Conditional Threshold Autoregression (CoTAR)

Kaiji Motegi* John W. Dennis† Shigeyuki Hamori‡

Kobe University IDA Yamato U. & Kobe U.

July 16, 2024

Abstract

We propose a new time series model called Conditional Threshold Autoregression (CoTAR), in which the threshold is specified as an empirical quantile of recent observations of a threshold variable. The conditional threshold is expected to trace economic and financial variables well, as a cut-off level of low and high regimes likely changes over time. The presence versus absence of conditional threshold effects can be tested via the wild bootstrap, and the out-of-sample predictive ability of CoTAR can be evaluated via the Diebold-Mariano test. We show that CoTAR satisfies desired statistical properties in both large and small samples. We apply the proposed model to the CBOE Volatility Indices of S&P 500 and major U.S. shares, obtaining desired in-sample and out-of-sample results.

JEL codes: C22, C24, C51.

Keywords: Nonlinear time series, profiling estimation, regime switch, self-exciting threshold autoregression (SETAR), threshold effect, Volatility Index (VIX).

*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

1 Introduction

It is well known that time series variables often have asymmetric properties below versus above a certain threshold. Such nonlinear responses are called *threshold effects*, and there is a vast literature on modeling and testing them. Threshold Autoregression (TAR) of Tong (1978) is one of the earliest and most well-known models in this field. In TAR, a target series y follows AR(p) processes with coefficients being different across regimes, and a regime switch is triggered when a threshold variable x crosses a threshold parameter μ.[1] Hansen (2000) proposed the threshold regression (TR) model to unify various forms of threshold effects. The TR model is designed for both cross-section and time series data, and includes TAR as a special case.

In TAR and TR models, the threshold parameter μ is constant over time. Recently, TAR and related models are extended in various ways to allow for a *time-varying* threshold μ_t. Time-varying threshold models are expected to fit economic and financial variables well, as a cut-off level of "low" and "high" regimes likely changes over time. A fixed magnitude of volatility shock in financial markets, for example, might surprise investors during a tranquil period but might be nothing surprising during a crisis period, since recent volatility levels are different across periods. This sort of relativity is common in many economic indicators, motivating the use of time-varying threshold models.

Time-varying or state-dependent threshold models can be categorized into several groups, depending on the specifications of thresholds. First, the threshold may be specified as a linear combination of observed covariates; TR is extended in this direction by Seo and Linton (2007) and Yu and Fan (2021); smooth transition autoregression (STAR) is extended by Dueker, Psaradakis, Sola, and Spagnolo (2013); regression kink models of Hansen (2017) are extended by Yang and Su (2018). Second, the threshold is assumed to follow a latent AR process; TR is extended in this direction by Zhu, Chen, and Lin (2019). Third, Yang, Lee, and Chen (2021) generalize TR by applying a Fourier approximation to the threshold. Fourth, Motegi, Cai, Hamori, and Xu (2020, MCHX2020) add time-varying threshold effects to heterogeneous autoregressive (HAR) models, where the thresholds are sample averages of recent observations of x for each sampling frequency.

We propose an alternative model called Conditional Threshold Autoregression (CoTAR), in which the threshold is an empirical quantile of recent observations of x. More precisely, the regime at time t is determined by whether $x_{t-d} < \mu_{t-d-1}(c)$

[1] Extensive surveys on TAR can be found at Hansen (2011) and Tong (2011, 2015), among others.

or $x_{t-d} \geq \mu_{t-d-1}(c)$, where the conditional threshold $\mu_t(c)$ is the $100c$-percentile of $\{x_t, x_{t-1}, \ldots, x_{t-m+1}\}$. The memory size m is chosen by the researcher, while the delay parameter d and the percentile parameter c can be estimated from data. The conditional threshold represents a "normal" level of recent x if c is around 0.5 and an "abnormal" level if c is far from 0.5. The conditional threshold traces the fluctuation of x, embodying the insight that recent economic conditions matter. When the threshold variable is chosen to be the lagged target variable (i.e., $x = y$), the model is called Self-Exciting CoTAR (SE-CoTAR).

