# Supplemental material for "*Conditional Threshold Autoregression (CoTAR)*"

Kaiji Motegi* John W. Dennis† Shigeyuki Hamori‡

Kobe University IDA Yamato U. & Kobe U.

This draft: July 16, 2024

## 1 Introduction

In this supplemental material, we present additional numerical and empirical studies that are omitted from the main paper. In Section 2, we present a full version of Monte Carlo simulations to complement the numerical evidence reported in the main paper. In Section 3, we analyze a pandemic of novel coronavirus disease 2019 (COVID-19) in Japan and the United States (U.S.). We use the following notation throughout: $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers, $\lfloor a \rfloor$ is the largest integer not larger than $a \in \mathbb{R}$, the Euclidean norm of any $k$-dimensional vector $\boldsymbol{c} \in \mathbb{R}^k$ is denoted as $\|\boldsymbol{c}\| = (\boldsymbol{c}^\top \boldsymbol{c})^{1/2}$, $\mathbf{1}(A)$ is the indicator function which equals 1 if event $A$ occurs and 0 otherwise, and convergence in probability is denoted by $\xrightarrow{p}$.

## 2 Monte Carlo simulation: A full version

In this section, we present full Monte Carlo simulations to evaluate the finite sample performance of our proposed model: *Self-Exciting Conditional Threshold Autoregression (SE-CoTAR)*. To keep this section self-contained, we review and extend the set-up of the simulation study in the main paper. In Section 2.1, we examine the performance of the profiling

*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

‡Faculty of Political Science and Economics, Yamato University. Kobe University (Professor Emeritus). E-mail: hamori.shigeyuki@yamato-u.ac.jp

estimator. In Section 2.2, we inspect the empirical size and power of the wild-bootstrap tests for the no-threshold-effect hypothesis $H_0^*$. In Section 2.3, we report rejection frequencies of the Diebold-Mariano tests for the equal predictive accuracy hypothesis $H_0^{eq}$ between SE-CoTAR and two benchmark models: AR and Tong's (1978) Self-Exciting Threshold Autoregression (SETAR).

## 2.1 Profiling estimation

Suppose that the data generating process (DGP) is SE-CoTAR with lag length $p = 1$:

$$y_t = \begin{cases} \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t & \text{if } \ y_{t-d_0} < \mu_{t-d_0-1}(c_0), \\ \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t & \text{if } \ y_{t-d_0} \geq \mu_{t-d_0-1}(c_0), \end{cases} \tag{1}$$

where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $d_0 = 1$, $c_0 = 0.5$, and $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. The conditional threshold $\mu_t(c_0)$ takes the $mc_0$-th smallest value (i.e., almost the median) of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. The memory size $m \in \{6, 18\}$ is assumed to be known. We considered only $m = 6$ in the main paper, and we are adding $m = 18$ here to understand the role of memory size.

The regression parameters in regime 2 are assumed to have a local drift toward those in regime 1:

$$\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}, \qquad \phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}, \tag{2}$$

where $\lambda_0 \neq 0$ and $\delta_0 \geq 0$. Let $\boldsymbol{\beta}_{10} = (\alpha_{10}, \phi_{10})^\top$ and $\boldsymbol{\beta}_{20n} = (\alpha_{20n}, \phi_{20n})^\top$, then $\boldsymbol{\beta}_{20n} - \boldsymbol{\beta}_{10} = O(n^{-\delta_0})$. As elaborated by Andrews and Cheng (2012), $\delta_0$ is a key quantity that determines the identification category of the nuisance parameter $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$. The smaller value of $\delta_0$ implies the slower convergence of $\boldsymbol{\beta}_{20n}$ to $\boldsymbol{\beta}_{10}$ and hence the greater identifiability of $\boldsymbol{\gamma}_0$. We consider five values for $\delta_0$:

**Case #1 (non-identification):** $\delta_0 \to \infty$, in which case $\boldsymbol{\gamma}_0$ is *unidentified* as $\boldsymbol{\beta}_{20n} = \boldsymbol{\beta}_{10}$ for any $n \in \mathbb{N}$.

**Case #2 (weak identification):** $\delta_0 = 0.75$, in which case $\boldsymbol{\gamma}_0$ is *weakly identified* as $\boldsymbol{\beta}_{20n} \to \boldsymbol{\beta}_{10}$ as $n \to \infty$ at a sufficiently fast rate.

**Case #3 (weak identification, boundary case):** $\delta_0 = 0.5$, which is the boundary case of weak identification.

**Case #4 (semi-strong identification):** $\delta_0 = 0.25$, in which case $\boldsymbol{\gamma}_0$ is *semi-strongly identified* as $\boldsymbol{\beta}_{20n} \to \boldsymbol{\beta}_{10}$ at a sufficiently slow rate.

**Case #5 (strong identification):** $\delta_0 = 0$, in which case $\boldsymbol{\gamma}_0$ is *strongly identified* as $\boldsymbol{\beta}_{20n} = \boldsymbol{\beta}_{10} + \lambda_0 \boldsymbol{\iota}_2$ for any $n \in \mathbb{N}$, where $\lambda_0 \neq 0$ and $\boldsymbol{\iota}_2 = (1,1)^\top$.

In the main paper, we consider only Case #1 (non-identification) and Case #5 (strong identification) out of the five cases to save space; Case #1 here matches Case 1 in the main paper, and Case #5 here matches Case 2 in the main paper. In this section, we additionally consider Cases #2–#4 to better understand the small-sample and large-sample behavior of our proposed methods across the identification categories.

The sample size is chosen from $n \in \{125, 250, 500, 1000, 2000, 4000, 8000\}$. We consider the various sample sizes to fully observe subtle differences across Cases #1–#5. For Case #1, the value of $\lambda_0$ is irrelevant since $\delta_0 \to \infty$. We set $\lambda_0 \in \{13.08, 3.91, 1.17, 0.35\}$ for Cases #2–#5, respectively. Given these values of $\lambda_0$, $\boldsymbol{\beta}_{20,125} \approx (0.35, 0.55)^\top$ for these four cases, which makes comparison easier. As $n$ grows, $\boldsymbol{\beta}_{20n}$ converges to $\boldsymbol{\beta}_{10} = (0, 0.2)^\top$ at rates $n^{0.75}$, $n^{0.5}$, and $n^{0.25}$ in Cases #2–#4, respectively; $\boldsymbol{\beta}_{20n}$ is constant at $(0.35, 0.55)$ in Case #5. These parameterizations are summarized in Table 1. Let $\Delta_{0n} = \|\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20n}\|$ be the Euclidean norm between the true regression parameters in regimes 1 and 2. It serves as an overall measure of heterogeneity between the two regimes. In Case #1, $\Delta_{0n} = 0$ for all $n$. We have that $\Delta_{0,8000} \in \{0.022, 0062, 0.175\}$ in Cases #2–#4, respectively. In Case #5, $\Delta_{0n} = 0.495$ for all $n$.

The SE-CoTAR model with $p = 1$ is specified as

$$y_t = \begin{cases} \alpha_1 + \phi_1 y_{t-1} + u_t & \text{if } \; y_{t-d} < \mu_{t-d-1}(c), \\ \alpha_2 + \phi_2 y_{t-1} + u_t & \text{if } \; y_{t-d} \geq \mu_{t-d-1}(c), \end{cases} \tag{3}$$

where $\mu_t(c)$ is the conditional threshold described in (1). The choice space of the delay parameter $d$ is $D = \{1, 2, 3\}$. The largest possible choice space of the percentile parameter $c$ would be $\bar{C} = \{1/m, 2/m, \ldots, 1\}$, but it might cause an identification issue in small samples since a certain regime might occur too few times. To restrict $\bar{C}$, define binary indicators which determine regimes:

$$I_{1t}(c) = \mathbf{1}\left\{y_t < \mu_{t-1}(c)\right\}, \qquad I_{2t}(c) = \mathbf{1}\left\{y_t \geq \mu_{t-1}(c)\right\}.$$

Let $\delta_r(c) = n^{-1}\sum_{t=1}^{n} I_{rt}(c)$ be the share of regime $r \in \{1,2\}$ to the entire sample. The choice space of the percentile parameter $c$ is specified as

$$C = \left\{ c \in \bar{C} \mid \min\{\delta_1(c), \delta_2(c)\} > 0.15 \right\}. \tag{4}$$

This specification ensures that each regime accounts for at least 15% of the entire sample, following a well-known suggestion of Andrews (1993).

We fit model (3) to each of $J = 1000$ Monte Carlo samples generated from DGP (1). We then estimate the regression parameters $\boldsymbol{\beta} = (\alpha_1, \phi_1, \alpha_2, \phi_2)^\top$ and the nuisance parameters $\boldsymbol{\gamma} = (d, c)^\top$ via profiling, as described in Section 3.1 of the main paper. The resulting estimator is denoted as $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$. We report the bias and standard deviation of the profiling estimator for each element of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. The bias is computed with respect to $\boldsymbol{\theta}_{0n} = (\boldsymbol{\beta}_{0n}^\top, \boldsymbol{\gamma}_0^\top)^\top$, where $\boldsymbol{\beta}_{0n} = (\boldsymbol{\beta}_{10}^\top, \boldsymbol{\beta}_{20n}^\top)^\top$. Inspection of the performance of the profiling estimator was omitted from the main paper, since we put more emphasis on testing the no-threshold effect hypothesis $H_0^*$ and the equal predictive accuracy hypothesis $H_0^{eq}$. We fill this gap here for completeness.

Simulation results for Cases #1–#5 are shown in Tables 2-6, respectively. The bias of $\hat{\boldsymbol{\beta}}$ is negligibly small for all cases considered, suggesting $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}_{0n}$. Further, the standard deviation of $\hat{\boldsymbol{\beta}}$ vanishes as $n$ grows. The standard deviation of $\hat{\phi}_2$ in Case 1 with $m = 6$, for example, is $\{0.255, 0.175, 0.123, \ldots, 0.030\}$ for $n \in \{125, 250, 500, \ldots, 8000\}$, respectively (Table 2). These results are in line with the implication of Theorem 1.$(i)$ of the main paper that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{0n} \xrightarrow{p} \mathbf{0}$ irrespective of identification categories.

The stronger threshold effect facilitates estimation of the regression parameter $\boldsymbol{\beta}_{0n}$. Fixing $m = 6$ and $n = 500$, the standard deviation of $\hat{\phi}_2$ is $\{0.123, 0.113, 0.110, 0.092, 0.065\}$ for Cases #1–#5, respectively (Tables 2-6). Another finding is that the larger value in the memory size has an adverse impact on the small sample performance of $\hat{\boldsymbol{\beta}}$, although the consistency of $\hat{\boldsymbol{\beta}}$ is preserved. Fixing $m = 18$ and $n = 500$, the standard deviation of $\hat{\phi}_2$ is $\{0.142, 0.133, 0.119, 0.105, 0.079\}$ for Cases #1–#5. The standard deviation increases, albeit slightly, for all five cases when $m$ increases from 6 to 18. A reason for the worse small sample performance is likely due to the fact that the larger memory size requires more initial observations for implementing the estimation.

The profiling estimator for the nuisance parameter, $\hat{\boldsymbol{\gamma}}$, does not converge to $\boldsymbol{\gamma}_0$ under Cases #1–#3 but it does converge under Cases #4–#5. Taking $m = 6$ as an in-

stance, the standard deviation of the estimated percentile parameter $\hat{c}$ under Case #1 is $\{0.261, 0.262, 0.249, \ldots, 0.268\}$ for $n \in \{125, 250, 500, \ldots, 8000\}$, respectively (Table 2). The standard deviation is almost constant across sample sizes, suggesting that $\hat{\boldsymbol{\gamma}} = O_p(1)$ under Case #1. Similar patterns appear in Cases #2–#3 (Tables 3-4). By contrast, the standard deviation is $\{0.218, 0.192, 0.183, \ldots, 0.049\}$ in Case #4 and $\{0.217, 0.139, 0.061, \ldots, 0.000\}$ in Case #5, monotonically approaching 0 (Tables 5-6).

