, 0.053, 0.108
0.00	exp-W	0.016, 0.064, 0.118	0.017, 0.061, 0.108	0.018, 0.059, 0.105	0.007, 0.047, 0.100
0.00	sup-LM	0.011, 0.047, 0.094	0.017, 0.051, 0.095	0.015, 0.052, 0.094	0.007, 0.044, 0.092
0.00	ave-LM	0.015, 0.058, 0.113	0.017, 0.064, 0.115	0.017, 0.072, 0.111	0.015, 0.052, 0.106
0.00	exp-LM	0.011, 0.049, 0.097	0.017, 0.054, 0.099	0.015, 0.057, 0.101	0.008, 0.045, 0.097
0.35	sup-W	0.020, 0.073, 0.128	0.026, 0.066, 0.122	0.011, 0.074, 0.126	0.013, 0.063, 0.119
0.35	ave-W	0.022, 0.082, 0.139	0.032, 0.077, 0.119	0.018, 0.075, 0.129	0.017, 0.070, 0.121
0.35	exp-W	0.016, 0.073, 0.130	0.027, 0.066, 0.111	0.011, 0.059, 0.118	0.015, 0.057, 0.118
0.35	sup-LM	0.005, 0.050, 0.103	0.017, 0.058, 0.103	0.010, 0.064, 0.120	0.013, 0.054, 0.111
0.35	ave-LM	0.015, 0.062, 0.116	0.027, 0.073, 0.111	0.018, 0.068, 0.118	0.018, 0.071, 0.113
0.35	exp-LM	0.005, 0.046, 0.100	0.019, 0.061, 0.102	0.007, 0.053, 0.108	0.014, 0.056, 0.113
0.70	sup-W	0.012, 0.074, 0.156	0.008, 0.064, 0.128	0.010, 0.049, 0.104	0.008, 0.050, 0.098
0.70	ave-W	0.023, 0.098, 0.179	0.016, 0.084, 0.127	0.018, 0.063, 0.105	0.021, 0.069, 0.122
0.70	exp-W	0.015, 0.078, 0.150	0.009, 0.059, 0.133	0.010, 0.048, 0.090	0.011, 0.048, 0.093
0.70	sup-LM	0.003, 0.024, 0.076	0.005, 0.042, 0.100	0.007, 0.038, 0.086	0.009, 0.044, 0.087
0.70	ave-LM	0.015, 0.061, 0.133	0.011, 0.065, 0.116	0.018, 0.053, 0.101	0.021, 0.066, 0.119
0.70	exp-LM	0.002, 0.030, 0.093	0.006, 0.042, 0.099	0.010, 0.042, 0.073	0.010, 0.048, 0.087

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 20: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.6$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.007, 0.050, 0.108	0.015, 0.056, 0.102	0.013, 0.054, 0.111	0.007, 0.044, 0.082
0.00	ave-W	0.016, 0.070, 0.142	0.033, 0.079, 0.132	0.017, 0.083, 0.137	0.018, 0.060, 0.107
0.00	exp-W	0.008, 0.052, 0.108	0.015, 0.060, 0.102	0.013, 0.060, 0.118	0.007, 0.044, 0.086
0.00	sup-LM	0.005, 0.032, 0.078	0.011, 0.049, 0.093	0.011, 0.051, 0.106	0.008, 0.044, 0.078
0.00	ave-LM	0.012, 0.052, 0.108	0.026, 0.072, 0.122	0.018, 0.083, 0.138	0.017, 0.058, 0.103
0.00	exp-LM	0.005, 0.039, 0.080	0.013, 0.054, 0.095	0.011, 0.056, 0.111	0.009, 0.042, 0.084
0.35	sup-W	0.018, 0.072, 0.136	0.010, 0.057, 0.105	0.008, 0.047, 0.098	0.012, 0.059, 0.108
0.35	ave-W	0.031, 0.087, 0.145	0.022, 0.073, 0.128	0.023, 0.086, 0.139	0.021, 0.077, 0.124
0.35	exp-W	0.017, 0.069, 0.133	0.013, 0.056, 0.104	0.006, 0.051, 0.113	0.012, 0.054, 0.108
0.35	sup-LM	0.007, 0.038, 0.085	0.007, 0.034, 0.088	0.005, 0.042, 0.086	0.012, 0.051, 0.103
0.35	ave-LM	0.019, 0.062, 0.115	0.016, 0.069, 0.113	0.019, 0.079, 0.132	0.020, 0.076, 0.120
0.35	exp-LM	0.010, 0.038, 0.089	0.008, 0.038, 0.088	0.005, 0.042, 0.105	0.012, 0.050, 0.103
0.70	sup-W	0.019, 0.071, 0.138	0.006, 0.069, 0.114	0.011, 0.063, 0.120	0.012, 0.069, 0.127
0.70	ave-W	0.031, 0.104, 0.174	0.027, 0.087, 0.145	0.034, 0.083, 0.134	0.035, 0.097, 0.158
0.70	exp-W	0.017, 0.074, 0.136	0.010, 0.062, 0.114	0.010, 0.061, 0.119	0.011, 0.069, 0.130
0.70	sup-LM	0.002, 0.027, 0.068	0.004, 0.038, 0.085	0.007, 0.044, 0.091	0.009, 0.058, 0.117
0.70	ave-LM	0.018, 0.067, 0.125	0.017, 0.068, 0.127	0.025, 0.074, 0.121	0.031, 0.092, 0.150
0.70	exp-LM	0.003, 0.033, 0.081	0.005, 0.043, 0.087	0.007, 0.044, 0.097	0.010, 0.064, 0.124

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 21: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.6$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.015, 0.054, 0.117	0.010, 0.049, 0.120	0.008, 0.053, 0.113	0.010, 0.045, 0.093
0.00	ave-W	0.021, 0.093, 0.162	0.028, 0.093, 0.146	0.024, 0.075, 0.133	0.028, 0.079, 0.134
0.00	exp-W	0.016, 0.056, 0.123	0.010, 0.054, 0.125	0.007, 0.055, 0.119	0.010, 0.046, 0.095
0.00	sup-LM	0.008, 0.034, 0.076	0.006, 0.037, 0.090	0.006, 0.048, 0.100	0.009, 0.045, 0.092
0.00	ave-LM	0.011, 0.069, 0.128	0.024, 0.077, 0.133	0.024, 0.065, 0.127	0.029, 0.077, 0.126
0.00	exp-LM	0.008, 0.031, 0.081	0.008, 0.041, 0.101	0.006, 0.053, 0.114	0.010, 0.044, 0.092
0.35	sup-W	0.017, 0.080, 0.143	0.017, 0.075, 0.129	0.009, 0.052, 0.095	0.006, 0.059, 0.106
0.35	ave-W	0.033, 0.099, 0.164	0.031, 0.094, 0.152	0.028, 0.081, 0.132	0.028, 0.096, 0.153
0.35	exp-W	0.015, 0.072, 0.140	0.015, 0.065, 0.123	0.009, 0.046, 0.101	0.006, 0.050, 0.115
0.35	sup-LM	0.007, 0.039, 0.084	0.008, 0.049, 0.098	0.008, 0.043, 0.082	0.005, 0.053, 0.099
0.35	ave-LM	0.020, 0.071, 0.129	0.024, 0.082, 0.137	0.025, 0.072, 0.125	0.025, 0.090, 0.147
0.35	exp-LM	0.007, 0.035, 0.083	0.007, 0.052, 0.095	0.007, 0.038, 0.091	0.005, 0.047, 0.107
0.70	sup-W	0.015, 0.081, 0.146	0.011, 0.054, 0.093	0.011, 0.060, 0.115	0.011, 0.063, 0.117
0.70	ave-W	0.040, 0.125, 0.190	0.038, 0.104, 0.153	0.033, 0.096, 0.153	0.049, 0.117, 0.173
0.70	exp-W	0.015, 0.086, 0.149	0.013, 0.045, 0.101	0.010, 0.059, 0.117	0.012, 0.064, 0.119
0.70	sup-LM	0.000, 0.023, 0.048	0.003, 0.024, 0.065	0.006, 0.045, 0.092	0.011, 0.051, 0.105
0.70	ave-LM	0.022, 0.073, 0.130	0.022, 0.081, 0.124	0.028, 0.083, 0.135	0.048, 0.108, 0.163
0.70	exp-LM	0.001, 0.024, 0.062	0.003, 0.031, 0.065	0.007, 0.047, 0.091	0.011, 0.054, 0.108