The CoTAR approach is related to the conditional average approach of MCHX2020. Our proposed conditional threshold can be an "abnormal" level of x by setting c to be close to 0 or 1, whereas the threshold of MCHX2020 is restricted to be an average level of x. Besides, our framework is more general because we do not impose HAR restrictions. We also establish asymptotic properties of CoTAR, whereas MCHX2020 reported numerical and empirical evidence only.

Statistical inference of CoTAR is analogous to TAR. Let $\boldsymbol{\beta}_r$ be a vector of the intercept and AR parameters for regime $r \in \{1, 2\}$. We estimate $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and the nuisance parameters $\boldsymbol{\gamma} = (d, c)^\top$ via a two-step method called *profiling*. The no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ can be tested with the wild-bootstrap tests of Hansen (1996). We test the equal predictive accuracy hypothesis between CoTAR and a benchmark model via a well-known test of Diebold and Mariano (1995).

To derive asymptotic properties of the profiling estimator and the bootstrap tests for H_0^*, we pay careful attention to an identification of $\boldsymbol{\gamma}$. Under $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, threshold effects are absent and $\boldsymbol{\gamma}$ is unidentified. Under $H_1^* : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, threshold effects are present and $\boldsymbol{\gamma}$ is identified. There are also a continuum of intermediate cases since we consider a local alternative hypothesis. We demonstrate that the profiling estimator and the bootstrap tests for H_0^* satisfy desired asymptotic properties under standard regularity conditions. Further, our Monte Carlo simulations indicate that the proposed methods operate well in small samples.

We apply the SE-CoTAR model to Volatility Index (VIX) calculated by Chicago Board Options Exchange, a well-known measure of U.S. stock market uncertainty. We analyze monthly log-VIX of S&P 500 Index and five major U.S. shares: Apple, Amazon, Google, Goldman Sachs, and IBM. The bootstrap test of Hansen (1996) rejects the no-threshold-effect hypothesis H_0^* for all cases but for Goldman Sachs, indicating the relevance of conditional threshold effects. The mean squared forecast error of SE-CoTAR is smaller than that of our benchmark models (i.e., constant-only, AR,

and SETAR models) for S&P 500, Apple, Amazon, and IBM. For some of them, the superior out-of-sample performance of SE-CoTAR is statistically significant according to the Diebold-Mariano test. Overall, our proposed model traces and predicts VIX quite well, a useful finding for market participants and policymakers.

The rest of this paper is organized as follows. The CoTAR model is proposed in Section 2. Procedures of estimation and hypothesis testing are described in Section 3. Asymptotic properties of the proposed methods are derived in Section 4. In Section 5, we evaluate the finite sample performance of SE-CoTAR via Monte Carlo simulations. We analyze the monthly log-VIX data in Section 6, and give some concluding remarks in Section 7. Omitted technical details are collected in Appendices. In a separate supplemental material, we present additional simulation results and another empirical study on the pandemic of novel coronavirus disease.

2 Conditional Threshold Autoregression

In this section, we propose the CoTAR model and rewrite it in a matrix form. We use the following notation throughout: $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers, $\#A$ is the number of elements of set A, and $\mathbf{1}(A)$ is the indicator function which equals 1 if event A occurs and 0 otherwise.

Let y_t and x_t be a target variable and a threshold variable at time $t \in \{1, \ldots, n\}$, respectively. Consider Tong's (1978) TAR model with two regimes:

$$y_t = \begin{cases} \alpha_1 + \sum_{k=1}^{p} \phi_{1k} y_{t-k} + u_t & \text{if } \ x_{t-d} < \mu, \\ \alpha_2 + \sum_{k=1}^{p} \phi_{2k} y_{t-k} + u_t & \text{if } \ x_{t-d} \geq \mu, \end{cases} \tag{1}$$

where $(\alpha_r, \phi_{r1}, \ldots, \phi_{rp})$ are regression parameters in regime $r \in \{1, 2\}$; $d \in \mathbb{N}$ is the delay parameter; $\mu \in \mathbb{R}$ is the threshold parameter. Usual conditions on the error term u_t include a martingale difference sequence and finite second moment; the assumptions on u_t are defined more precisely in Section 4. Self-Exciting TAR (SETAR) is a special case of (1) with $x_t = y_t$.