Intuitively, the consistency of $\hat{\boldsymbol{\gamma}}$ holds if there exist sufficiently strong threshold effects (i.e., if $\boldsymbol{\beta}_{20n}$ converges to $\boldsymbol{\beta}_{10}$ at a sufficiently slow rate). Our simulation results are indicative of non-convergence under non-identification and weak identification (i.e., $\hat{\boldsymbol{\gamma}} = O_p(1)$ for $\delta_0 \in [0.5, \infty)$), and indicative of convergence under semi-strong and strong identification (i.e., $\hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$ for $\delta_0 \in [0, 0.5)$). We observe similar patterns when the memory size is $m = 18$. These results are in line with the implication of Theorem 1.$(ii)$ of the main paper that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}_{10})$ is *not* asymptotically normal under non- and weak identification, where $\bar{\boldsymbol{\beta}}_{10} = (\boldsymbol{\beta}_{10}^\top, \boldsymbol{\beta}_{10}^\top)^\top$. In summary, the simulation results in Tables 2-6 are all reasonable, and we can conclude that the profiling estimator performs well in small and large samples.

## 2.2 Testing the no-threshold-effect hypothesis

As in the previous section, suppose that the DGP is given by (1)-(2) and the model is given by (3). Consider testing the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, where $\boldsymbol{\beta}_r = (\alpha_r, \phi_r)^\top$ is the regression parameters in regime $r \in \{1, 2\}$. We adopt the wild-bootstrap tests of Hansen (1996) since $\boldsymbol{\gamma}_0$ is unidentified under $H_0^*$. For comparison, the sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM tests are all implemented. We use the heteroscedasticity-robust covariance matrix estimator when computing the conditional Wald or LM test statistic, although the true error $\epsilon$ is homoscedastic. The exact procedure of the bootstrap tests is described in Section 3.2 of the main paper.

Rejection frequencies of each test are computed across $J = 1000$ Monte Carlo samples, where the nominal size is $a = 0.05$ and the number of bootstrap samples is $B = 500$. The rejection frequencies can be interpreted as empirical size for Case #1 (non-identification) and empirical power for Cases #2–#5. We expect that the rejection frequencies should increase as we proceed from Case #1 to #5, since threshold effects become stronger.

Rejection frequencies associated with Cases #1–#5 are reported in Tables 7-11, respectively. Under Case #1, the empirical size of the six tests all converges to the nominal size as $n \to \infty$, confirming Theorem 2.$(i)$ of the main paper. The Wald tests are over-

sized when $n \leq 250$, although the size distortions diminish as $n$ grows. The empirical size of the exp-Wald test with $m = 6$, for instance, is $\{0.191, 0.099, 0.071, \ldots, 0.048\}$ for $n \in \{125, 250, 500, \ldots, 8000\}$, respectively (Table 7). The sup-Wald test leads to similar size distortions, while the ave-Wald test is slightly better for $n = 125$. The LM tests, by contrast, achieve sharp empirical size for all sample sizes considered. The empirical size of the exp-LM test with $m = 6$, for example, is $\{0.046, 0.028, 0.047, \ldots, 0.046\}$ . The sup-LM and ave-LM tests are roughly comparable with the exp-LM test in terms of empirical size. Hence, it is advised to use the LM tests instead of the Wald tests to control the type-I error rate in small samples.

We now discuss empirical power of the LM tests. For each fixed sample size, the empirical power is lowest in Case #2 and highest in Case #5. This result agrees with our conjecture, since threshold effects become stronger as we proceed from #2 to #5. Focusing on $n = 500$ and $m = 6$, the empirical power of the exp-LM test is $\{0.104, 0.218, 0.552, 0.973\}$ for Cases #2–#5, respectively (Tables 7-11).

The empirical power of each test *decreases* as $n$ grows under Cases #2–#3, while it approaches 1 under Cases #4–#5. The decreasing power in Cases #2–#3 reflects the sufficiently fast convergence of $\boldsymbol{\beta}_{20n}$ to $\boldsymbol{\beta}_{10}$, while the increasing power in Cases #4–#5 reflects the sufficiently slow convergence (#4) or non-convergence (#5). These results are in line with Theorem 2.$(ii)$ of the main paper, which states that the bootstrap tests have power approaching 1 under semi-strong and strong identification. The contrast between #2–#3 and #4–#5 also agrees with our previous observation that $\hat{\boldsymbol{\gamma}}$ does not converge to $\boldsymbol{\gamma}_0$ when $\delta_0 \geq 0.5$ but it does converge when $\delta_0 < 0.5$ (Section 2.1). Focusing on $n = 8000$ and $m = 6$, the empirical power of the exp-LM test is $\{0.057, 0.180, 0.993, 1.000\}$ for Cases #2–#5, respectively (Tables 7-11).

When the memory size increases from $m = 6$ to $m = 18$, we observe the following results. First, the Wald tests keep producing size distortions in small samples, while the LM tests keep accurate empirical size (Table 7). Second, the empirical power of each test becomes slightly lower. It is not surprising that the larger value of $m$ negatively affects the small sample performance, as we need more initial observations to execute the test. Taking Case #4 and $n = 500$ as an example, the empirical power of the exp-LM test is 0.552 for $m = 6$ and 0.479 for $m = 18$ (Table 10). Overall, the simulation results in Tables 7-11 are all reasonable, and we can conclude that the LM tests achieve desired size and power properties in small and large samples.

## 2.3 Testing the equal predictive accuracy hypothesis

In this section, we compare the out-of-sample forecast performance of the SE-CoTAR model (3) and two benchmark models: AR and SETAR. The AR model is specified as $y_t = \alpha + \phi y_{t-1} + u_t$, and the SETAR model is specified as

$$y_t = \begin{cases} \alpha_1 + \phi_1 y_{t-1} + u_t & \text{if } \ y_{t-d} < \mu, \\ \alpha_2 + \phi_2 y_{t-1} + u_t & \text{if } \ y_{t-d} \geq \mu. \end{cases}$$

We consider three scenarios to make fair comparison between SE-CoTAR and the two benchmarks. First, the DGP is SE-CoTAR (1)-(2) and the benchmark model is AR. Second, the DGP is SE-CoTAR and the benchmark model is SETAR. Third, the DGP is SETAR and the benchmark model is also SETAR. For the third scenario, the DGP is given by

$$y_t = \begin{cases} \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t & \text{if } \ y_{t-d_0} < \mu_0, \\ \alpha_{20} + \phi_{20} y_{t-1} + \epsilon_t & \text{if } \ y_{t-d_0} \geq \mu_0, \end{cases}$$

where $\mu_0 = 0.1$ and the other parameters are analogous to (1)-(2).

We perform rolling window one-step-ahead prediction with the whole sample size $N \in \{125, 250, 500, 1000, 2000, 4000\}$. The window size is fixed at $n = 0.8N$. For each of $J = 1000$ Monte Carlo samples, we perform the asymptotic $S_1$-test of Diebold and Mariano (1995) [DM1995] with nominal size $a = 0.05$. The precise procedures of the TAR model, rolling window analysis, and the DM test are provided in Section 3.3 of the main paper.

Our conjecture on the first scenario is as follows. When the true DGP is AR (Case #1), the two-regime structure of the SE-CoTAR model is redundant and the AR model is exactly specified. Hence, we expect under Case #1 that rejecting $H_0^{eq} : MSE^{ar} = MSE^{secotar}$ in favor of $H_1^{ar} : MSE^{ar} < MSE^{secotar}$ should occur more frequently than rejecting $H_0^{eq}$ in favor of $H_1^{secotar} : MSE^{ar} > MSE^{secotar}$. When there are strong threshold effects (Case #5), SE-CoTAR is correctly specified and AR is misspecified. Hence, rejection frequencies in favor of $H_1^{eq} : MSE^{ar} \neq MSE^{secotar}$ and $H_1^{secotar}$ should approach 1 as $N$ grows. Similarly, rejection frequencies in favor of $H_1^{ar}$ should approach 0 under Case #5. We expect to observe gradual transition of rejection rates as we proceed from Case #1 to #5.

The following results are expected to appear in the second scenario. Under Case #1,

both SE-CoTAR and SETAR models are correctly specified relative to the true AR process, although their two-regime structures are redundant. We therefore expect under Case #1 that rejection rates of $H_0^{eq} : MSE^{setar} = MSE^{secotar}$ should be close to 0 whichever the alternative hypothesis is. Under Case #5, the SE-CoTAR model is correctly specified and the SETAR model is misspecified as its constant threshold deviates from the truth of conditional threshold. We therefore expect under Case #5 that rejection rates in favor of $H_1^{eq} : MSE^{setar} \neq MSE^{secotar}$ and $H_1^{secotar} : MSE^{setar} > MSE^{secotar}$ should approach 1 and those in favor of $H_1^{setar} : MSE^{setar} < MSE^{secotar}$ should approach 0 as $N$ grows.

In the third scenario, the null and alternative hypotheses of the DM test are specified in the same way as the second scenario. Under Case #1, rejection rates of $H_0^{eq}$ should be close to 0 whichever the alternative hypothesis is. Under Case #5, the SETAR model is correctly specified and the SE-CoTAR model is misspecified as its conditional threshold deviates from the truth of constant threshold. Hence, under Case #5, rejection rates in favor of $H_1^{eq}$ and $H_1^{setar}$ should approach 1 and those in favor of $H_1^{secotar}$ should approach 0 as $N$ grows.

Rejection frequencies of the DM test associated with scenarios 1-3 are reported in Tables 12-14, respectively. We begin with the first scenario, where the DGP is SE-CoTAR and the benchmark model is AR. We focus on $m = 6$, as results with $m = 18$ are almost identical. Under Case #1 (non-identification), the DM test chooses $H_1^{ar}$ more often than it chooses $H_1^{secotar}$, reflecting the correct and parsimonious specification of AR. This result is consistent with our conjecture that the redundant two-regime structure of SE-CoTAR should be penalized by the DM test. Fixing $N = 4000$, the rejection frequency of $H_0^{eq}$ is $\{0.105, 0.007, 0.159\}$ for $\{H_1^{eq}, H_1^{secotar}, H_1^{ar}\}$, respectively (Table 12).

Rejections of $H_0^{eq}$ do not happen frequently for Cases #2–#4. This suggests that, when threshold effects are asymptotically vanishing, the DM test does not decisively choose either $H_1^{secotar}$ or $H_1^{ar}$. Focusing on $N = 4000$ and $H_1^{eq}$, the rejection frequency is $\{0.081, 0.076, 0.115\}$ for Cases #2–#4, respectively (Table 12). Under Case #5 (strong identification), the DM test chooses $H_1^{secotar}$ with probability approaching 1 as $N$ grows. This is a desired result since SE-CoTAR is correctly specified and AR is misspecified under Case #5. Fixing $N = 1000$, the rejection frequency of $H_0^{eq}$ is $\{0.329, 0.474, 0.000\}$ for $\{H_1^{eq}, H_1^{secotar}, H_1^{ar}\}$, respectively. When we raise the sample size from 1000 to 4000, the rejection frequency becomes $\{0.925, 0.958, 0.000\}$.

We next discuss the second scenario, where the DGP is SE-CoTAR and the benchmark model is SETAR. For Case #1, rejection frequencies are all below 10% (Table 13). This

is reasonable since the true DGP is AR and there is no reason why one of $MSE^{secotar}$ and $MSE^{setar}$ should be smaller than the other. For Cases #2–#4, rejection frequencies are generally low. This result indicates that asymptotically vanishing conditional threshold effects are too weak for the DM test to detect the subtle difference between SE-CoTAR and SETAR. Focusing on $H_1^{eq}$ with $(m, N) = (6, 4000)$, the rejection frequency is $\{0.042, 0.047, 0.140\}$ for Cases #2–#4, respectively. For Case #5, the DM test selects $H_1^{secotar}$ with probability approaching 1 as $N$ grows. Focusing on $m = 6$, the rejection frequency in favor of $H_1^{secotar}$ is $\{0.469, 0.663, 0.892\}$ for $N \in \{1000, 2000, 4000\}$, respectively.