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 22: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.9$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.022, 0.081, 0.152	0.016, 0.054, 0.110	0.015, 0.061, 0.107	0.011, 0.048, 0.103
0.00	ave-W	0.022, 0.081, 0.152	0.016, 0.054, 0.110	0.015, 0.061, 0.107	0.011, 0.048, 0.103
0.00	exp-W	0.022, 0.081, 0.152	0.016, 0.054, 0.110	0.015, 0.061, 0.107	0.011, 0.048, 0.103
0.00	sup-LM	0.014, 0.069, 0.128	0.011, 0.046, 0.100	0.012, 0.058, 0.104	0.010, 0.048, 0.101
0.00	ave-LM	0.014, 0.069, 0.128	0.011, 0.046, 0.100	0.012, 0.058, 0.104	0.010, 0.048, 0.101
0.00	exp-LM	0.014, 0.069, 0.128	0.011, 0.046, 0.100	0.012, 0.058, 0.104	0.010, 0.048, 0.101
0.35	sup-W	0.016, 0.081, 0.145	0.010, 0.064, 0.125	0.013, 0.077, 0.116	0.009, 0.046, 0.099
0.35	ave-W	0.012, 0.065, 0.126	0.011, 0.059, 0.125	0.015, 0.061, 0.112	0.012, 0.045, 0.097
0.35	exp-W	0.017, 0.069, 0.128	0.011, 0.067, 0.119	0.010, 0.072, 0.116	0.011, 0.043, 0.091
0.35	sup-LM	0.006, 0.053, 0.118	0.006, 0.055, 0.116	0.010, 0.070, 0.112	0.009, 0.042, 0.094
0.35	ave-LM	0.005, 0.048, 0.105	0.008, 0.051, 0.120	0.012, 0.056, 0.104	0.009, 0.044, 0.093
0.35	exp-LM	0.005, 0.049, 0.108	0.006, 0.053, 0.108	0.011, 0.068, 0.109	0.011, 0.042, 0.089
0.70	sup-W	0.019, 0.097, 0.178	0.019, 0.082, 0.141	0.018, 0.064, 0.110	0.012, 0.061, 0.113
0.70	ave-W	0.016, 0.075, 0.155	0.021, 0.071, 0.122	0.017, 0.059, 0.113	0.011, 0.051, 0.103
0.70	exp-W	0.017, 0.085, 0.153	0.017, 0.076, 0.123	0.014, 0.057, 0.118	0.012, 0.060, 0.107
0.70	sup-LM	0.004, 0.047, 0.108	0.011, 0.058, 0.113	0.011, 0.048, 0.100	0.012, 0.056, 0.107
0.70	ave-LM	0.011, 0.049, 0.107	0.013, 0.062, 0.108	0.015, 0.050, 0.105	0.008, 0.051, 0.097
0.70	exp-LM	0.002, 0.045, 0.107	0.011, 0.056, 0.107	0.013, 0.051, 0.105	0.009, 0.056, 0.104

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 23: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.9$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.016, 0.061, 0.111	0.016, 0.061, 0.100	0.009, 0.032, 0.083	0.006, 0.041, 0.078
0.00	ave-W	0.031, 0.096, 0.155	0.024, 0.088, 0.141	0.018, 0.072, 0.123	0.023, 0.070, 0.119
0.00	exp-W	0.016, 0.068, 0.120	0.017, 0.063, 0.113	0.011, 0.040, 0.097	0.007, 0.048, 0.089
0.00	sup-LM	0.009, 0.053, 0.094	0.012, 0.054, 0.092	0.009, 0.033, 0.082	0.005, 0.041, 0.076
0.00	ave-LM	0.023, 0.073, 0.138	0.020, 0.080, 0.134	0.018, 0.070, 0.123	0.020, 0.066, 0.116
0.00	exp-LM	0.010, 0.056, 0.105	0.012, 0.057, 0.103	0.010, 0.040, 0.094	0.006, 0.047, 0.085
0.35	sup-W	0.024, 0.075, 0.141	0.015, 0.045, 0.109	0.012, 0.056, 0.105	0.013, 0.054, 0.115
0.35	ave-W	0.040, 0.102, 0.147	0.034, 0.086, 0.141	0.025, 0.094, 0.151	0.035, 0.103, 0.156
0.35	exp-W	0.024, 0.076, 0.140	0.015, 0.055, 0.110	0.015, 0.053, 0.121	0.016, 0.060, 0.117
0.35	sup-LM	0.007, 0.046, 0.099	0.015, 0.038, 0.092	0.012, 0.052, 0.101	0.012, 0.049, 0.110
0.35	ave-LM	0.034, 0.082, 0.129	0.027, 0.077, 0.129	0.026, 0.089, 0.152	0.032, 0.098, 0.151
0.35	exp-LM	0.012, 0.055, 0.104	0.012, 0.051, 0.091	0.014, 0.051, 0.108	0.015, 0.058, 0.115
0.70	sup-W	0.025, 0.084, 0.139	0.013, 0.055, 0.113	0.009, 0.038, 0.097	0.014, 0.051, 0.100
0.70	ave-W	0.044, 0.120, 0.188	0.033, 0.094, 0.152	0.028, 0.082, 0.143	0.027, 0.089, 0.142
0.70	exp-W	0.027, 0.088, 0.146	0.012, 0.059, 0.116	0.008, 0.044, 0.102	0.014, 0.054, 0.109
0.70	sup-LM	0.006, 0.036, 0.085	0.002, 0.036, 0.079	0.008, 0.029, 0.076	0.013, 0.040, 0.092
0.70	ave-LM	0.025, 0.088, 0.146	0.025, 0.082, 0.133	0.022, 0.072, 0.139	0.027, 0.086, 0.141
0.70	exp-LM	0.008, 0.044, 0.105	0.002, 0.043, 0.085	0.006, 0.038, 0.090	0.014, 0.049, 0.101