In model (1), y has different autocorrelation structures below versus above the unconditional threshold μ. The term "unconditional" means that μ is time-independent and chosen from the entire memory $\mathcal{X}_1^n = \{x_1, \ldots, x_n\}$. We propose to replace μ with a conditional threshold μ_t which is time-dependent and chosen from a local memory $\mathcal{X}_{t-m+1}^t = \{x_{t-m+1}, \ldots, x_t\}$. Specifically, we propose the *Conditional Threshold*

Autoregressive (CoTAR) model:

$$y_t = \begin{cases} \alpha_1 + \sum_{k=1}^{p} \phi_{1k} y_{t-k} + u_t & \text{if } \ x_{t-d} < \mu_{t-d-1}(c), \\ \alpha_2 + \sum_{k=1}^{p} \phi_{2k} y_{t-k} + u_t & \text{if } \ x_{t-d} \geq \mu_{t-d-1}(c), \end{cases} \tag{2}$$

where the conditional threshold $\mu_t(c)$ is the mc-th smallest value (or equivalently the $100c$-percentile) of $\mathcal{X}_{t-m+1}^{t}$; $m = \#\mathcal{X}_{t-m+1}^{t}$ is the size of the local memory; $c \in \{1/m, 2/m, \ldots, 1\}$ signifies the relevant percentile, where the possible values c can take are restricted to be discrete so that $mc \in \{1, \ldots, m\}$. Self-Exciting CoTAR (SE-CoTAR) is a special case of (2) with $x_t = y_t$.

A unique feature of CoTAR lies in the specification of the conditional threshold $\mu_{t-d-1}(c)$. If $m = 1$, then $c = 1$ by construction and hence $\mu_{t-d-1}(c) = x_{t-d-1}$. In this special case, (2) reduces to (1) with the threshold variable $\Delta x_{t-d} = x_{t-d} - x_{t-d-1}$ and the known threshold $\mu = 0$. If $c = (2m)^{-1}(m+1)$, then $\mu_{t-d-1}(c)$ (almost) coincides with the median of $\mathcal{X}_{t-d-m}^{t-d-1}$. In this case, regime 1 arises when x is below the "normal" level given the local memory, while regime 2 arises when x is above it.

If the value of c is close to the lower bound $1/m$ or the upper bound 1, then a regime switch is triggered by a rare event of x crossing an "abnormal" level given the local memory. Suppose, for example, that we fit SE-CoTAR with $m = 12$ and $c = 10/12 = 0.833$ to monthly log-VIX of S&P 500. Then, the conditional threshold is the third largest observation of the log-VIX in recent 12 months, hence regime 2 is a phase of extremely high volatility (in the relative sense). The conditional threshold approach is intuitively reasonable, since investors likely judge current market conditions relative to the recent past, not to a constant cut-off value. Similar arguments might well apply to other variables such as economic growth, public debt, and the spread of novel coronavirus disease.