When we raise the memory size $m$ from 6 to 18, the difference between SE-CoTAR and SETAR becomes less salient. This is reasonable since the larger memory size regulates the fluctuation of conditional threshold $\mu_t(c)$. In fact, SETAR can be interpreted as a limit case of SE-CoTAR with $m \to \infty$. Replacing $m = 6$ with $m = 18$ in the previous example, the rejection frequencies in favor of $H_1^{secotar}$ are $\{0.333, 0.514, 0.758\}$ (Table 13). By switching from $m = 6$ to $m = 18$, the rejection rate decreases by approximately 15% for each $N \in \{1000, 2000, 4000\}$.

We now discuss the third scenario, where the DGP is SETAR and the benchmark model is SETAR. For Cases #1–#4, rejection frequencies are all below 10%, indicating that $MSE^{secotar}$ and $MSE^{setar}$ are sufficiently close to each other under missing or vanishing constant threshold effects (Table 14). Curiously, we observe low rejection rates even for Case #5. The small sample predictive ability of the SE-CoTAR model is comparable with SETAR even when the true DGP is SETAR, indicating a practical value of SE-CoTAR. For Case #5 with $m = 6$, the rejection rates of $H_0^{eq}$ in favor of $H_1^{setar}$ are $\{0.122, 0.260, 0.413\}$ for $N \in \{1000, 2000, 4000\}$, respectively. These are much lower than the corresponding rejection rates in favor of $H_1^{secotar}$ in the second scenario: $\{0.469, 0.663, 0.892\}$ (Table 13).

As in the second scenario, the difference between SE-CoTAR and SETAR becomes less noticeable when $m$ increases. For Case #5 with $m = 18$, rejection rates in favor of $H_1^{setar}$ are $\{0.109, 0.210, 0.313\}$ for $N \in \{1000, 2000, 4000\}$ (Table 14). In summary, the simulation results in Tables 12-14 are all reasonable, and the Diebold-Mariano test is an appropriate approach for comparing the out-of-sample MSEs of the SE-CoTAR model and the benchmark AR or SETAR model. In particular, it is worth emphasizing that the small sample predictive ability of SE-CoTAR is comparable with SETAR even when the truth is constant threshold.

# 3 An empirical application on COVID-19

In Section 6 of the main paper, we analyze the Volatility Index of S&P 500 and major U.S. firms, showing a practical value of the SE-CoTAR model. This section presents another empirical application of SE-CoTAR where we study the COVID-19 pandemic in Japan and the U.S. The two applications are different in terms of data but analogous in terms of statistical procedures, showing the wide applicability of our proposed methods.

Modelling and predicting the spread of the novel coronavirus were of utmost concern around 2020–2022, and are still an urgent research topic as of this writing. There is a growing literature in which time series methods are applied to COVID-19 data (e.g., Chimmula and Zhang, 2020, Zeroual, Harrou, Dairi, and Sun, 2020, Medeiros, Street, Valladão, Vasconcelos, and Zilberman, 2022, Petropoulos, Makridakis, and Stylianou, 2022). In particular, Aidoo, Ampofo, Awashie, Appiah, and Adebanji (2021) fitted smooth transition autoregressive (STAR) models to daily new confirmed COVID-19 cases in the African sub-region, detecting nonlinear effects.

The number of new confirmed cases is expected to follow a self-exciting process, since individuals feel tempted to take tests in response to an emerging pandemic. Such a built-in acceleration mechanism likely produces time-varying threshold effects, motivating the use of SE-CoTAR. We therefore fit the SE-CoTAR model to daily new confirmed COVID-19 cases. In Section 3.1, we present our data and some preliminary analyses. In Section 3.2, we perform an in-sample analysis with a special emphasis on testing the no-threshold-effect hypothesis $H_0^*$. In Section 3.3, we perform an out-of-sample analysis using the Diebold-Mariano test for the equal predictive accuracy hypothesis $H_0^{eq}$.

## 3.1 Data and preliminary analysis

Our target variable is daily new confirmed COVID-19 cases per million people in Japan and the U.S. The data are publicly available at Our World in Data (OWID). The raw data, `new_cases_per_million`, have strong weekday effects (e.g., the number of new confirmed cases tends to be smaller on weekends and holidays due to the smaller number of tests).[1] A seasonally adjusted version (`new_cases_smoothed_per_million`) is also available at OWID, in which the weekday effects are smoothed out. We analyze the seasonally adjusted version,

[1]There might exist some other seasonal patterns at longer horizons such as summer or winter effects, but they are not considered in this paper given the fact that it has been only a few years since the pandemic began in 2020.

denoted by $\{w_t\}_{t=1}^n$, from April 4, 2020 through June 23, 2021 ($n = 446$ days).[2]

In Figure 1, we plot the log-difference of the new confirmed cases: $y_t = \ln w_t - \ln w_{t-1}$. It exhibits rather complex fluctuations with persistent swings and temporary noise being combined, which is indicative of nonlinear effects. Sample statistics of $y_t$ are reported in Table 15. The mean and median are positive for Japan and negative for the U.S. The Japanese series is more volatile than the U.S. series in terms of range (i.e., spread between the maximum and minimum), standard deviation, and kurtosis. We observe a curious contrast that the Japanese series is negatively skewed whereas the U.S. series is positively skewed.

## 3.2 Testing the no-threshold-effect hypothesis

For each country, we fit a SE-CoTAR model with lag length $p = 2$:

$$y_t = \begin{cases} \alpha_1 + \phi_{11} y_{t-1} + \phi_{12} y_{t-2} + u_t & \text{if } \ y_{t-d} < \mu_{t-d-1}(c), \\ \alpha_2 + \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + u_t & \text{if } \ y_{t-d} \geq \mu_{t-d-1}(c), \end{cases} \tag{5}$$

where $y$ is the log-difference of daily new confirmed COVID-19 cases. The conditional threshold $\mu_t(c)$ takes the $mc$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. The memory size is set to be $m = 14$ days, and the choice space of the delay parameter $d$ is $D = \{1, 2, \ldots, 14\}$. These configurations correspond to a common perception that people's activity affects the status of pandemic roughly in two weeks. The choice space of the percentile parameter $c$ is given by (4) so that each regime accounts for at least 15% of the entire sample.

Based on model (5), we test the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, where $\boldsymbol{\beta}_1 = (\alpha_1, \phi_{11}, \phi_{12})^\top$ and $\boldsymbol{\beta}_2 = (\alpha_2, \phi_{21}, \phi_{22})^\top$. As described in Section 3.2 of the main paper, we adopt the wild-bootstrap tests of Hansen (1996) since the nuisance parameters $\boldsymbol{\gamma} = (d, c)^\top$ are unidentified under $H_0^*$. We use the exp-LM test as it exhibits the better size and power properties than other tests in our simulation study (Section 2.2). To address potential heteroscedasticity, we use the robust covariance matrix estimator. The number of bootstrap samples is $B = 5000$.

For comparison, we replace the SE-CoTAR model with SETAR, and perform the bootstrap exp-LM test for $H_0^*$. The SETAR model is obtained by replacing the conditional threshold $\mu_t(c)$ in (5) with the constant threshold $\mu$. The choice space for $\mu$ is designed in a

[2]The seasonal adjustment might affect the nonlinear dynamics of the raw series, but we take the seasonally adjusted series as given.

way that each regime accounts for at least 15% of the whole sample; see Section 3.3 of the main paper for details. The remaining procedure is the same as SE-CoTAR.

Bootstrap p-values associated with $H_0^*$ are reported in Table 16. For Japan, the p-value is 0.075 for the SE-CoTAR model and 0.003 for the SETAR model. Conditional threshold effects are significant at the 10% level but not at the 5% level, while constant threshold effects are significant at the 1% level. Hence, the validity of constant threshold effects relative to non-threshold effects seems slightly stronger than the validity of conditional threshold effects.

For the U.S., the p-value is 0.042 for SE-CoTAR and 0.123 for SETAR (Table 16). Conditional threshold effects are significant at the 5% level, while constant threshold effects are insignificant at the 10% level. The validity of conditional threshold effects relative to non-threshold effects seems slightly stronger than the validity of constant threshold effects. For both countries, the in-sample explanatory power likely improves by incorporating threshold effects, whether they are constant or conditional. The contrast in the relative validity of constant versus conditional threshold effects suggests that Japan and the U.S. are heterogeneous in terms of the COVID-19 pandemic.

We add the estimated conditional threshold $\mu_t(\hat{c})$ and estimated constant threshold $\hat{\mu}$ to Figure 1. We see $\mu_t(\hat{c})$ tracing the actual path of the log-difference of the new confirmed cases, which is a key feature of the SE-CoTAR model. For the SETAR model, $\hat{\mu}$ is often far away from $y_t$ and hence regime changes do not occur frequently.

## 3.3 Testing the equal predictive accuracy hypothesis

In this section, we conduct out-of-sample analyses in accordance with Section 3.3 of the main paper. Our main model is SE-CoTAR with $p = 2$, and we consider a constant-only model $y_t = \alpha + u_t$, an AR(2) model, and the SETAR model with $p = 2$ as benchmarks. For each model, we perform rolling window one-step-ahead forecast and compute root mean squared forecast error (RMSE). Further, we execute the asymptotic $S_1$-test of DM1995 for two scenarios. First, we test $H_0^{eq} : MSE^{ar} = MSE^{secotar}$ (i.e., the equal predictive accuracy hypothesis between AR and SE-CoTAR). Second, we test $H_0^{eq} : MSE^{setar} = MSE^{secotar}$.

Empirical results of the out-of-sample analyses are summarized in Table 16. For Japan, RMSEs of the four models are as follows.

$$RMSE_{jp}^{const} = 0.0423, \qquad RMSE_{jp}^{ar} = 0.0191,$$
$$RMSE_{jp}^{setar} = 0.0184, \qquad RMSE_{jp}^{secotar} = 0.0182.$$

The constant-only model leads to by far the largest MSE out of four, hence it is strongly advised to include AR terms to improve forecast precision. $RMSE_{jp}^{secotar}$ attains the minimum value out of four, indicating a practical value of SE-CoTAR.

The asymptotic p-value of the DM test for the equal predictive accuracy hypothesis $H_0^{eq}$ between AR and SE-CoTAR is $\{0.146, 0.927, 0.073\}$ for $\{H_1^{eq}, H_1^{ar}, H_1^{secotar}\}$, respectively (Table 16). The one-sided test rejects $H_0^{eq}$ in favor of $H_1^{secotar}$ at the 10% level, suggesting that the spread between $RMSE_{jp}^{ar} = 0.0191$ and $RMSE_{jp}^{secotar} = 0.0182$ is significant. The p-value between SETAR and SE-CoTAR is $\{0.796, 0.602, 0.398\}$ for $\{H_1^{eq}, H_1^{setar}, H_1^{secotar}\}$. This result suggests that the spread between $RMSE_{jp}^{setar} = 0.0184$ and $RMSE_{jp}^{secotar} = 0.0182$ is insignificant. These out-of-sample findings are consistent with the in-sample findings that both constant and conditional threshold effects are significant for Japan.

For the U.S., RMSEs of the four models are as follows (Table 16).

$$RMSE_{us}^{const} = 0.0321, \qquad RMSE_{us}^{ar} = 0.0214,$$
$$RMSE_{us}^{setar} = 0.0219, \qquad RMSE_{us}^{secotar} = 0.0204.$$

$RMSE_{us}^{secotar}$ attains by far the smallest value out of four, highlighting the relevance of conditional threshold effects. This finding is in line with the in-sample results, and supported by the DM test. The p-value between AR and SE-CoTAR is $\{0.033, 0.984, 0.016\}$ for $\{H_1^{eq}, H_1^{ar}, H_1^{secotar}\}$, respectively. Further, the p-value between SETAR and SE-CoTAR is $\{0.043, 0.979, 0.022\}$ for $\{H_1^{eq}, H_1^{setar}, H_1^{secotar}\}$. For both benchmark models, the two-sided test rejects $H_0^{eq}$ at the 5% level, and the one-sided test rejects $H_0^{eq}$ in favor of $H_1^{secotar}$ at the 5% level. These are strong evidence that the SE-CoTAR model outperforms the AR and SETAR models in the out-of-sample prediction.