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 24: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.9$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.011, 0.053, 0.105	0.008, 0.058, 0.101	0.010, 0.038, 0.079	0.010, 0.051, 0.092
0.00	ave-W	0.050, 0.111, 0.174	0.055, 0.121, 0.177	0.044, 0.103, 0.158	0.064, 0.124, 0.170
0.00	exp-W	0.016, 0.062, 0.116	0.013, 0.064, 0.115	0.012, 0.048, 0.095	0.012, 0.066, 0.106
0.00	sup-LM	0.006, 0.038, 0.077	0.006, 0.041, 0.085	0.008, 0.033, 0.077	0.010, 0.050, 0.090
0.00	ave-LM	0.039, 0.091, 0.153	0.044, 0.110, 0.162	0.043, 0.099, 0.153	0.063, 0.121, 0.169
0.00	exp-LM	0.006, 0.052, 0.092	0.007, 0.052, 0.106	0.012, 0.045, 0.090	0.012, 0.065, 0.106
0.35	sup-W	0.019, 0.077, 0.131	0.012, 0.060, 0.113	0.012, 0.058, 0.098	0.012, 0.036, 0.069
0.35	ave-W	0.068, 0.150, 0.205	0.064, 0.137, 0.193	0.063, 0.116, 0.159	0.047, 0.105, 0.149
0.35	exp-W	0.018, 0.084, 0.136	0.017, 0.066, 0.119	0.015, 0.058, 0.107	0.014, 0.046, 0.079
0.35	sup-LM	0.004, 0.035, 0.085	0.008, 0.045, 0.092	0.009, 0.049, 0.083	0.011, 0.033, 0.065
0.35	ave-LM	0.052, 0.126, 0.179	0.057, 0.121, 0.181	0.061, 0.115, 0.155	0.049, 0.103, 0.145
0.35	exp-LM	0.004, 0.046, 0.109	0.012, 0.052, 0.104	0.012, 0.056, 0.101	0.010, 0.043, 0.072
0.70	sup-W	0.014, 0.072, 0.146	0.014, 0.060, 0.097	0.008, 0.032, 0.072	0.009, 0.047, 0.098
0.70	ave-W	0.084, 0.152, 0.219	0.070, 0.142, 0.198	0.056, 0.124, 0.169	0.071, 0.136, 0.171
0.70	exp-W	0.016, 0.076, 0.157	0.016, 0.065, 0.112	0.009, 0.048, 0.093	0.014, 0.057, 0.109
0.70	sup-LM	0.002, 0.023, 0.065	0.004, 0.039, 0.070	0.005, 0.023, 0.060	0.008, 0.041, 0.088
0.70	ave-LM	0.061, 0.128, 0.169	0.062, 0.122, 0.179	0.053, 0.115, 0.158	0.064, 0.132, 0.169
0.70	exp-LM	0.006, 0.039, 0.080	0.007, 0.048, 0.091	0.006, 0.037, 0.081	0.009, 0.052, 0.103

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4,\}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 25: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.9$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.016, 0.065, 0.128	0.009, 0.050, 0.108	0.010, 0.037, 0.075	0.005, 0.046, 0.081
0.00	ave-W	0.091, 0.175, 0.234	0.071, 0.142, 0.203	0.058, 0.121, 0.170	0.076, 0.133, 0.184
0.00	exp-W	0.021, 0.084, 0.149	0.010, 0.070, 0.131	0.013, 0.045, 0.098	0.012, 0.053, 0.102
0.00	sup-LM	0.006, 0.044, 0.091	0.007, 0.043, 0.084	0.008, 0.036, 0.068	0.005, 0.043, 0.077
0.00	ave-LM	0.066, 0.150, 0.204	0.057, 0.136, 0.186	0.057, 0.117, 0.166	0.073, 0.133, 0.181
0.00	exp-LM	0.009, 0.053, 0.112	0.007, 0.058, 0.105	0.011, 0.041, 0.088	0.010, 0.048, 0.098
0.35	sup-W	0.019, 0.072, 0.129	0.009, 0.051, 0.102	0.013, 0.054, 0.091	0.005, 0.045, 0.076
0.35	ave-W	0.102, 0.181, 0.231	0.088, 0.161, 0.220	0.093, 0.152, 0.207	0.099, 0.155, 0.192
0.35	exp-W	0.020, 0.083, 0.145	0.013, 0.066, 0.112	0.014, 0.061, 0.106	0.009, 0.055, 0.111
0.35	sup-LM	0.006, 0.040, 0.075	0.008, 0.035, 0.074	0.010, 0.039, 0.080	0.004, 0.038, 0.075
0.35	ave-LM	0.071, 0.152, 0.210	0.082, 0.148, 0.207	0.089, 0.147, 0.201	0.096, 0.150, 0.188
0.35	exp-LM	0.008, 0.047, 0.093	0.008, 0.045, 0.092	0.014, 0.054, 0.100	0.005, 0.050, 0.106
0.70	sup-W	0.013, 0.081, 0.148	0.010, 0.048, 0.083	0.010, 0.050, 0.089	0.010, 0.042, 0.080
0.70	ave-W	0.105, 0.193, 0.266	0.084, 0.138, 0.191	0.083, 0.138, 0.186	0.092, 0.172, 0.206
0.70	exp-W	0.019, 0.084, 0.162	0.009, 0.051, 0.103	0.009, 0.059, 0.093	0.013, 0.056, 0.105
0.70	sup-LM	0.002, 0.018, 0.049	0.003, 0.026, 0.058	0.007, 0.034, 0.076	0.007, 0.036, 0.072
0.70	ave-LM	0.073, 0.141, 0.202	0.071, 0.122, 0.164	0.078, 0.131, 0.178	0.088, 0.163, 0.200
0.70	exp-LM	0.002, 0.025, 0.059	0.003, 0.033, 0.066	0.006, 0.044, 0.080	0.012, 0.050, 0.095

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 26: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.3$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.014, 0.066, 0.119	0.012, 0.049, 0.098	0.011, 0.064, 0.111	0.011, 0.060, 0.099
0.00	ave-W	0.014, 0.066, 0.119	0.012, 0.049, 0.098	0.011, 0.064, 0.111	0.011, 0.060, 0.099
0.00	exp-W	0.014, 0.066, 0.119	0.012, 0.049, 0.098	0.011, 0.064, 0.111	0.011, 0.060, 0.099
0.00	sup-LM	0.008, 0.057, 0.106	0.009, 0.046, 0.090	0.009, 0.061, 0.109	0.009, 0.058, 0.096
0.00	ave-LM	0.008, 0.057, 0.106	0.009, 0.046, 0.090	0.009, 0.061, 0.109	0.009, 0.058, 0.096
0.00	exp-LM	0.008, 0.057, 0.106	0.009, 0.046, 0.090	0.009, 0.061, 0.109	0.009, 0.058, 0.096
0.35	sup-W	0.017, 0.065, 0.122	0.011, 0.064, 0.119	0.014, 0.059, 0.117	0.011, 0.058, 0.108
0.35	ave-W	0.017, 0.057, 0.112	0.012, 0.059, 0.107	0.013, 0.055, 0.114	0.011, 0.050, 0.104
0.35	exp-W	0.016, 0.054, 0.113	0.009, 0.056, 0.111	0.016, 0.056, 0.112	0.010, 0.051, 0.101
0.35	sup-LM	0.005, 0.048, 0.088	0.009, 0.053, 0.112	0.012, 0.057, 0.113	0.010, 0.051, 0.106
0.35	ave-LM	0.012, 0.045, 0.097	0.010, 0.049, 0.101	0.013, 0.054, 0.111	0.011, 0.048, 0.101
0.35	exp-LM	0.007, 0.045, 0.089	0.009, 0.049, 0.101	0.012, 0.053, 0.105	0.009, 0.047, 0.098
0.70	sup-W	0.023, 0.073, 0.150	0.007, 0.068, 0.134	0.016, 0.061, 0.127	0.014, 0.054, 0.100
0.70	ave-W	0.014, 0.074, 0.135	0.008, 0.057, 0.110	0.011, 0.062, 0.130	0.011, 0.056, 0.105
0.70	exp-W	0.023, 0.071, 0.143	0.008, 0.061, 0.125	0.012, 0.059, 0.122	0.014, 0.050, 0.101
0.70	sup-LM	0.001, 0.043, 0.086	0.005, 0.045, 0.108	0.010, 0.051, 0.117	0.011, 0.051, 0.098
0.70	ave-LM	0.004, 0.050, 0.107	0.004, 0.046, 0.094	0.009, 0.057, 0.121	0.009, 0.056, 0.103
0.70	exp-LM	0.003, 0.041, 0.096	0.006, 0.041, 0.108	0.010, 0.051, 0.107	0.011, 0.047, 0.097