One might be tempted to specify the conditional threshold as a conditional average of x (i.e., $\mu_t = m^{-1} \sum_{\ell=0}^{m-1} x_{t-\ell}$); indeed, MCHX2020 took this approach in the HAR framework. This is an intuitively plausible specification, and there is a computational advantage that the percentile parameter c disappears. A possible limitation, however, is that the threshold level cannot be an "abnormal" level unlike the conditional quantile specification with c being away from 0.5. A more general specification than the simple average would be a weighted average: $\mu_t = \sum_{\ell=0}^{m-1} w_\ell x_{t-\ell}$ with $w_\ell \geq 0$ and $\sum_{\ell=0}^{m-1} w_\ell = 1$. If the entire weighting scheme $\boldsymbol{w} = (w_0, \ldots, w_{m-1})^\top$ is estimated without any restrictions, then parameter proliferation is likely to occur. If $\boldsymbol{w}$

is prescribed or tightly parameterized, then there is a higher risk of misspecification. Further, whether $\boldsymbol{w}$ is estimated or not, an identification problem arises since two distinct weighting schemes can lead to an identical profile of regimes for all time periods. Hence, the proposed conditional quantile approach is better balanced between flexible specification and practical implementation than the conditional average approaches.

For elaborating the statistical properties of CoTAR, it is of use to rewrite model (2) in a matrix form. Stack the parameters as

$$\boldsymbol{\beta}_1 = \begin{bmatrix} \alpha_1 \\ \phi_{11} \\ \vdots \\ \phi_{1p} \end{bmatrix}, \quad \boldsymbol{\beta}_2 = \begin{bmatrix} \alpha_2 \\ \phi_{21} \\ \vdots \\ \phi_{2p} \end{bmatrix}, \quad \underbrace{\boldsymbol{\beta}}_{K\times 1} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} d \\ c \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix}, \tag{3}$$

where $K = 2(p+1)$. The target parameters $\boldsymbol{\theta}$ are partitioned into the regression parameters $\boldsymbol{\beta}$ and the nuisance parameters $\boldsymbol{\gamma}$. One or both elements of $\boldsymbol{\gamma}$ could be pre-specified by the researcher, but we estimate both of them to avoid misspecification. To focus on the estimation of (d, c), we sidestep lag selection issues by assuming that (p, m) are known.[2] Define binary variables which determine the regime: $I_{1t}(c) = \mathbf{1}\{x_t < \mu_{t-1}(c)\}$ and $I_{2t}(c) = \mathbf{1}\{x_t \geq \mu_{t-1}(c)\}$. Stack the regressors as

$$\underbrace{\boldsymbol{z}_{t-1}}_{(p+1)\times 1} = (1, y_{t-1}, \ldots, y_{t-p})^\top, \qquad \underbrace{\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})}_{K\times 1} = \begin{bmatrix} \boldsymbol{z}_{t-1} I_{1,t-d}(c) \\ \boldsymbol{z}_{t-1} I_{2,t-d}(c) \end{bmatrix}. \tag{4}$$

Then, model (2) can be rewritten as a single equation:

$$y_t = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top \boldsymbol{\beta} + u_t. \tag{5}$$

3 Statistical inference of CoTAR

We describe profiling estimation of $\boldsymbol{\theta}$ in Section 3.1, testing the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ in Section 3.2, and testing the equal predictive accuracy hypothesis H_0^{eq} between the CoTAR model and a benchmark model in Section 3.3.

[2] If the model is misspecified, statistical inference may produce a spurious result. For example, consider setting $p = 1$ when the truth is $p > 1$. Then, estimation of $\boldsymbol{\beta}$ will be biased in general.

3.1 Profiling estimation of parameters

The quadratic loss function of the CoTAR model is written as $\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{n}\{y_t - \mathrm{E}_{\boldsymbol{\theta}}(y_t \,|\, y_{t-1}, \ldots, y_1)\}^2$, where $\boldsymbol{\theta}$ is defined in (3) and $\mathrm{E}_{\boldsymbol{\theta}}(y_t \,|\, y_{t-1}, \ldots, y_1)$ is the conditional expectation of y_t given $\boldsymbol{\theta}$ and the past observations $\{y_{t-1}, \ldots, y_1\}$. The least squares (LS) estimator for $\boldsymbol{\theta}$ is defined as the minimizer of $\mathcal{L}(\boldsymbol{\theta})$, and it can be computed via a two-step approach called *profiling*. To this end, we define choice spaces of the nuisance parameters $\boldsymbol{\gamma} = (d, c)^\top$. Let $D \subseteq \mathbb{N}$ be the choice space of the delay parameter d. Taking the memory size m as given, the largest possible space of c is given by $\bar{C} = \{1/m, 2/m, \ldots, 1\}$. 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. For some $c \in \bar{C}$, $\delta_r(c)$ may be too small to identify both regimes in finite samples. A practical compromise which originates in a suggestion by Andrews (1993) is to restrict the choice space so that both regimes account for at least 15% of the entire sample:

$$C = \left\{c \in \bar{C} \,|\, \min\{\delta_1(c), \delta_2(c)\} > 0.15\right\}. \tag{6}$$

Using the Cartesian product, the space of $\boldsymbol{\gamma}$ is written as $\Gamma = D \times C$.

Fixing $\boldsymbol{\gamma} \in \Gamma$, the regressors $\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})$ in (5) can be computed from data; hence, the LS estimator for $\boldsymbol{\beta}$ conditional on $\boldsymbol{\gamma}$ can be computed by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left\{\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top\right\}^{-1}\left\{\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})y_t\right\}. \tag{7}$$

The resulting residual is given by $\hat{u}_t(\boldsymbol{\gamma}) = y_t - \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$. The LS estimator for $\boldsymbol{\gamma}$ is given by

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}\in\Gamma}{\operatorname{argmin}} \sum_{t=1}^{n} \hat{u}_t(\boldsymbol{\gamma})^2.$$

The LS estimator for $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}})$. Finally, write $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\gamma}}^\top)^\top$.

3.2 Testing the no-threshold-effect hypothesis

Consider testing the null hypothesis of no threshold effects against a fixed alternative hypothesis:

$$H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 \quad \text{vs.} \quad H_1^* : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2,$$

where $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ are defined in (3). Clearly, H_0^* is a special case of general linear restrictions $H_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{q}$ with

$$\underbrace{\boldsymbol{R}}_{(p+1)\times K} = (\boldsymbol{I}_{p+1},\ -\boldsymbol{I}_{p+1}), \qquad \boldsymbol{q} = \boldsymbol{0}_{(p+1)\times 1}, \tag{8}$$

where $\boldsymbol{I}_{p+1}$ is the identity matrix of dimension $p+1$. Under H_0^*, the CoTAR model reduces to the single-regime AR(p) which does not depend on the threshold variable x. Hence, $\boldsymbol{\gamma} = (d, c)^\top$ is unidentified under H_0^* and identified under H_1^*. This identification issue makes the testing problem non-standard, hence the no-threshold-effect hypothesis H_0^* is a particularly important class of linear restrictions H_0.

More generally, we can consider a local alternative hypothesis in which $\boldsymbol{\beta}_2$ approaches $\boldsymbol{\beta}_1$ at an arbitrary rate as $n \to \infty$. The rate of convergence determines what Andrews and Cheng (2012) call the *identification category*, and it affects the asymptotic power of our proposed test. We consider the local alternative hypothesis when deriving asymptotic theory in Section 4 and when performing Monte Carlo simulations in the separate supplemental material.

In what follows, we describe the wild-bootstrap tests of Hansen (1996) for H_0^*. To proceed, it is of use to define some key quantities conditional on $\boldsymbol{\gamma} \in \Gamma$. The regression score associated with model (5) is given by

$$\boldsymbol{s}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) u_t. \tag{9}$$

The estimated regression score under H_1^* is given by

$$\hat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) \hat{u}_t(\boldsymbol{\gamma}). \tag{10}$$

The Wald test statistic associated with H_0^* is formulated as

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

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ is defined in (7). The heteroscedasticity-robust covariance matrix estimator is given by

$$\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}) \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}, \tag{11}$$

where

$$\hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{t=1}^{n} \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma})\hat{\boldsymbol{s}}_t(\boldsymbol{\gamma})^\top, \qquad \boldsymbol{M}_n(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^\top. \tag{12}$$