In summary, the SE-CoTAR model achieves desired in-sample and out-of-sample performance for daily new confirmed COVID-19 cases of Japan and the U.S. The results are especially striking for the U.S., in which the SE-CoTAR model has higher predictive ability than the AR and SETAR models. This is a useful finding which helps healthcare professionals and policymakers better monitor the status of the pandemic. We also find some contrasting results between Japan and the U.S.; for example, the validity of constant threshold effects relative to conditional threshold effects is stronger for Japan than for the U.S. in both in-sample and out-of-sample frameworks. Potential sources of these contrasts include people's perception on the pandemic, the speed of vaccine roll-out, and the stringency of

lock-down policies. Identifying sources of the asymmetry is left as a future task.

# References


Aidoo, E. N., R. T. Ampofo, G. E. Awashie, S. K. Appiah, and A. O. Adebanji (2021): "Modelling COVID-19 incidence in the African sub-region using smooth transition autoregressive model," *Modeling Earth Systems and Environment*, https://doi.org/10.1007/s40808-021-01136-1.

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

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80(5), 2153–2211.

Chimmula, V. K. R., and L. Zhang (2020): "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons and Fractals*, 135, #109864.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

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

Medeiros, M. C., A. Street, D. Valladão, G. Vasconcelos, and E. Zilberman (2022): "Short-term Covid-19 forecast for latecomers," *International Journal of Forecasting*, 38, 467–488.

Petropoulos, F., S. Makridakis, and N. Stylianou (2022): "COVID-19: Forecasting confirmed cases and deaths with a simple time series model," *International Journal of Forecasting*, 38, 439–452.

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

Zeroual, A., F. Harrou, A. Dairi, and Y. Sun (2020): "Deep learning methods for forecasting COVID-19 time-series data: A comparative study," *Chaos, Solitons and Fractals*, 140, #110121.

Table 1: True values of regression parameters in regime 2 ($\alpha_{10} = 0$, $\phi_{10} = 0.2$)

| | | Case #1 | Case #2 | Case #3 | Case #4 | Case #5 |
|---|---|---|---|---|---|---|
| $n$ | $(\delta_0, \lambda_0)$ | $(\infty, \text{n.a.})$ | $(0.75, 13.08)$ | $(0.5, 3.91)$ | $(0.25, 1.17)$ | $(0, 0.35)$ |
| 125 | $\alpha_{20n}$ | 0 | 0.350 | 0.350 | 0.350 | 0.350 |
| 125 | $\phi_{20n}$ | 0.2 | 0.550 | 0.550 | 0.550 | 0.550 |
| 125 | $\Delta_{0n}$ | 0 | 0.495 | 0.495 | 0.495 | 0.495 |
| 250 | $\alpha_{20n}$ | 0 | 0.208 | 0.247 | 0.294 | 0.350 |
| 250 | $\phi_{20n}$ | 0.2 | 0.408 | 0.447 | 0.494 | 0.550 |
| 250 | $\Delta_{0n}$ | 0 | 0.294 | 0.350 | 0.416 | 0.495 |
| 500 | $\alpha_{20n}$ | 0 | 0.124 | 0.175 | 0.247 | 0.350 |
| 500 | $\phi_{20n}$ | 0.2 | 0.324 | 0.375 | 0.447 | 0.550 |
| 500 | $\Delta_{0n}$ | 0 | 0.175 | 0.247 | 0.350 | 0.495 |
| 1000 | $\alpha_{20n}$ | 0 | 0.074 | 0.124 | 0.208 | 0.350 |
| 1000 | $\phi_{20n}$ | 0.2 | 0.274 | 0.324 | 0.408 | 0.550 |
| 1000 | $\Delta_{0n}$ | 0 | 0.104 | 0.175 | 0.294 | 0.495 |
| 2000 | $\alpha_{20n}$ | 0 | 0.044 | 0.087 | 0.175 | 0.350 |
| 2000 | $\phi_{20n}$ | 0.2 | 0.244 | 0.287 | 0.375 | 0.550 |
| 2000 | $\Delta_{0n}$ | 0 | 0.062 | 0.124 | 0.247 | 0.495 |
| 4000 | $\alpha_{20n}$ | 0 | 0.026 | 0.062 | 0.147 | 0.350 |
| 4000 | $\phi_{20n}$ | 0.2 | 0.226 | 0.262 | 0.347 | 0.550 |
| 4000 | $\Delta_{0n}$ | 0 | 0.037 | 0.087 | 0.208 | 0.495 |
| 8000 | $\alpha_{20n}$ | 0 | 0.015 | 0.044 | 0.124 | 0.350 |
| 8000 | $\phi_{20n}$ | 0.2 | 0.215 | 0.244 | 0.324 | 0.550 |
| 8000 | $\Delta_{0n}$ | 0 | 0.022 | 0.062 | 0.175 | 0.495 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-1} < \mu_{t-2}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-1} \geq \mu_{t-2}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $m \in \{6, 18\}$, $c_0 = 0.5$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Regression parameters in regime 1 are $\boldsymbol{\beta}_{10} = (\alpha_{10}, \phi_{10})^\top$, whereas those in regime 2 are $\boldsymbol{\beta}_{20n} = (\alpha_{20n}, \phi_{20n})^\top$. Distance between the two regimes is measured as the Euclidean norm $\Delta_{0n} = \|\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20n}\|$. Sample size: $n \in \{125, 250, \ldots, 4000, 8000\}$. Case #1: non-identification ($\delta_0 \to \infty$, $\lambda_0 = $ n.a.). Case #2: weak identification ($\delta_0 = 0.75$, $\lambda_0 = 13.08$). Case #3: boundary case of weak identification ($\delta_0 = 0.5$, $\lambda_0 = 3.91$). Case #4: semi-strong identification ($\delta_0 = 0.25$, $\lambda_0 = 1.17$). Case #5: strong identification ($\delta_0 = 0$, $\lambda_0 = 0.35$).

Table 2: Simulation results on the profiling estimation (Case #1: non-identification)

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 6 | $\alpha_1$ | Bias | −0.023 | −0.003 | 0.002 | 0.006 | 0.000 | −0.002 | 0.002 |
| 6 | $\alpha_1$ | Stdev | 0.295 | 0.187 | 0.136 | 0.096 | 0.066 | 0.051 | 0.036 |
| 6 | $\phi_1$ | Bias | −0.029 | −0.012 | −0.008 | −0.001 | 0.000 | −0.002 | 0.001 |
| 6 | $\phi_1$ | Stdev | 0.259 | 0.175 | 0.121 | 0.082 | 0.058 | 0.044 | 0.030 |
| 6 | $\alpha_2$ | Bias | 0.028 | 0.000 | 0.003 | 0.008 | 0.003 | 0.001 | −0.001 |
| 6 | $\alpha_2$ | Stdev | 0.297 | 0.197 | 0.145 | 0.106 | 0.070 | 0.049 | 0.035 |
| 6 | $\phi_2$ | Bias | −0.025 | −0.007 | −0.004 | −0.003 | −0.005 | −0.001 | −0.001 |
| 6 | $\phi_2$ | Stdev | 0.255 | 0.175 | 0.123 | 0.089 | 0.058 | 0.042 | 0.030 |
| 6 | $d$ | Bias | 1.255 | 1.299 | 1.265 | 1.285 | 1.256 | 1.282 | 1.303 |
| 6 | $d$ | Stdev | 0.802 | 0.782 | 0.804 | 0.790 | 0.804 | 0.794 | 0.784 |
| 6 | $c$ | Bias | 0.088 | 0.082 | 0.092 | 0.090 | 0.092 | 0.077 | 0.083 |
| 6 | $c$ | Stdev | 0.261 | 0.262 | 0.249 | 0.264 | 0.250 | 0.259 | 0.268 |

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 18 | $\alpha_1$ | Bias | −0.015 | −0.000 | −0.002 | −0.008 | −0.001 | −0.001 | −0.002 |
| 18 | $\alpha_1$ | Stdev | 0.391 | 0.260 | 0.190 | 0.122 | 0.090 | 0.070 | 0.047 |
| 18 | $\phi_1$ | Bias | −0.032 | −0.020 | −0.007 | −0.005 | −0.001 | 0.002 | −0.001 |
| 18 | $\phi_1$ | Stdev | 0.299 | 0.203 | 0.148 | 0.096 | 0.075 | 0.055 | 0.036 |
| 18 | $\alpha_2$ | Bias | 0.018 | 0.001 | 0.005 | 0.007 | 0.001 | 0.001 | 0.001 |
| 18 | $\alpha_2$ | Stdev | 0.363 | 0.236 | 0.174 | 0.114 | 0.082 | 0.064 | 0.041 |
| 18 | $\phi_2$ | Bias | −0.037 | −0.015 | −0.008 | −0.008 | −0.002 | 0.000 | −0.001 |
| 18 | $\phi_2$ | Stdev | 0.287 | 0.194 | 0.142 | 0.094 | 0.066 | 0.049 | 0.034 |
| 18 | $d$ | Bias | 1.292 | 1.287 | 1.295 | 1.303 | 1.301 | 1.234 | 1.302 |
| 18 | $d$ | Stdev | 0.789 | 0.780 | 0.789 | 0.774 | 0.776 | 0.811 | 0.779 |
| 18 | $c$ | Bias | 0.032 | 0.022 | 0.013 | 0.027 | 0.004 | 0.020 | 0.010 |
| 18 | $c$ | Stdev | 0.240 | 0.246 | 0.252 | 0.248 | 0.256 | 0.254 | 0.257 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #1: $\delta_0 \to \infty$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is unidentified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. The profiling estimator for $(\alpha_1, \phi_1, \alpha_2, \phi_2, d, c)$ is computed. For each parameter, bias and standard deviation across $J = 1000$ Monte Carlo samples are reported.

Table 3: Simulation results on the profiling estimation (Case #2: weak identification)

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 6 | $\alpha_1$ | Bias | 0.053 | 0.072 | 0.046 | 0.031 | 0.019 | 0.015 | 0.008 |
| 6 | $\alpha_1$ | Stdev | 0.268 | 0.220 | 0.162 | 0.110 | 0.078 | 0.053 | 0.036 |
| 6 | $\phi_1$ | Bias | 0.027 | 0.050 | 0.036 | 0.024 | 0.017 | 0.013 | 0.007 |
| 6 | $\phi_1$ | Stdev | 0.249 | 0.198 | 0.137 | 0.096 | 0.065 | 0.044 | 0.031 |
| 6 | $\alpha_2$ | Bias | 0.031 | −0.016 | 0.001 | −0.004 | 0.001 | −0.001 | −0.000 |
| 6 | $\alpha_2$ | Stdev | 0.349 | 0.231 | 0.148 | 0.102 | 0.068 | 0.049 | 0.034 |
| 6 | $\phi_2$ | Bias | −0.033 | −0.006 | −0.008 | −0.008 | −0.003 | −0.001 | −0.000 |
| 6 | $\phi_2$ | Stdev | 0.190 | 0.152 | 0.113 | 0.082 | 0.060 | 0.042 | 0.030 |
| 6 | $d$ | Bias | 0.572 | 0.928 | 1.047 | 1.174 | 1.145 | 1.208 | 1.288 |
| 6 | $d$ | Stdev | 0.832 | 0.893 | 0.877 | 0.832 | 0.827 | 0.819 | 0.796 |
| 6 | $c$ | Bias | 0.046 | 0.051 | 0.076 | 0.072 | 0.071 | 0.076 | 0.091 |
| 6 | $c$ | Stdev | 0.213 | 0.253 | 0.262 | 0.270 | 0.266 | 0.261 | 0.267 |