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 27: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.3$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.016, 0.060, 0.109	0.011, 0.058, 0.103	0.004, 0.051, 0.096	0.008, 0.052, 0.107
0.00	ave-W	0.017, 0.062, 0.117	0.012, 0.069, 0.121	0.009, 0.053, 0.097	0.013, 0.055, 0.105
0.00	exp-W	0.016, 0.058, 0.109	0.011, 0.061, 0.111	0.004, 0.047, 0.096	0.008, 0.054, 0.107
0.00	sup-LM	0.010, 0.049, 0.093	0.008, 0.051, 0.098	0.004, 0.044, 0.092	0.007, 0.050, 0.109
0.00	ave-LM	0.010, 0.046, 0.102	0.012, 0.062, 0.110	0.008, 0.045, 0.090	0.011, 0.054, 0.104
0.00	exp-LM	0.010, 0.048, 0.096	0.008, 0.054, 0.104	0.004, 0.048, 0.089	0.007, 0.050, 0.106
0.35	sup-W	0.014, 0.072, 0.123	0.018, 0.062, 0.122	0.007, 0.051, 0.112	0.008, 0.042, 0.104
0.35	ave-W	0.009, 0.053, 0.117	0.015, 0.059, 0.113	0.013, 0.059, 0.112	0.012, 0.046, 0.105
0.35	exp-W	0.007, 0.059, 0.114	0.015, 0.053, 0.119	0.008, 0.053, 0.101	0.007, 0.044, 0.097
0.35	sup-LM	0.004, 0.039, 0.094	0.012, 0.049, 0.097	0.007, 0.047, 0.103	0.007, 0.041, 0.096
0.35	ave-LM	0.005, 0.038, 0.091	0.011, 0.050, 0.100	0.013, 0.053, 0.108	0.012, 0.044, 0.100
0.35	exp-LM	0.003, 0.035, 0.085	0.010, 0.040, 0.096	0.007, 0.043, 0.095	0.006, 0.041, 0.089
0.70	sup-W	0.014, 0.083, 0.157	0.010, 0.073, 0.122	0.009, 0.064, 0.117	0.013, 0.052, 0.117
0.70	ave-W	0.030, 0.087, 0.151	0.018, 0.080, 0.139	0.007, 0.056, 0.113	0.014, 0.066, 0.126
0.70	exp-W	0.011, 0.083, 0.144	0.012, 0.068, 0.118	0.008, 0.055, 0.110	0.014, 0.054, 0.116
0.70	sup-LM	0.006, 0.027, 0.084	0.005, 0.053, 0.094	0.006, 0.051, 0.101	0.009, 0.044, 0.104
0.70	ave-LM	0.011, 0.056, 0.111	0.011, 0.065, 0.122	0.007, 0.051, 0.101	0.012, 0.060, 0.123
0.70	exp-LM	0.005, 0.036, 0.094	0.007, 0.053, 0.099	0.005, 0.042, 0.102	0.014, 0.048, 0.106

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 28: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.3$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.014, 0.067, 0.123	0.009, 0.052, 0.101	0.007, 0.057, 0.115	0.015, 0.055, 0.113
0.00	ave-W	0.024, 0.069, 0.121	0.010, 0.046, 0.101	0.015, 0.058, 0.115	0.023, 0.066, 0.128
0.00	exp-W	0.014, 0.071, 0.125	0.010, 0.054, 0.104	0.007, 0.059, 0.118	0.015, 0.056, 0.117
0.00	sup-LM	0.006, 0.047, 0.102	0.004, 0.043, 0.094	0.005, 0.051, 0.111	0.014, 0.053, 0.110
0.00	ave-LM	0.012, 0.052, 0.102	0.009, 0.042, 0.089	0.012, 0.053, 0.111	0.021, 0.064, 0.127
0.00	exp-LM	0.006, 0.051, 0.101	0.005, 0.043, 0.095	0.005, 0.053, 0.109	0.015, 0.055, 0.116
0.35	sup-W	0.012, 0.071, 0.147	0.017, 0.056, 0.118	0.012, 0.050, 0.115	0.011, 0.056, 0.109
0.35	ave-W	0.019, 0.074, 0.150	0.020, 0.072, 0.124	0.014, 0.061, 0.102	0.011, 0.055, 0.107
0.35	exp-W	0.010, 0.068, 0.131	0.014, 0.052, 0.111	0.013, 0.046, 0.115	0.013, 0.048, 0.111
0.35	sup-LM	0.003, 0.037, 0.084	0.010, 0.043, 0.089	0.011, 0.042, 0.099	0.008, 0.050, 0.101
0.35	ave-LM	0.013, 0.049, 0.107	0.011, 0.061, 0.111	0.014, 0.055, 0.093	0.009, 0.052, 0.104
0.35	exp-LM	0.002, 0.038, 0.088	0.010, 0.040, 0.096	0.009, 0.038, 0.104	0.013, 0.048, 0.108
0.70	sup-W	0.012, 0.070, 0.147	0.012, 0.064, 0.119	0.014, 0.061, 0.132	0.013, 0.054, 0.107
0.70	ave-W	0.024, 0.089, 0.167	0.016, 0.064, 0.128	0.018, 0.062, 0.139	0.012, 0.056, 0.110
0.70	exp-W	0.018, 0.079, 0.149	0.014, 0.058, 0.128	0.014, 0.062, 0.119	0.012, 0.051, 0.093
0.70	sup-LM	0.001, 0.023, 0.063	0.008, 0.034, 0.087	0.012, 0.051, 0.106	0.009, 0.049, 0.095
0.70	ave-LM	0.012, 0.046, 0.103	0.009, 0.048, 0.099	0.014, 0.058, 0.120	0.008, 0.052, 0.103
0.70	exp-LM	0.002, 0.023, 0.081	0.009, 0.035, 0.087	0.007, 0.051, 0.098	0.010, 0.046, 0.088