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 18 | $\alpha_1$ | Bias | 0.045 | 0.062 | 0.038 | 0.020 | 0.020 | 0.012 | 0.007 |
| 18 | $\alpha_1$ | Stdev | 0.354 | 0.259 | 0.186 | 0.141 | 0.098 | 0.067 | 0.045 |
| 18 | $\phi_1$ | Bias | 0.028 | 0.040 | 0.034 | 0.018 | 0.015 | 0.009 | 0.006 |
| 18 | $\phi_1$ | Stdev | 0.329 | 0.230 | 0.156 | 0.111 | 0.079 | 0.054 | 0.037 |
| 18 | $\alpha_2$ | Bias | 0.075 | 0.013 | 0.012 | 0.005 | −0.005 | 0.001 | 0.001 |
| 18 | $\alpha_2$ | Stdev | 0.451 | 0.279 | 0.188 | 0.131 | 0.082 | 0.061 | 0.042 |
| 18 | $\phi_2$ | Bias | −0.058 | −0.023 | −0.015 | −0.007 | −0.001 | −0.001 | −0.002 |
| 18 | $\phi_2$ | Stdev | 0.231 | 0.175 | 0.133 | 0.098 | 0.066 | 0.048 | 0.033 |
| 18 | $d$ | Bias | 0.682 | 0.923 | 1.042 | 1.115 | 1.180 | 1.164 | 1.262 |
| 18 | $d$ | Stdev | 0.843 | 0.880 | 0.863 | 0.858 | 0.835 | 0.820 | 0.798 |
| 18 | $c$ | Bias | 0.007 | 0.008 | 0.019 | 0.015 | 0.001 | 0.021 | 0.006 |
| 18 | $c$ | Stdev | 0.200 | 0.221 | 0.242 | 0.246 | 0.248 | 0.247 | 0.248 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #2: $\delta_0 = 0.75$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is weakly identified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. The profiling estimator for $(\alpha_1, \phi_1, \alpha_2, \phi_2, d, c)$ is computed. For each parameter, bias and standard deviation across $J = 1000$ Monte Carlo samples are reported.

Table 4: Simulation results on the profiling estimation (Case #3: weak identification, boundary)

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 6 | $\alpha_1$ | Bias | 0.058 | 0.049 | 0.053 | 0.028 | 0.027 | 0.020 | 0.012 |
| 6 | $\alpha_1$ | Stdev | 0.268 | 0.206 | 0.156 | 0.122 | 0.089 | 0.061 | 0.047 |
| 6 | $\phi_1$ | Bias | 0.021 | 0.032 | 0.036 | 0.026 | 0.025 | 0.017 | 0.011 |
| 6 | $\phi_1$ | Stdev | 0.250 | 0.195 | 0.143 | 0.108 | 0.075 | 0.053 | 0.040 |
| 6 | $\alpha_2$ | Bias | 0.016 | 0.005 | −0.000 | −0.005 | −0.003 | −0.000 | −0.000 |
| 6 | $\alpha_2$ | Stdev | 0.341 | 0.217 | 0.157 | 0.096 | 0.071 | 0.050 | 0.034 |
| 6 | $\phi_2$ | Bias | −0.023 | −0.012 | −0.003 | −0.004 | −0.001 | −0.003 | −0.001 |
| 6 | $\phi_2$ | Stdev | 0.184 | 0.146 | 0.110 | 0.077 | 0.058 | 0.042 | 0.029 |
| 6 | $d$ | Bias | 0.600 | 0.690 | 0.779 | 0.824 | 0.869 | 0.887 | 0.829 |
| 6 | $d$ | Stdev | 0.837 | 0.858 | 0.870 | 0.878 | 0.889 | 0.873 | 0.868 |
| 6 | $c$ | Bias | 0.037 | 0.049 | 0.043 | 0.040 | 0.062 | 0.061 | 0.058 |
| 6 | $c$ | Stdev | 0.209 | 0.237 | 0.238 | 0.240 | 0.248 | 0.252 | 0.256 |

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 18 | $\alpha_1$ | Bias | 0.025 | 0.043 | 0.023 | 0.035 | 0.024 | 0.019 | 0.014 |
| 18 | $\alpha_1$ | Stdev | 0.328 | 0.262 | 0.199 | 0.135 | 0.105 | 0.075 | 0.050 |
| 18 | $\phi_1$ | Bias | 0.006 | 0.034 | 0.021 | 0.029 | 0.021 | 0.015 | 0.011 |
| 18 | $\phi_1$ | Stdev | 0.313 | 0.229 | 0.173 | 0.115 | 0.089 | 0.060 | 0.041 |
| 18 | $\alpha_2$ | Bias | 0.059 | 0.026 | 0.017 | 0.009 | −0.001 | 0.001 | 0.003 |
| 18 | $\alpha_2$ | Stdev | 0.417 | 0.285 | 0.178 | 0.128 | 0.081 | 0.062 | 0.042 |
| 18 | $\phi_2$ | Bias | −0.050 | −0.026 | −0.013 | −0.012 | −0.003 | −0.002 | −0.003 |
| 18 | $\phi_2$ | Stdev | 0.224 | 0.171 | 0.119 | 0.090 | 0.063 | 0.046 | 0.033 |
| 18 | $d$ | Bias | 0.717 | 0.769 | 0.807 | 0.869 | 0.920 | 0.905 | 0.922 |
| 18 | $d$ | Stdev | 0.867 | 0.860 | 0.885 | 0.876 | 0.882 | 0.873 | 0.890 |
| 18 | $c$ | Bias | 0.011 | 0.007 | −0.011 | 0.012 | −0.002 | −0.001 | 0.013 |
| 18 | $c$ | Stdev | 0.200 | 0.210 | 0.213 | 0.227 | 0.232 | 0.225 | 0.228 |

DGP: $y_t = \alpha_{10}+\phi_{10}y_{t-1}+\epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n}+\phi_{20n}y_{t-1}+\epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10}=0$, $\phi_{10}=0.2$, $\alpha_{20n}=\alpha_{10}+\lambda_0 n^{-\delta_0}$, $\phi_{20n}=\phi_{10}+\lambda_0 n^{-\delta_0}$, $d_0=1$, $c_0=0.5$, $m \in \{6,18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #3: $\delta_0=0.5$, hence $\boldsymbol{\gamma}_0=(d_0,c_0)^\top$ is weakly identified (boundary case). Sample size: $n \in \{125,\ldots,8000\}$. Model: $y_t=\alpha_1+\phi_1 y_{t-1}+u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t=\alpha_2+\phi_2 y_{t-1}+u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. The profiling estimator for $(\alpha_1,\phi_1,\alpha_2,\phi_2,d,c)$ is computed. For each parameter, bias and standard deviation across $J=1000$ Monte Carlo samples are reported.

Table 5: Simulation results on the profiling estimation (Case #4: semi-strong identification)

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 6 | $\alpha_1$ | Bias | 0.046 | 0.026 | 0.023 | 0.014 | 0.002 | −0.000 | −0.001 |
| 6 | $\alpha_1$ | Stdev | 0.273 | 0.201 | 0.153 | 0.099 | 0.060 | 0.039 | 0.025 |
| 6 | $\phi_1$ | Bias | 0.023 | 0.019 | 0.019 | 0.007 | −0.001 | 0.001 | −0.001 |
| 6 | $\phi_1$ | Stdev | 0.241 | 0.188 | 0.137 | 0.093 | 0.056 | 0.036 | 0.024 |
| 6 | $\alpha_2$ | Bias | 0.020 | −0.000 | 0.003 | 0.001 | 0.005 | −0.001 | −0.000 |
| 6 | $\alpha_2$ | Stdev | 0.340 | 0.201 | 0.137 | 0.082 | 0.052 | 0.033 | 0.022 |
| 6 | $\phi_2$ | Bias | −0.027 | −0.010 | −0.005 | −0.001 | −0.002 | 0.000 | −0.000 |
| 6 | $\phi_2$ | Stdev | 0.185 | 0.125 | 0.092 | 0.060 | 0.042 | 0.029 | 0.019 |
| 6 | $d$ | Bias | 0.578 | 0.475 | 0.375 | 0.237 | 0.087 | 0.026 | 0.004 |
| 6 | $d$ | Stdev | 0.837 | 0.779 | 0.723 | 0.601 | 0.387 | 0.203 | 0.077 |
| 6 | $c$ | Bias | 0.047 | 0.012 | 0.026 | 0.012 | 0.004 | 0.002 | −0.001 |
| 6 | $c$ | Stdev | 0.218 | 0.192 | 0.183 | 0.145 | 0.097 | 0.069 | 0.049 |

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n=125$ | $n=250$ | $n=500$ | $n=1000$ | $n=2000$ | $n=4000$ | $n=8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 18 | $\alpha_1$ | Bias | 0.035 | 0.033 | 0.015 | 0.004 | −0.002 | −0.001 | −0.003 |
| 18 | $\alpha_1$ | Stdev | 0.330 | 0.246 | 0.172 | 0.122 | 0.078 | 0.045 | 0.029 |
| 18 | $\phi_1$ | Bias | 0.015 | 0.016 | 0.006 | −0.001 | −0.004 | −0.002 | −0.002 |
| 18 | $\phi_1$ | Stdev | 0.311 | 0.228 | 0.152 | 0.106 | 0.070 | 0.043 | 0.028 |
| 18 | $\alpha_2$ | Bias | 0.073 | 0.010 | 0.010 | 0.003 | −0.001 | 0.003 | 0.001 |
| 18 | $\alpha_2$ | Stdev | 0.440 | 0.260 | 0.166 | 0.101 | 0.063 | 0.043 | 0.027 |
| 18 | $\phi_2$ | Bias | −0.052 | −0.015 | −0.010 | −0.004 | 0.001 | −0.001 | −0.001 |
| 18 | $\phi_2$ | Stdev | 0.232 | 0.150 | 0.105 | 0.070 | 0.050 | 0.035 | 0.023 |
| 18 | $d$ | Bias | 0.698 | 0.595 | 0.397 | 0.259 | 0.117 | 0.034 | 0.009 |
| 18 | $d$ | Stdev | 0.866 | 0.830 | 0.723 | 0.610 | 0.437 | 0.225 | 0.130 |
| 18 | $c$ | Bias | 0.010 | −0.007 | −0.007 | −0.014 | −0.010 | −0.002 | −0.006 |
| 18 | $c$ | Stdev | 0.203 | 0.191 | 0.168 | 0.143 | 0.117 | 0.080 | 0.049 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #4: $\delta_0 = 0.25$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is semi-strongly identified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. The profiling estimator for $(\alpha_1, \phi_1, \alpha_2, \phi_2, d, c)$ is computed. For each parameter, bias and standard deviation across $J = 1000$ Monte Carlo samples are reported.

Table 6: Simulation results on the profiling estimation (Case #5: strong identification)

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 6 | $\alpha_1$ | Bias | 0.041 | 0.012 | 0.000 | −0.001 | 0.001 | −0.000 | 0.000 |
| 6 | $\alpha_1$ | Stdev | 0.268 | 0.153 | 0.086 | 0.053 | 0.037 | 0.025 | 0.018 |
| 6 | $\phi_1$ | Bias | 0.019 | −0.004 | −0.009 | −0.002 | −0.001 | −0.001 | 0.001 |
| 6 | $\phi_1$ | Stdev | 0.259 | 0.147 | 0.082 | 0.052 | 0.038 | 0.027 | 0.019 |
| 6 | $\alpha_2$ | Bias | 0.039 | 0.019 | 0.007 | 0.006 | 0.006 | 0.001 | 0.000 |
| 6 | $\alpha_2$ | Stdev | 0.346 | 0.193 | 0.105 | 0.070 | 0.050 | 0.034 | 0.025 |
| 6 | $\phi_2$ | Bias | −0.039 | −0.014 | −0.006 | −0.004 | −0.004 | −0.001 | −0.001 |
| 6 | $\phi_2$ | Stdev | 0.194 | 0.110 | 0.065 | 0.045 | 0.031 | 0.022 | 0.016 |
| 6 | $d$ | Bias | 0.540 | 0.201 | 0.021 | 0.000 | 0.000 | 0.000 | 0.000 |
| 6 | $d$ | Stdev | 0.804 | 0.561 | 0.180 | 0.000 | 0.000 | 0.000 | 0.000 |
| 6 | $c$ | Bias | 0.043 | 0.010 | 0.001 | −0.000 | 0.000 | 0.000 | 0.000 |
| 6 | $c$ | Stdev | 0.217 | 0.139 | 0.061 | 0.007 | 0.000 | 0.000 | 0.000 |

| $m$ | $\boldsymbol{\theta}$ | Quantity | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|---|
| 18 | $\alpha_1$ | Bias | 0.052 | 0.008 | −0.012 | −0.003 | 0.000 | 0.000 | 0.001 |
| 18 | $\alpha_1$ | Stdev | 0.347 | 0.204 | 0.102 | 0.053 | 0.036 | 0.025 | 0.019 |
| 18 | $\phi_1$ | Bias | 0.028 | 0.000 | −0.018 | −0.006 | −0.000 | 0.000 | −0.000 |
| 18 | $\phi_1$ | Stdev | 0.311 | 0.191 | 0.103 | 0.060 | 0.041 | 0.029 | 0.021 |
| 18 | $\alpha_2$ | Bias | 0.059 | 0.048 | 0.020 | 0.012 | 0.002 | 0.002 | 0.001 |
| 18 | $\alpha_2$ | Stdev | 0.456 | 0.248 | 0.145 | 0.089 | 0.062 | 0.043 | 0.031 |
| 18 | $\phi_2$ | Bias | −0.050 | −0.026 | −0.012 | −0.009 | −0.002 | −0.002 | −0.001 |
| 18 | $\phi_2$ | Stdev | 0.233 | 0.133 | 0.079 | 0.054 | 0.037 | 0.025 | 0.018 |
| 18 | $d$ | Bias | 0.709 | 0.324 | 0.047 | 0.000 | 0.000 | 0.000 | 0.000 |
| 18 | $d$ | Stdev | 0.872 | 0.675 | 0.281 | 0.000 | 0.000 | 0.000 | 0.000 |
| 18 | $c$ | Bias | 0.010 | 0.001 | −0.005 | −0.001 | −0.000 | 0.000 | 0.000 |
| 18 | $c$ | Stdev | 0.197 | 0.147 | 0.075 | 0.021 | 0.006 | 0.002 | 0.000 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #5: $\delta_0 = 0$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is strongly identified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. The profiling estimator for $(\alpha_1, \phi_1, \alpha_2, \phi_2, d, c)$ is computed. For each parameter, bias and standard deviation across $J = 1000$ Monte Carlo samples are reported.

Table 7: Rejection frequencies of the bootstrap tests for the no-threshold-effect hypothesis $H_0^*$ (Case #1: non-identification)

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 6 | sup-Wald | 0.197 | 0.100 | 0.070 | 0.068 | 0.048 | 0.045 | 0.050 |
| 6 | ave-Wald | 0.120 | 0.085 | 0.064 | 0.058 | 0.055 | 0.052 | 0.053 |
| 6 | exp-Wald | 0.191 | 0.099 | 0.071 | 0.066 | 0.050 | 0.052 | 0.048 |
| 6 | sup-LM | 0.045 | 0.026 | 0.041 | 0.054 | 0.041 | 0.044 | 0.048 |
| 6 | ave-LM | 0.040 | 0.044 | 0.046 | 0.049 | 0.051 | 0.047 | 0.050 |
| 6 | exp-LM | 0.046 | 0.028 | 0.047 | 0.052 | 0.044 | 0.047 | 0.046 |

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 18 | sup-Wald | 0.207 | 0.124 | 0.084 | 0.058 | 0.062 | 0.069 | 0.056 |
| 18 | ave-Wald | 0.114 | 0.073 | 0.059 | 0.056 | 0.064 | 0.059 | 0.055 |
| 18 | exp-Wald | 0.200 | 0.109 | 0.084 | 0.050 | 0.062 | 0.064 | 0.054 |
| 18 | sup-LM | 0.030 | 0.029 | 0.047 | 0.044 | 0.050 | 0.062 | 0.056 |
| 18 | ave-LM | 0.040 | 0.040 | 0.045 | 0.048 | 0.060 | 0.057 | 0.053 |
| 18 | exp-LM | 0.029 | 0.031 | 0.043 | 0.039 | 0.056 | 0.058 | 0.051 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #1: $\delta_0 \to \infty$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is unidentified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. We report rejection frequencies of the wild-bootstrap tests for the no-threshold-effect hypothesis $H_0^* : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $\alpha = 0.05$; the number of bootstrap samples is $B = 500$; the number of Monte Carlo samples is $J = 1000$.

Table 8: Rejection frequencies of the bootstrap tests for the no-threshold-effect hypothesis $H_0^*$ (Case #2: weak identification)

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 6 | sup-Wald | 0.573 | 0.269 | 0.153 | 0.094 | 0.071 | 0.061 | 0.056 |
| 6 | ave-Wald | 0.496 | 0.245 | 0.147 | 0.111 | 0.064 | 0.073 | 0.059 |
| 6 | exp-Wald | 0.581 | 0.275 | 0.160 | 0.107 | 0.070 | 0.072 | 0.060 |
| 6 | sup-LM | 0.212 | 0.130 | 0.100 | 0.076 | 0.054 | 0.060 | 0.054 |
| 6 | ave-LM | 0.255 | 0.152 | 0.108 | 0.098 | 0.060 | 0.072 | 0.056 |
| 6 | exp-LM | 0.235 | 0.133 | 0.104 | 0.089 | 0.060 | 0.069 | 0.057 |

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 18 | sup-Wald | 0.521 | 0.273 | 0.137 | 0.108 | 0.067 | 0.084 | 0.067 |
| 18 | ave-Wald | 0.392 | 0.243 | 0.147 | 0.119 | 0.087 | 0.084 | 0.080 |
| 18 | exp-Wald | 0.500 | 0.275 | 0.132 | 0.121 | 0.073 | 0.080 | 0.067 |
| 18 | sup-LM | 0.171 | 0.124 | 0.090 | 0.081 | 0.054 | 0.079 | 0.067 |
| 18 | ave-LM | 0.250 | 0.170 | 0.126 | 0.107 | 0.086 | 0.083 | 0.078 |
| 18 | exp-LM | 0.199 | 0.136 | 0.092 | 0.090 | 0.064 | 0.077 | 0.066 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #2: $\delta_0 = 0.75$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is weakly identified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. We report rejection frequencies of the wild-bootstrap tests for the no-threshold-effect hypothesis $H_0^* : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $\alpha = 0.05$; the number of bootstrap samples is $B = 500$; the number of Monte Carlo samples is $J = 1000$.

Table 9: Rejection frequencies of the bootstrap tests for the no-threshold-effect hypothesis $H_0^*$ (Case #3: weak identification, boundary)

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 6 | sup-Wald | 0.584 | 0.389 | 0.283 | 0.235 | 0.206 | 0.202 | 0.171 |
| 6 | ave-Wald | 0.513 | 0.364 | 0.278 | 0.253 | 0.200 | 0.203 | 0.186 |
| 6 | exp-Wald | 0.593 | 0.408 | 0.301 | 0.239 | 0.220 | 0.207 | 0.184 |
| 6 | sup-LM | 0.228 | 0.214 | 0.207 | 0.192 | 0.193 | 0.190 | 0.171 |
| 6 | ave-LM | 0.261 | 0.260 | 0.230 | 0.230 | 0.187 | 0.196 | 0.186 |
| 6 | exp-LM | 0.249 | 0.234 | 0.218 | 0.205 | 0.204 | 0.203 | 0.180 |

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 18 | sup-Wald | 0.515 | 0.344 | 0.252 | 0.202 | 0.192 | 0.174 | 0.168 |
| 18 | ave-Wald | 0.403 | 0.303 | 0.231 | 0.237 | 0.229 | 0.215 | 0.193 |
| 18 | exp-Wald | 0.509 | 0.342 | 0.254 | 0.223 | 0.213 | 0.190 | 0.172 |
| 18 | sup-LM | 0.162 | 0.170 | 0.168 | 0.169 | 0.174 | 0.163 | 0.165 |
| 18 | ave-LM | 0.234 | 0.221 | 0.188 | 0.224 | 0.223 | 0.210 | 0.191 |
| 18 | exp-LM | 0.183 | 0.191 | 0.183 | 0.190 | 0.201 | 0.183 | 0.169 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #3: $\delta_0 = 0.5$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is weakly identified (boundary case). Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. We report rejection frequencies of the wild-bootstrap tests for the no-threshold-effect hypothesis $H_0^* : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $\alpha = 0.05$; the number of bootstrap samples is $B = 500$; the number of Monte Carlo samples is $J = 1000$.

Table 10: Rejection frequencies of the bootstrap tests for the no-threshold-effect hypothesis $H_0^*$ (Case #4: semi-strong identification)

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 6 | sup-Wald | 0.551 | 0.561 | 0.615 | 0.722 | 0.866 | 0.957 | 0.992 |
| 6 | ave-Wald | 0.493 | 0.521 | 0.547 | 0.643 | 0.795 | 0.898 | 0.981 |
| 6 | exp-Wald | 0.559 | 0.570 | 0.633 | 0.731 | 0.871 | 0.960 | 0.994 |
| 6 | sup-LM | 0.192 | 0.371 | 0.545 | 0.685 | 0.851 | 0.953 | 0.992 |
| 6 | ave-LM | 0.235 | 0.390 | 0.478 | 0.619 | 0.782 | 0.894 | 0.978 |
| 6 | exp-LM | 0.202 | 0.399 | 0.552 | 0.701 | 0.862 | 0.956 | 0.993 |

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 18 | sup-Wald | 0.497 | 0.508 | 0.546 | 0.670 | 0.829 | 0.931 | 0.987 |
| 18 | ave-Wald | 0.411 | 0.453 | 0.506 | 0.648 | 0.799 | 0.891 | 0.960 |
| 18 | exp-Wald | 0.491 | 0.507 | 0.557 | 0.687 | 0.837 | 0.936 | 0.986 |
| 18 | sup-LM | 0.147 | 0.312 | 0.460 | 0.629 | 0.819 | 0.927 | 0.987 |
| 18 | ave-LM | 0.232 | 0.348 | 0.449 | 0.611 | 0.788 | 0.887 | 0.959 |
| 18 | exp-LM | 0.183 | 0.345 | 0.479 | 0.655 | 0.827 | 0.933 | 0.986 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #4: $\delta_0 = 0.25$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is semi-strongly identified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. We report rejection frequencies of the wild-bootstrap tests for the no-threshold-effect hypothesis $H_0^* : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $\alpha = 0.05$; the number of bootstrap samples is $B = 500$; the number of Monte Carlo samples is $J = 1000$.