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 29: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.3$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.012, 0.061, 0.117	0.016, 0.062, 0.123	0.011, 0.068, 0.122	0.011, 0.050, 0.108
0.00	ave-W	0.010, 0.069, 0.142	0.018, 0.063, 0.123	0.022, 0.069, 0.131	0.011, 0.066, 0.114
0.00	exp-W	0.013, 0.064, 0.123	0.015, 0.058, 0.116	0.010, 0.071, 0.124	0.012, 0.053, 0.107
0.00	sup-LM	0.006, 0.043, 0.087	0.012, 0.047, 0.103	0.008, 0.059, 0.110	0.009, 0.049, 0.105
0.00	ave-LM	0.006, 0.051, 0.103	0.012, 0.055, 0.109	0.019, 0.054, 0.123	0.010, 0.062, 0.113
0.00	exp-LM	0.005, 0.041, 0.094	0.012, 0.050, 0.105	0.008, 0.064, 0.115	0.009, 0.047, 0.105
0.35	sup-W	0.020, 0.082, 0.171	0.009, 0.066, 0.136	0.013, 0.059, 0.112	0.016, 0.056, 0.103
0.35	ave-W	0.022, 0.081, 0.161	0.019, 0.071, 0.128	0.018, 0.064, 0.126	0.013, 0.061, 0.120
0.35	exp-W	0.018, 0.076, 0.146	0.010, 0.062, 0.122	0.012, 0.051, 0.098	0.016, 0.052, 0.109
0.35	sup-LM	0.005, 0.036, 0.083	0.004, 0.046, 0.096	0.011, 0.049, 0.094	0.015, 0.051, 0.093
0.35	ave-LM	0.011, 0.054, 0.101	0.013, 0.060, 0.107	0.015, 0.057, 0.115	0.011, 0.060, 0.118
0.35	exp-LM	0.005, 0.036, 0.089	0.004, 0.041, 0.094	0.008, 0.043, 0.088	0.016, 0.046, 0.100
0.70	sup-W	0.013, 0.069, 0.129	0.011, 0.056, 0.116	0.010, 0.054, 0.107	0.015, 0.062, 0.110
0.70	ave-W	0.020, 0.077, 0.146	0.018, 0.073, 0.134	0.017, 0.069, 0.130	0.012, 0.051, 0.098
0.70	exp-W	0.013, 0.067, 0.114	0.013, 0.052, 0.117	0.011, 0.056, 0.108	0.012, 0.057, 0.111
0.70	sup-LM	0.003, 0.016, 0.049	0.003, 0.030, 0.074	0.007, 0.038, 0.081	0.012, 0.049, 0.103
0.70	ave-LM	0.003, 0.041, 0.086	0.010, 0.048, 0.107	0.012, 0.058, 0.115	0.012, 0.040, 0.088
0.70	exp-LM	0.004, 0.022, 0.055	0.002, 0.031, 0.069	0.007, 0.041, 0.082	0.011, 0.051, 0.095

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.3$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 30: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.6$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.016, 0.073, 0.120	0.015, 0.058, 0.114	0.011, 0.051, 0.117	0.014, 0.054, 0.099
0.00	ave-W	0.016, 0.073, 0.120	0.015, 0.058, 0.114	0.011, 0.051, 0.117	0.014, 0.054, 0.099
0.00	exp-W	0.016, 0.073, 0.120	0.015, 0.058, 0.114	0.011, 0.051, 0.117	0.014, 0.054, 0.099
0.00	sup-LM	0.010, 0.058, 0.108	0.015, 0.052, 0.110	0.011, 0.047, 0.106	0.010, 0.054, 0.099
0.00	ave-LM	0.010, 0.058, 0.108	0.015, 0.052, 0.110	0.011, 0.047, 0.106	0.010, 0.054, 0.099
0.00	exp-LM	0.010, 0.058, 0.108	0.015, 0.052, 0.110	0.011, 0.047, 0.106	0.010, 0.054, 0.099
0.35	sup-W	0.023, 0.084, 0.133	0.015, 0.059, 0.117	0.009, 0.065, 0.119	0.004, 0.036, 0.104
0.35	ave-W	0.022, 0.065, 0.133	0.010, 0.059, 0.109	0.006, 0.053, 0.104	0.003, 0.046, 0.095
0.35	exp-W	0.021, 0.071, 0.123	0.011, 0.057, 0.113	0.008, 0.050, 0.109	0.004, 0.041, 0.096
0.35	sup-LM	0.013, 0.055, 0.112	0.008, 0.046, 0.103	0.009, 0.050, 0.109	0.006, 0.036, 0.100
0.35	ave-LM	0.012, 0.053, 0.105	0.008, 0.052, 0.103	0.007, 0.043, 0.095	0.005, 0.042, 0.092
0.35	exp-LM	0.015, 0.054, 0.106	0.008, 0.051, 0.104	0.007, 0.047, 0.100	0.004, 0.040, 0.096
0.70	sup-W	0.022, 0.095, 0.161	0.010, 0.069, 0.130	0.012, 0.052, 0.105	0.014, 0.067, 0.130
0.70	ave-W	0.022, 0.087, 0.154	0.012, 0.067, 0.110	0.012, 0.050, 0.104	0.018, 0.076, 0.120
0.70	exp-W	0.022, 0.093, 0.157	0.012, 0.065, 0.116	0.012, 0.052, 0.098	0.012, 0.069, 0.127
0.70	sup-LM	0.006, 0.051, 0.102	0.005, 0.048, 0.107	0.006, 0.045, 0.095	0.010, 0.059, 0.125
0.70	ave-LM	0.006, 0.063, 0.129	0.009, 0.057, 0.096	0.012, 0.047, 0.094	0.012, 0.073, 0.115
0.70	exp-LM	0.005, 0.054, 0.108	0.006, 0.049, 0.095	0.007, 0.041, 0.089	0.012, 0.059, 0.119

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 31: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.6$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.017, 0.068, 0.117	0.013, 0.062, 0.120	0.012, 0.054, 0.107	0.012, 0.054, 0.115
0.00	ave-W	0.022, 0.079, 0.131	0.018, 0.071, 0.129	0.018, 0.068, 0.122	0.025, 0.069, 0.110
0.00	exp-W	0.017, 0.070, 0.117	0.013, 0.065, 0.119	0.012, 0.054, 0.109	0.013, 0.059, 0.113
0.00	sup-LM	0.008, 0.060, 0.098	0.012, 0.052, 0.110	0.011, 0.049, 0.098	0.012, 0.054, 0.112
0.00	ave-LM	0.014, 0.066, 0.107	0.013, 0.062, 0.123	0.015, 0.062, 0.118	0.021, 0.070, 0.110
0.00	exp-LM	0.009, 0.058, 0.100	0.012, 0.053, 0.107	0.011, 0.050, 0.106	0.012, 0.058, 0.112
0.35	sup-W	0.019, 0.082, 0.140	0.017, 0.072, 0.133	0.014, 0.064, 0.113	0.008, 0.051, 0.103
0.35	ave-W	0.020, 0.066, 0.128	0.027, 0.087, 0.153	0.015, 0.077, 0.126	0.020, 0.065, 0.109
0.35	exp-W	0.015, 0.073, 0.128	0.013, 0.076, 0.130	0.012, 0.062, 0.117	0.010, 0.057, 0.097
0.35	sup-LM	0.009, 0.047, 0.103	0.010, 0.055, 0.111	0.011, 0.058, 0.104	0.007, 0.050, 0.099
0.35	ave-LM	0.010, 0.056, 0.109	0.016, 0.078, 0.130	0.015, 0.075, 0.120	0.017, 0.065, 0.110
0.35	exp-LM	0.006, 0.046, 0.099	0.011, 0.060, 0.122	0.007, 0.052, 0.107	0.009, 0.053, 0.095
0.70	sup-W	0.020, 0.088, 0.151	0.014, 0.067, 0.122	0.013, 0.064, 0.111	0.007, 0.036, 0.081
0.70	ave-W	0.033, 0.107, 0.161	0.019, 0.072, 0.125	0.029, 0.071, 0.121	0.016, 0.066, 0.120
0.70	exp-W	0.024, 0.084, 0.157	0.013, 0.070, 0.118	0.013, 0.061, 0.116	0.008, 0.040, 0.092
0.70	sup-LM	0.003, 0.038, 0.085	0.007, 0.045, 0.099	0.009, 0.054, 0.106	0.005, 0.033, 0.079
0.70	ave-LM	0.015, 0.075, 0.126	0.015, 0.065, 0.108	0.026, 0.068, 0.113	0.012, 0.063, 0.115
0.70	exp-LM	0.006, 0.040, 0.095	0.008, 0.049, 0.099	0.006, 0.055, 0.106	0.005, 0.037, 0.085