Table 11: Rejection frequencies of the bootstrap tests for the no-threshold-effect hypothesis $H_0^*$ (Case #5: strong identification)

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 6 | sup-Wald | 0.550 | 0.805 | 0.986 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6 | ave-Wald | 0.484 | 0.730 | 0.949 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6 | exp-Wald | 0.548 | 0.811 | 0.988 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6 | sup-LM | 0.208 | 0.630 | 0.968 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6 | ave-LM | 0.245 | 0.558 | 0.922 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6 | exp-LM | 0.224 | 0.648 | 0.973 | 1.000 | 1.000 | 1.000 | 1.000 |

| $m$ | test | $n = 125$ | $n = 250$ | $n = 500$ | $n = 1000$ | $n = 2000$ | $n = 4000$ | $n = 8000$ |
|---|---|---|---|---|---|---|---|---|
| 18 | sup-Wald | 0.499 | 0.713 | 0.948 | 1.000 | 1.000 | 1.000 | 1.000 |
| 18 | ave-Wald | 0.421 | 0.655 | 0.901 | 0.999 | 1.000 | 1.000 | 1.000 |
| 18 | exp-Wald | 0.489 | 0.718 | 0.953 | 0.999 | 1.000 | 1.000 | 1.000 |
| 18 | sup-LM | 0.145 | 0.505 | 0.925 | 0.998 | 1.000 | 1.000 | 1.000 |
| 18 | ave-LM | 0.226 | 0.536 | 0.878 | 0.999 | 1.000 | 1.000 | 1.000 |
| 18 | exp-LM | 0.187 | 0.522 | 0.922 | 0.999 | 1.000 | 1.000 | 1.000 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20n} + \phi_{20n} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20n} = \alpha_{10} + \lambda_0 n^{-\delta_0}$, $\phi_{20n} = \phi_{10} + \lambda_0 n^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Case #5: $\delta_0 = 0$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is strongly identified. Sample size: $n \in \{125, \ldots, 8000\}$. Model: $y_t = \alpha_1 + \phi_1 y_{t-1} + u_t$ if $y_{t-d} < \mu_{t-d-1}(c)$ and $y_t = \alpha_2 + \phi_2 y_{t-1} + u_t$ if $y_{t-d} \geq \mu_{t-d-1}(c)$. We report rejection frequencies of the wild-bootstrap tests for the no-threshold-effect hypothesis $H_0^* : (\alpha_1, \phi_1) = (\alpha_2, \phi_2)$, where the nominal size is $\alpha = 0.05$; the number of bootstrap samples is $B = 500$; the number of Monte Carlo samples is $J = 1000$.

Table 12: Rejection frequencies of the Diebold-Mariano test for the equal predictive accuracy hypothesis $H_0^{eq}$ under Scenario 1 (DGP: SE-CoTAR, models: SE-CoTAR vs. AR)

| | | Memory size: $m = 6$ | | | | | Memory size: $m = 18$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $H_1$ | $N$ | #1 | #2 | #3 | #4 | #5 | #1 | #2 | #3 | #4 | #5 |
| $H_1^{eq}$ | 125 | 0.108 | 0.074 | 0.077 | 0.079 | 0.057 | 0.103 | 0.075 | 0.096 | 0.078 | 0.084 |
| $H_1^{eq}$ | 250 | 0.082 | 0.072 | 0.059 | 0.059 | 0.063 | 0.109 | 0.091 | 0.077 | 0.071 | 0.063 |
| $H_1^{eq}$ | 500 | 0.103 | 0.062 | 0.072 | 0.068 | 0.134 | 0.102 | 0.085 | 0.081 | 0.060 | 0.098 |
| $H_1^{eq}$ | 1000 | 0.108 | 0.075 | 0.067 | 0.064 | 0.329 | 0.091 | 0.093 | 0.081 | 0.074 | 0.266 |
| $H_1^{eq}$ | 2000 | 0.083 | 0.101 | 0.074 | 0.088 | 0.595 | 0.092 | 0.085 | 0.074 | 0.082 | 0.562 |
| $H_1^{eq}$ | 4000 | 0.105 | 0.081 | 0.076 | 0.115 | 0.925 | 0.110 | 0.100 | 0.088 | 0.088 | 0.863 |
| $H_1^{secotar}$ | 125 | 0.009 | 0.037 | 0.031 | 0.034 | 0.037 | 0.009 | 0.026 | 0.028 | 0.026 | 0.019 |
| $H_1^{secotar}$ | 250 | 0.011 | 0.010 | 0.023 | 0.034 | 0.088 | 0.008 | 0.019 | 0.018 | 0.033 | 0.059 |
| $H_1^{secotar}$ | 500 | 0.010 | 0.011 | 0.018 | 0.050 | 0.206 | 0.009 | 0.013 | 0.023 | 0.042 | 0.181 |
| $H_1^{secotar}$ | 1000 | 0.012 | 0.009 | 0.018 | 0.096 | 0.474 | 0.008 | 0.009 | 0.021 | 0.076 | 0.393 |
| $H_1^{secotar}$ | 2000 | 0.011 | 0.011 | 0.016 | 0.125 | 0.713 | 0.005 | 0.009 | 0.020 | 0.103 | 0.682 |
| $H_1^{secotar}$ | 4000 | 0.007 | 0.005 | 0.018 | 0.177 | 0.958 | 0.004 | 0.007 | 0.016 | 0.138 | 0.912 |
| $H_1^{ar}$ | 125 | 0.175 | 0.102 | 0.103 | 0.113 | 0.085 | 0.172 | 0.124 | 0.134 | 0.131 | 0.145 |
| $H_1^{ar}$ | 250 | 0.146 | 0.110 | 0.099 | 0.078 | 0.040 | 0.178 | 0.146 | 0.121 | 0.096 | 0.052 |
| $H_1^{ar}$ | 500 | 0.166 | 0.122 | 0.121 | 0.061 | 0.010 | 0.167 | 0.138 | 0.116 | 0.081 | 0.011 |
| $H_1^{ar}$ | 1000 | 0.148 | 0.137 | 0.103 | 0.033 | 0.000 | 0.164 | 0.165 | 0.134 | 0.062 | 0.000 |
| $H_1^{ar}$ | 2000 | 0.136 | 0.154 | 0.115 | 0.028 | 0.000 | 0.163 | 0.145 | 0.135 | 0.032 | 0.000 |
| $H_1^{ar}$ | 4000 | 0.159 | 0.158 | 0.131 | 0.011 | 0.000 | 0.170 | 0.156 | 0.136 | 0.018 | 0.000 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20N} + \phi_{20N} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20N} = \alpha_{10} + \lambda_0 N^{-\delta_0}$, $\phi_{20N} = \phi_{10} + \lambda_0 N^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Sample size: $N \in \{125, \ldots, 4000\}$. Case #1: $\delta_0 \to \infty$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is unidentified. Case #2: $\delta_0 = 0.75$, hence $\boldsymbol{\gamma}_0$ is weakly identified. Case #3: $\delta_0 = 0.5$, hence $\boldsymbol{\gamma}_0$ is weakly identified (boundary case). Case #4: $\delta_0 = 0.25$, hence $\boldsymbol{\gamma}_0$ is semi-strongly identified. Case #5: $\delta_0 = 0$, hence $\boldsymbol{\gamma}_0$ is strongly identified. The rolling window out-of-sample prediction is performed, where the window size is fixed at $n = 0.8N$. The one-step-ahead forecasts based on the SE-CoTAR and AR models are compared by the asymptotic Diebold-Mariano test. $H_0^{eq}$: $MSE^{ar} = MSE^{secotar}$. $H_1^{eq}$: $MSE^{ar} \neq MSE^{secotar}$. $H_1^{secotar}$: $MSE^{ar} > MSE^{secotar}$. $H_1^{ar}$: $MSE^{ar} < MSE^{secotar}$. Rejection frequencies across $J = 1000$ Monte Carlo samples are reported, where the nominal size is $a = 0.05$.

Table 13: Rejection frequencies of the Diebold-Mariano test for the equal predictive accuracy hypothesis $H_0^{eq}$ under Scenario 2 (DGP: SE-CoTAR, models: SE-CoTAR vs. SETAR)

| | | Memory size: $m = 6$ | | | | | Memory size: $m = 18$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $H_1$ | $N$ | #1 | #2 | #3 | #4 | #5 | #1 | #2 | #3 | #4 | #5 |
| $H_1^{eq}$ | 125 | 0.056 | 0.049 | 0.069 | 0.058 | 0.079 | 0.053 | 0.058 | 0.065 | 0.055 | 0.042 |
| $H_1^{eq}$ | 250 | 0.041 | 0.043 | 0.054 | 0.060 | 0.088 | 0.037 | 0.038 | 0.051 | 0.046 | 0.067 |
| $H_1^{eq}$ | 500 | 0.040 | 0.048 | 0.056 | 0.065 | 0.160 | 0.047 | 0.042 | 0.047 | 0.050 | 0.113 |
| $H_1^{eq}$ | 1000 | 0.041 | 0.045 | 0.058 | 0.075 | 0.344 | 0.040 | 0.038 | 0.046 | 0.066 | 0.203 |
| $H_1^{eq}$ | 2000 | 0.037 | 0.049 | 0.044 | 0.106 | 0.535 | 0.042 | 0.044 | 0.045 | 0.081 | 0.398 |
| $H_1^{eq}$ | 4000 | 0.048 | 0.042 | 0.047 | 0.140 | 0.811 | 0.041 | 0.042 | 0.051 | 0.106 | 0.666 |
| $H_1^{secotar}$ | 125 | 0.065 | 0.089 | 0.100 | 0.084 | 0.103 | 0.056 | 0.058 | 0.062 | 0.069 | 0.055 |
| $H_1^{secotar}$ | 250 | 0.056 | 0.066 | 0.085 | 0.099 | 0.157 | 0.044 | 0.061 | 0.064 | 0.063 | 0.111 |
| $H_1^{secotar}$ | 500 | 0.075 | 0.063 | 0.086 | 0.104 | 0.254 | 0.054 | 0.050 | 0.066 | 0.078 | 0.171 |
| $H_1^{secotar}$ | 1000 | 0.062 | 0.070 | 0.078 | 0.142 | 0.469 | 0.070 | 0.047 | 0.056 | 0.097 | 0.333 |
| $H_1^{secotar}$ | 2000 | 0.058 | 0.066 | 0.073 | 0.177 | 0.663 | 0.075 | 0.054 | 0.072 | 0.129 | 0.514 |
| $H_1^{secotar}$ | 4000 | 0.069 | 0.070 | 0.076 | 0.225 | 0.892 | 0.051 | 0.054 | 0.064 | 0.174 | 0.758 |
| $H_1^{setar}$ | 125 | 0.038 | 0.022 | 0.036 | 0.026 | 0.029 | 0.063 | 0.060 | 0.063 | 0.046 | 0.032 |
| $H_1^{setar}$ | 250 | 0.028 | 0.022 | 0.029 | 0.014 | 0.012 | 0.050 | 0.031 | 0.044 | 0.034 | 0.027 |
| $H_1^{setar}$ | 500 | 0.023 | 0.033 | 0.027 | 0.014 | 0.003 | 0.035 | 0.035 | 0.030 | 0.032 | 0.010 |
| $H_1^{setar}$ | 1000 | 0.030 | 0.033 | 0.033 | 0.013 | 0.000 | 0.024 | 0.056 | 0.029 | 0.020 | 0.002 |
| $H_1^{setar}$ | 2000 | 0.025 | 0.032 | 0.018 | 0.006 | 0.000 | 0.027 | 0.035 | 0.040 | 0.014 | 0.000 |
| $H_1^{setar}$ | 4000 | 0.024 | 0.025 | 0.024 | 0.003 | 0.000 | 0.035 | 0.026 | 0.032 | 0.007 | 0.001 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_{t-d_0-1}(c_0)$ and $y_t = \alpha_{20N} + \phi_{20N} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20N} = \alpha_{10} + \lambda_0 N^{-\delta_0}$, $\phi_{20N} = \phi_{10} + \lambda_0 N^{-\delta_0}$, $d_0 = 1$, $c_0 = 0.5$, $m \in \{6, 18\}$, and $\mu_t(c_0)$ is the $mc_0$-th smallest value of $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Sample size: $N \in \{125, \ldots, 4000\}$. Case #1: $\delta_0 \to \infty$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is unidentified. Case #2: $\delta_0 = 0.75$, hence $\boldsymbol{\gamma}_0$ is weakly identified. Case #3: $\delta_0 = 0.5$, hence $\boldsymbol{\gamma}_0$ is weakly identified (boundary case). Case #4: $\delta_0 = 0.25$, hence $\boldsymbol{\gamma}_0$ is semi-strongly identified. Case #5: $\delta_0 = 0$, hence $\boldsymbol{\gamma}_0$ is strongly identified. The rolling window out-of-sample prediction is performed, where the window size is fixed at $n = 0.8N$. The one-step-ahead forecasts based on the SE-CoTAR and SETAR models are compared by the asymptotic Diebold-Mariano test. $H_0^{eq}$: $MSE^{setar} = MSE^{secotar}$. $H_1^{eq}$: $MSE^{setar} \neq MSE^{secotar}$. $H_1^{secotar}$: $MSE^{setar} > MSE^{secotar}$. $H_1^{setar}$: $MSE^{setar} < MSE^{secotar}$. Rejection frequencies across $J = 1000$ Monte Carlo samples are reported, where the nominal size is $a = 0.05$.