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 32: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.6$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.020, 0.074, 0.123	0.012, 0.059, 0.105	0.013, 0.050, 0.099	0.010, 0.043, 0.086
0.00	ave-W	0.022, 0.076, 0.143	0.020, 0.069, 0.115	0.018, 0.076, 0.124	0.018, 0.055, 0.105
0.00	exp-W	0.022, 0.076, 0.131	0.013, 0.063, 0.108	0.013, 0.053, 0.106	0.010, 0.046, 0.093
0.00	sup-LM	0.007, 0.054, 0.097	0.007, 0.047, 0.095	0.011, 0.048, 0.096	0.008, 0.040, 0.084
0.00	ave-LM	0.017, 0.058, 0.122	0.015, 0.059, 0.105	0.018, 0.071, 0.120	0.018, 0.055, 0.102
0.00	exp-LM	0.009, 0.056, 0.098	0.007, 0.053, 0.097	0.011, 0.048, 0.103	0.008, 0.045, 0.091
0.35	sup-W	0.021, 0.099, 0.171	0.014, 0.075, 0.137	0.014, 0.056, 0.112	0.008, 0.055, 0.108
0.35	ave-W	0.049, 0.118, 0.179	0.039, 0.105, 0.160	0.031, 0.076, 0.120	0.024, 0.078, 0.126
0.35	exp-W	0.019, 0.097, 0.161	0.012, 0.077, 0.137	0.016, 0.056, 0.109	0.007, 0.054, 0.102
0.35	sup-LM	0.010, 0.053, 0.119	0.007, 0.053, 0.118	0.011, 0.048, 0.098	0.006, 0.050, 0.102
0.35	ave-LM	0.033, 0.099, 0.145	0.028, 0.100, 0.147	0.029, 0.073, 0.115	0.021, 0.078, 0.121
0.35	exp-LM	0.007, 0.062, 0.116	0.006, 0.057, 0.119	0.012, 0.050, 0.098	0.006, 0.047, 0.095
0.70	sup-W	0.016, 0.078, 0.151	0.014, 0.068, 0.135	0.011, 0.052, 0.106	0.009, 0.049, 0.104
0.70	ave-W	0.039, 0.125, 0.188	0.036, 0.105, 0.169	0.027, 0.084, 0.146	0.031, 0.081, 0.134
0.70	exp-W	0.015, 0.075, 0.148	0.015, 0.070, 0.124	0.013, 0.053, 0.104	0.006, 0.051, 0.097
0.70	sup-LM	0.001, 0.022, 0.064	0.010, 0.038, 0.089	0.007, 0.048, 0.083	0.003, 0.046, 0.093
0.70	ave-LM	0.016, 0.075, 0.139	0.027, 0.089, 0.145	0.026, 0.076, 0.139	0.028, 0.074, 0.131
0.70	exp-LM	0.003, 0.026, 0.070	0.010, 0.049, 0.098	0.005, 0.042, 0.088	0.004, 0.045, 0.091

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 33: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.6$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.016, 0.080, 0.149	0.020, 0.067, 0.106	0.008, 0.057, 0.103	0.011, 0.049, 0.097
0.00	ave-W	0.030, 0.095, 0.169	0.036, 0.107, 0.154	0.026, 0.084, 0.146	0.022, 0.077, 0.133
0.00	exp-W	0.017, 0.082, 0.145	0.021, 0.069, 0.119	0.009, 0.057, 0.115	0.011, 0.052, 0.108
0.00	sup-LM	0.007, 0.044, 0.113	0.013, 0.057, 0.097	0.006, 0.051, 0.093	0.009, 0.047, 0.093
0.00	ave-LM	0.018, 0.073, 0.122	0.034, 0.092, 0.138	0.025, 0.078, 0.142	0.019, 0.074, 0.132
0.00	exp-LM	0.011, 0.050, 0.108	0.014, 0.061, 0.103	0.006, 0.051, 0.108	0.010, 0.048, 0.105
0.35	sup-W	0.015, 0.072, 0.142	0.017, 0.074, 0.141	0.011, 0.050, 0.104	0.009, 0.060, 0.112
0.35	ave-W	0.032, 0.106, 0.177	0.040, 0.109, 0.168	0.018, 0.080, 0.132	0.034, 0.096, 0.167
0.35	exp-W	0.015, 0.066, 0.137	0.020, 0.070, 0.135	0.011, 0.050, 0.104	0.009, 0.053, 0.118
0.35	sup-LM	0.007, 0.034, 0.071	0.008, 0.053, 0.100	0.008, 0.038, 0.089	0.008, 0.050, 0.106
0.35	ave-LM	0.010, 0.071, 0.138	0.032, 0.093, 0.147	0.017, 0.072, 0.120	0.032, 0.095, 0.159
0.35	exp-LM	0.005, 0.037, 0.078	0.011, 0.055, 0.099	0.009, 0.040, 0.094	0.009, 0.050, 0.112
0.70	sup-W	0.011, 0.077, 0.138	0.009, 0.058, 0.111	0.013, 0.051, 0.100	0.003, 0.052, 0.111
0.70	ave-W	0.046, 0.130, 0.206	0.032, 0.117, 0.160	0.037, 0.093, 0.146	0.033, 0.078, 0.130
0.70	exp-W	0.013, 0.075, 0.137	0.007, 0.051, 0.110	0.011, 0.057, 0.110	0.006, 0.048, 0.099
0.70	sup-LM	0.000, 0.012, 0.044	0.001, 0.025, 0.065	0.006, 0.035, 0.078	0.001, 0.047, 0.101
0.70	ave-LM	0.017, 0.077, 0.139	0.025, 0.089, 0.139	0.025, 0.085, 0.135	0.028, 0.076, 0.124
0.70	exp-LM	0.000, 0.017, 0.055	0.003, 0.030, 0.067	0.009, 0.046, 0.095	0.007, 0.039, 0.090