Table 14: Rejection frequencies of the Diebold-Mariano test for the equal predictive accuracy hypothesis $H_0^{eq}$ under Scenario 3 (DGP: SETAR, models: SE-CoTAR vs. SETAR)

| | | Memory size: $m = 6$ | | | | | Memory size: $m = 18$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $H_1$ | $N$ | #1 | #2 | #3 | #4 | #5 | #1 | #2 | #3 | #4 | #5 |
| $H_1^{eq}$ | 125 | 0.055 | 0.070 | 0.057 | 0.044 | 0.047 | 0.052 | 0.069 | 0.047 | 0.048 | 0.047 |
| $H_1^{eq}$ | 250 | 0.052 | 0.043 | 0.050 | 0.045 | 0.048 | 0.037 | 0.042 | 0.045 | 0.059 | 0.051 |
| $H_1^{eq}$ | 500 | 0.046 | 0.038 | 0.038 | 0.059 | 0.071 | 0.038 | 0.038 | 0.042 | 0.052 | 0.049 |
| $H_1^{eq}$ | 1000 | 0.046 | 0.038 | 0.053 | 0.044 | 0.072 | 0.040 | 0.054 | 0.041 | 0.047 | 0.056 |
| $H_1^{eq}$ | 2000 | 0.049 | 0.049 | 0.041 | 0.042 | 0.180 | 0.038 | 0.040 | 0.049 | 0.036 | 0.129 |
| $H_1^{eq}$ | 4000 | 0.046 | 0.045 | 0.040 | 0.065 | 0.279 | 0.048 | 0.045 | 0.036 | 0.047 | 0.218 |
| $H_1^{secotar}$ | 125 | 0.063 | 0.073 | 0.051 | 0.070 | 0.063 | 0.054 | 0.049 | 0.047 | 0.052 | 0.045 |
| $H_1^{secotar}$ | 250 | 0.073 | 0.048 | 0.063 | 0.053 | 0.053 | 0.053 | 0.050 | 0.056 | 0.054 | 0.047 |
| $H_1^{secotar}$ | 500 | 0.068 | 0.046 | 0.054 | 0.062 | 0.046 | 0.048 | 0.052 | 0.053 | 0.038 | 0.031 |
| $H_1^{secotar}$ | 1000 | 0.056 | 0.059 | 0.064 | 0.042 | 0.011 | 0.071 | 0.058 | 0.056 | 0.050 | 0.021 |
| $H_1^{secotar}$ | 2000 | 0.059 | 0.073 | 0.055 | 0.028 | 0.006 | 0.054 | 0.060 | 0.063 | 0.030 | 0.007 |
| $H_1^{secotar}$ | 4000 | 0.067 | 0.065 | 0.062 | 0.026 | 0.001 | 0.067 | 0.062 | 0.054 | 0.036 | 0.001 |
| $H_1^{setar}$ | 125 | 0.039 | 0.062 | 0.050 | 0.038 | 0.044 | 0.065 | 0.068 | 0.062 | 0.064 | 0.043 |
| $H_1^{setar}$ | 250 | 0.034 | 0.037 | 0.043 | 0.035 | 0.043 | 0.040 | 0.052 | 0.047 | 0.055 | 0.049 |
| $H_1^{setar}$ | 500 | 0.029 | 0.028 | 0.037 | 0.040 | 0.075 | 0.037 | 0.031 | 0.043 | 0.060 | 0.062 |
| $H_1^{setar}$ | 1000 | 0.022 | 0.025 | 0.039 | 0.051 | 0.122 | 0.033 | 0.043 | 0.038 | 0.058 | 0.109 |
| $H_1^{setar}$ | 2000 | 0.031 | 0.029 | 0.034 | 0.067 | 0.260 | 0.028 | 0.028 | 0.035 | 0.048 | 0.210 |
| $H_1^{setar}$ | 4000 | 0.026 | 0.029 | 0.026 | 0.091 | 0.413 | 0.037 | 0.024 | 0.029 | 0.083 | 0.313 |

DGP: $y_t = \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t$ if $y_{t-d_0} < \mu_0$ and $y_t = \alpha_{20N} + \phi_{20N} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_0$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $\alpha_{20N} = \alpha_{10} + \lambda_0 N^{-\delta_0}$, $\phi_{20N} = \phi_{10} + \lambda_0 N^{-\delta_0}$, $d_0 = 1$, and $\mu_0 = 0.1$. Sample size: $N \in \{125, \ldots, 4000\}$. Case #1: $\delta_0 \to \infty$, hence $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ is unidentified. Case #2: $\delta_0 = 0.75$, hence $\boldsymbol{\gamma}_0$ is weakly identified. Case #3: $\delta_0 = 0.5$, hence $\boldsymbol{\gamma}_0$ is weakly identified (boundary case). Case #4: $\delta_0 = 0.25$, hence $\boldsymbol{\gamma}_0$ is semi-strongly identified. Case #5: $\delta_0 = 0$, hence $\boldsymbol{\gamma}_0$ is strongly identified. The rolling window out-of-sample prediction is performed, where the window size is fixed at $n = 0.8N$. The one-step-ahead forecasts based on the SE-CoTAR and SETAR models are compared by the asymptotic Diebold-Mariano test. $H_0^{eq}$: $MSE^{setar} = MSE^{secotar}$. $H_1^{eq}$: $MSE^{setar} \neq MSE^{secotar}$. $H_1^{secotar}$: $MSE^{setar} > MSE^{secotar}$. $H_1^{setar}$: $MSE^{setar} < MSE^{secotar}$. Rejection frequencies across $J = 1000$ Monte Carlo samples are reported, where the nominal size is $a = 0.05$.

Table 15: Sample statistics of the log-difference of daily new confirmed COVID-19 cases

| country | $n$ | mean | median | min | max | stdev | skew | kurt |
|---|---|---|---|---|---|---|---|---|
| Japan | 446 | 0.005 | 0.009 | −0.308 | 0.265 | 0.059 | −0.239 | 6.326 |
| U.S. | 446 | −0.001 | −0.003 | −0.126 | 0.124 | 0.028 | 0.124 | 4.027 |

In this table, we report summary statistics of the log-difference of daily new confirmed COVID-19 cases of Japan and the U.S. Sample period: April 4, 2020 – June 23, 2021 ($n = 446$ days).

Table 16: Empirical results of the in-sample and out-of-sample analyses on the log-difference of new confirmed COVID-19 cases

Japan

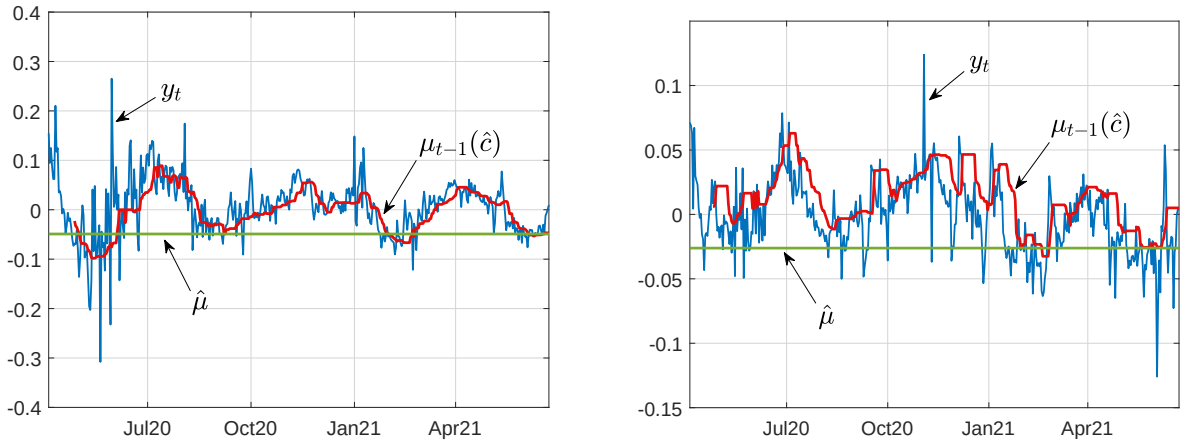
| $\hat{p}^{secotar}(H_0^*)$ | $\hat{p}^{setar}(H_0^*)$ | $RMSE^{const}$ | $RMSE^{ar}$ | $RMSE^{setar}$ | $RMSE^{secotar}$ |
|---|---|---|---|---|---|
| 0.075* | 0.003*** | 0.0423 | 0.0191 | 0.0184 | 0.0182 |
| $\hat{p}^{ar}(H_1^{eq})$ | $\hat{p}^{ar}(H_1^{ar})$ | $\hat{p}^{ar}(H_1^{secotar})$ | $\hat{p}^{setar}(H_1^{eq})$ | $\hat{p}^{setar}(H_1^{setar})$ | $\hat{p}^{setar}(H_1^{secotar})$ |
| 0.146 | 0.927 | 0.073* | 0.796 | 0.602 | 0.398 |

United States

| $\hat{p}^{secotar}(H_0^*)$ | $\hat{p}^{setar}(H_0^*)$ | $RMSE^{const}$ | $RMSE^{ar}$ | $RMSE^{setar}$ | $RMSE^{secotar}$ |
|---|---|---|---|---|---|
| 0.042** | 0.123 | 0.0321 | 0.0214 | 0.0219 | 0.0204 |
| $\hat{p}^{ar}(H_1^{eq})$ | $\hat{p}^{ar}(H_1^{ar})$ | $\hat{p}^{ar}(H_1^{secotar})$ | $\hat{p}^{setar}(H_1^{eq})$ | $\hat{p}^{setar}(H_1^{setar})$ | $\hat{p}^{setar}(H_1^{secotar})$ |
| 0.033** | 0.984 | 0.016** | 0.043** | 0.979 | 0.022** |

The target variable $y$ is the log-difference of the daily new confirmed COVID-19 cases. Sample period: April 4, 2020 – June 23, 2021 ($n = 446$ days). For the in-sample analysis, we fit the SE-CoTAR and SETAR models and test the no-threshold-effect hypothesis $H_0^*$ via the bootstrap exp-LM test. Resulting p-values $\hat{p}(H_0^*)$ are reported in the table. For the out-of-sample analysis, we compare the constant-only, AR, SETAR, and SE-CoTAR models in the rolling window framework and compute RMSEs. Further, we perform the asymptotic Diebold-Mariano test for $H_0^{eq} : MSE^{ar} = MSE^{secotar}$. $H_1^{eq} : MSE^{ar} \neq MSE^{secotar}$. $H_1^{ar} : MSE^{ar} < MSE^{secotar}$. $H_1^{secotar} : MSE^{ar} > MSE^{secotar}$. We perform the same test after changing the benchmark model from AR to SETAR. Resulting p-values $\hat{p}(H_1)$ are reported in the table. Asterisks ***, **, and * indicate a rejection of the null hypothesis at the 1%, 5%, and 10% levels, respectively.

Figure 1: The log-difference of the number of daily new confirmed COVID-19 cases




Japan United States

In this figure, we plot the log-difference of daily new confirmed COVID-19 cases of Japan and the U.S. $\mu_t(\hat{c})$ = estimated conditional threshold based on the SE-CoTAR model. Month $t + \hat{d}$ is in regime 1 if $y_t < \mu_{t-1}(\hat{c})$, and it is in regime 2 if $y_t \geq \mu_{t-1}(\hat{c})$. $\hat{\mu}$ = estimated constant threshold based on the SETAR model. Sample period: April 4, 2020 – June 23, 2021 ($n = 446$ days).