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.6$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 34: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.9$, $\bar{d} = 1$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.015, 0.063, 0.128	0.017, 0.069, 0.123	0.010, 0.061, 0.123	0.013, 0.051, 0.108
0.00	ave-W	0.015, 0.063, 0.128	0.017, 0.069, 0.123	0.010, 0.061, 0.123	0.013, 0.051, 0.108
0.00	exp-W	0.015, 0.063, 0.128	0.017, 0.069, 0.123	0.010, 0.061, 0.123	0.013, 0.051, 0.108
0.00	sup-LM	0.011, 0.053, 0.115	0.013, 0.058, 0.119	0.009, 0.061, 0.117	0.012, 0.050, 0.108
0.00	ave-LM	0.011, 0.053, 0.115	0.013, 0.058, 0.119	0.009, 0.061, 0.117	0.012, 0.050, 0.108
0.00	exp-LM	0.011, 0.053, 0.115	0.013, 0.058, 0.119	0.009, 0.061, 0.117	0.012, 0.050, 0.108
0.35	sup-W	0.030, 0.100, 0.164	0.020, 0.087, 0.152	0.016, 0.065, 0.123	0.011, 0.040, 0.091
0.35	ave-W	0.024, 0.076, 0.146	0.020, 0.069, 0.132	0.014, 0.053, 0.116	0.011, 0.042, 0.093
0.35	exp-W	0.025, 0.083, 0.154	0.017, 0.072, 0.132	0.013, 0.055, 0.119	0.011, 0.040, 0.090
0.35	sup-LM	0.013, 0.072, 0.129	0.007, 0.071, 0.133	0.013, 0.055, 0.118	0.011, 0.037, 0.088
0.35	ave-LM	0.018, 0.060, 0.119	0.013, 0.056, 0.116	0.012, 0.050, 0.108	0.009, 0.041, 0.090
0.35	exp-LM	0.012, 0.061, 0.126	0.009, 0.060, 0.119	0.015, 0.053, 0.110	0.011, 0.038, 0.091
0.70	sup-W	0.033, 0.135, 0.222	0.017, 0.074, 0.124	0.012, 0.065, 0.128	0.016, 0.060, 0.123
0.70	ave-W	0.032, 0.124, 0.203	0.015, 0.069, 0.122	0.009, 0.068, 0.121	0.012, 0.061, 0.121
0.70	exp-W	0.031, 0.125, 0.213	0.014, 0.069, 0.115	0.007, 0.061, 0.121	0.014, 0.060, 0.125
0.70	sup-LM	0.004, 0.057, 0.147	0.006, 0.048, 0.102	0.007, 0.053, 0.110	0.014, 0.053, 0.113
0.70	ave-LM	0.013, 0.090, 0.160	0.011, 0.061, 0.101	0.005, 0.055, 0.117	0.012, 0.062, 0.117
0.70	exp-LM	0.006, 0.067, 0.152	0.009, 0.052, 0.100	0.006, 0.049, 0.112	0.012, 0.053, 0.121

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1\}$ (i.e., $\bar{d} = 1$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 35: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.9$, $\bar{d} = 2$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.020, 0.069, 0.127	0.016, 0.058, 0.107	0.009, 0.059, 0.112	0.009, 0.048, 0.088
0.00	ave-W	0.043, 0.099, 0.160	0.031, 0.089, 0.144	0.026, 0.094, 0.159	0.026, 0.080, 0.120
0.00	exp-W	0.022, 0.073, 0.135	0.017, 0.062, 0.114	0.011, 0.067, 0.125	0.010, 0.053, 0.102
0.00	sup-LM	0.010, 0.055, 0.101	0.013, 0.050, 0.094	0.008, 0.055, 0.112	0.008, 0.046, 0.089
0.00	ave-LM	0.037, 0.082, 0.143	0.027, 0.080, 0.141	0.022, 0.088, 0.155	0.026, 0.080, 0.120
0.00	exp-LM	0.012, 0.060, 0.112	0.015, 0.056, 0.103	0.008, 0.064, 0.123	0.008, 0.053, 0.100
0.35	sup-W	0.027, 0.097, 0.166	0.015, 0.060, 0.120	0.011, 0.053, 0.106	0.012, 0.050, 0.084
0.35	ave-W	0.047, 0.123, 0.195	0.033, 0.090, 0.146	0.030, 0.089, 0.152	0.026, 0.087, 0.131
0.35	exp-W	0.031, 0.090, 0.162	0.013, 0.073, 0.114	0.009, 0.057, 0.110	0.010, 0.054, 0.099
0.35	sup-LM	0.010, 0.055, 0.116	0.007, 0.049, 0.097	0.008, 0.040, 0.098	0.010, 0.048, 0.082
0.35	ave-LM	0.037, 0.095, 0.160	0.023, 0.086, 0.134	0.027, 0.083, 0.144	0.027, 0.085, 0.130
0.35	exp-LM	0.010, 0.057, 0.122	0.006, 0.054, 0.101	0.008, 0.055, 0.105	0.008, 0.052, 0.094
0.70	sup-W	0.024, 0.092, 0.172	0.014, 0.065, 0.140	0.011, 0.044, 0.088	0.012, 0.049, 0.102
0.70	ave-W	0.064, 0.138, 0.211	0.046, 0.114, 0.170	0.030, 0.094, 0.153	0.036, 0.083, 0.125
0.70	exp-W	0.024, 0.098, 0.161	0.017, 0.072, 0.126	0.010, 0.046, 0.109	0.011, 0.057, 0.100
0.70	sup-LM	0.001, 0.041, 0.085	0.007, 0.049, 0.100	0.008, 0.033, 0.075	0.011, 0.042, 0.093
0.70	ave-LM	0.038, 0.108, 0.166	0.034, 0.095, 0.154	0.027, 0.087, 0.140	0.032, 0.080, 0.122
0.70	exp-LM	0.002, 0.048, 0.108	0.009, 0.046, 0.105	0.008, 0.040, 0.094	0.009, 0.052, 0.097

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2\}$ (i.e., $\bar{d} = 2$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 36: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.9$, $\bar{d} = 4$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.017, 0.069, 0.146	0.018, 0.062, 0.112	0.008, 0.041, 0.086	0.010, 0.048, 0.087
0.00	ave-W	0.069, 0.157, 0.214	0.065, 0.131, 0.181	0.041, 0.116, 0.174	0.056, 0.115, 0.153
0.00	exp-W	0.020, 0.083, 0.168	0.021, 0.073, 0.132	0.009, 0.053, 0.106	0.012, 0.059, 0.109
0.00	sup-LM	0.006, 0.050, 0.102	0.010, 0.055, 0.100	0.007, 0.039, 0.083	0.009, 0.047, 0.087
0.00	ave-LM	0.054, 0.135, 0.197	0.053, 0.120, 0.171	0.041, 0.109, 0.169	0.055, 0.115, 0.150
0.00	exp-LM	0.010, 0.059, 0.132	0.014, 0.066, 0.114	0.006, 0.048, 0.104	0.010, 0.056, 0.107
0.35	sup-W	0.023, 0.089, 0.152	0.007, 0.066, 0.119	0.009, 0.060, 0.108	0.009, 0.035, 0.070
0.35	ave-W	0.091, 0.167, 0.221	0.064, 0.135, 0.191	0.064, 0.138, 0.193	0.059, 0.115, 0.166
0.35	exp-W	0.027, 0.107, 0.158	0.009, 0.058, 0.130	0.011, 0.066, 0.122	0.010, 0.047, 0.097
0.35	sup-LM	0.009, 0.047, 0.097	0.002, 0.051, 0.088	0.006, 0.053, 0.094	0.007, 0.033, 0.064
0.35	ave-LM	0.065, 0.143, 0.199	0.051, 0.119, 0.181	0.061, 0.134, 0.186	0.055, 0.110, 0.162
0.35	exp-LM	0.011, 0.059, 0.127	0.007, 0.051, 0.112	0.006, 0.061, 0.116	0.009, 0.043, 0.095
0.70	sup-W	0.020, 0.093, 0.168	0.012, 0.064, 0.118	0.002, 0.042, 0.089	0.007, 0.043, 0.085
0.70	ave-W	0.087, 0.178, 0.239	0.073, 0.159, 0.214	0.066, 0.138, 0.184	0.070, 0.131, 0.177
0.70	exp-W	0.021, 0.099, 0.173	0.014, 0.068, 0.136	0.005, 0.057, 0.107	0.011, 0.045, 0.105
0.70	sup-LM	0.001, 0.029, 0.065	0.003, 0.030, 0.081	0.002, 0.032, 0.073	0.005, 0.037, 0.075
0.70	ave-LM	0.058, 0.128, 0.201	0.055, 0.144, 0.189	0.059, 0.131, 0.181	0.067, 0.126, 0.174
0.70	exp-LM	0.003, 0.039, 0.095	0.007, 0.047, 0.104	0.003, 0.043, 0.093	0.009, 0.042, 0.095

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, \}$ (i.e., $\bar{d} = 4$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 37: Empirical size of the bootstrap tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.9$, $\psi_0 = 0.9$, $\bar{d} = 8$)

		$n = 125$	$n = 250$	$n = 500$	$n = 1000$
κ	test	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
0.00	sup-W	0.019, 0.068, 0.126	0.013, 0.049, 0.105	0.013, 0.050, 0.093	0.011, 0.041, 0.084
0.00	ave-W	0.105, 0.182, 0.237	0.077, 0.153, 0.200	0.081, 0.150, 0.203	0.075, 0.136, 0.173
0.00	exp-W	0.024, 0.084, 0.168	0.014, 0.066, 0.139	0.020, 0.065, 0.115	0.011, 0.054, 0.104
0.00	sup-LM	0.008, 0.043, 0.098	0.009, 0.038, 0.091	0.011, 0.046, 0.086	0.010, 0.038, 0.082
0.00	ave-LM	0.080, 0.159, 0.210	0.069, 0.149, 0.190	0.078, 0.137, 0.200	0.074, 0.133, 0.172
0.00	exp-LM	0.010, 0.054, 0.127	0.010, 0.050, 0.121	0.018, 0.058, 0.104	0.011, 0.054, 0.097
0.35	sup-W	0.027, 0.098, 0.177	0.013, 0.057, 0.105	0.020, 0.056, 0.097	0.008, 0.042, 0.081
0.35	ave-W	0.126, 0.222, 0.285	0.093, 0.168, 0.208	0.084, 0.158, 0.207	0.088, 0.142, 0.203
0.35	exp-W	0.037, 0.113, 0.189	0.012, 0.064, 0.126	0.017, 0.066, 0.111	0.014, 0.052, 0.102
0.35	sup-LM	0.005, 0.042, 0.091	0.007, 0.034, 0.075	0.015, 0.043, 0.084	0.008, 0.035, 0.076
0.35	ave-LM	0.099, 0.188, 0.247	0.082, 0.153, 0.194	0.080, 0.152, 0.202	0.088, 0.142, 0.199
0.35	exp-LM	0.008, 0.058, 0.118	0.010, 0.044, 0.097	0.016, 0.055, 0.102	0.010, 0.050, 0.095
0.70	sup-W	0.023, 0.091, 0.170	0.011, 0.060, 0.112	0.011, 0.045, 0.093	0.004, 0.037, 0.087
0.70	ave-W	0.141, 0.247, 0.330	0.106, 0.176, 0.227	0.108, 0.174, 0.224	0.091, 0.158, 0.194
0.70	exp-W	0.022, 0.104, 0.184	0.017, 0.077, 0.131	0.012, 0.060, 0.112	0.011, 0.059, 0.104
0.70	sup-LM	0.003, 0.020, 0.061	0.004, 0.033, 0.065	0.008, 0.030, 0.073	0.002, 0.029, 0.075
0.70	ave-LM	0.089, 0.189, 0.253	0.097, 0.158, 0.203	0.102, 0.163, 0.211	0.089, 0.153, 0.192
0.70	exp-LM	0.004, 0.029, 0.076	0.005, 0.040, 0.097	0.009, 0.050, 0.096	0.008, 0.054, 0.098

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.9$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n \in \{125, 250, 500, 1000\}$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter d is $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e., $\bar{d} = 8$). Space of μ: $\mathcal{X}_{\kappa,n} = \{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \dots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\}$, where $x_{[1]} \leq \cdots \leq x_{[n]}$ is a sorted version of x and $\kappa \in \{0.00, 0.35, 0.70\}$. We perform the bootstrap sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.

Table 38: Empirical size of the bootstrap average tests for the no-threshold-effect hypothesis based on the TAR model ($\phi_0 = 0.6$, $\psi_0 = 0.9$, $\kappa = 0$, $n = 1000$)

		Average Wald test			Average LM test		
Choice space D	$\#D$	1%	5%	10%	1%	5%	10%
$\{1\}$	1	0.011	0.048	0.103	0.010	0.048	0.101
$\{4\}$	1	0.010	0.058	0.110	0.010	0.054	0.107
$\{8\}$	1	0.014	0.057	0.112	0.014	0.054	0.112
$\{1,2\}$	2	0.023	0.070	0.119	0.020	0.066	0.116
$\{4,5\}$	2	0.017	0.072	0.128	0.015	0.069	0.128
$\{7,8\}$	2	0.033	0.079	0.121	0.031	0.076	0.118
$\{1,2,3,4\}$	4	0.064	0.124	0.170	0.063	0.121	0.169
$\{2,4,6,8\}$	4	0.041	0.091	0.150	0.040	0.092	0.148
$\{5,6,7,8\}$	4	0.036	0.090	0.135	0.036	0.090	0.131
$\{1,2,3,4,5,6,7,8\}$	8	0.076	0.133	0.184	0.073	0.133	0.181

DGP: $y_t = \phi_0 y_{t-1} + \epsilon_t$, $x_t = \psi_0 x_{t-1} + \nu_t$, $\phi_0 = 0.6$, $\psi_0 = 0.9$, $(\epsilon_t, \nu_t)^\top \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Sample size: $n = 1000$. TAR Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $x_{t-d} < \mu$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $x_{t-d} \geq \mu$. The choice space of the delay parameter, D, takes various values. The threshold parameter μ is fixed at the meadian value of x (i.e., $\kappa = 0$). We perform the bootstrap ave-Wald and ave-LM tests for the no-threshold-effect hypothesis $H_0 : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $a \in \{0.01, 0.05, 0.10\}$; the number of bootstrap samples is $B = 500$. This table reports the empirical size of the tests across $J = 1000$ Monte Carlo samples.