Common ways for incorporating all possible values of $\boldsymbol{\gamma}$ include the supremum, average, and exponential transformations:

$$\sup\mathcal{W}_n \equiv \sup_{\boldsymbol{\gamma}\in\Gamma} \mathcal{W}_n(\boldsymbol{\gamma}), \qquad \text{ave}\mathcal{W}_n \equiv \frac{1}{\#\Gamma}\sum_{\boldsymbol{\gamma}\in\Gamma} \mathcal{W}_n(\boldsymbol{\gamma}), \tag{13}$$

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

Since $\boldsymbol{\gamma}$ is unidentified under H_0^*, the asymptotic distributions of the test statistics (13) under H_0^* are non-standard; hence, we adopt the wild bootstrap of Hansen (1996). Let $g(\mathcal{W}_n)$ denote either $\sup\mathcal{W}_n$, $\text{ave}\mathcal{W}_n$, 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

$$\mathcal{W}_n^{(b)}(\boldsymbol{\gamma}) = \hat{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma})^\top \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\boldsymbol{R}^\top\left\{\boldsymbol{R}\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma})\boldsymbol{R}^\top\right\}^{-1}\boldsymbol{R}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\hat{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}); \tag{14}$$

$$\hat{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma})\xi_t^{(b)}; \tag{15}$$

$\boldsymbol{R}$, $\hat{\boldsymbol{s}}_t(\boldsymbol{\gamma})$, $\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma})$, and $\boldsymbol{M}_n(\boldsymbol{\gamma})$ are defined in (8), (10), (11), and (12), respectively.

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 the 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 LS residual from model (5) with H_0^* being

imposed (i.e., the single-regime AR(p) model). The estimated regression score under H_0^* is given by $\tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})\tilde{u}_t$. The heteroscedasticity-robust covariance matrix estimator is given by

$$\tilde{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\tilde{\boldsymbol{S}}_n(\boldsymbol{\gamma})\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}, \qquad \tilde{\boldsymbol{S}}_n(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{t=1}^{n}\tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma})\tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma})^\top. \tag{16}$$

The conditional LM test statistic is given by

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

The transformed LM test statistics are obtained via (13), where $\mathcal{W}_n(\boldsymbol{\gamma})$ is replaced with $\mathcal{LM}_n(\boldsymbol{\gamma})$. Steps 1-4 are executed with (14)-(15) being replaced with

$$\mathcal{LM}_n^{(b)}(\boldsymbol{\gamma}) = \tilde{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma})^\top \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\boldsymbol{R}^\top\left\{\boldsymbol{R}\tilde{\boldsymbol{V}}_n(\boldsymbol{\gamma})\boldsymbol{R}^\top\right\}^{-1}\boldsymbol{R}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\tilde{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}),$$
$$\tilde{\boldsymbol{v}}_n^{(b)}(\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\boldsymbol{s}}_t(\boldsymbol{\gamma})\xi_t^{(b)}.$$

3.3 Testing the equal predictive accuracy hypothesis

In this section, we evaluate the out-of-sample predictive ability of the CoTAR model. There are two natural competitors to CoTAR. The first one is a single-regime AR(p) model: $y_t = \alpha + \sum_{k=1}^{p}\phi_k y_{t-k} + u_t$. The CoTAR and AR models are nested; the former includes the latter as a special case with $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Hence, comparing the forecast performance of CoTAR and AR amounts to evaluating the relevance of conditional threshold effects relative to non-threshold effects in the out-of-sample sense.

The second competitor is the TAR model. In view of (1) and (2), the only difference between TAR and CoTAR is whether the threshold is constant μ or conditional $\mu_t(c)$. Hence, comparing the forecast performance of CoTAR and TAR amounts to evaluating the relevance of conditional threshold effects relative to constant threshold effects in the out-of-sample sense.

To implement the profiling estimation of TAR, we need to define choice space of μ. Let $x_{[1]} \leq \cdots \leq x_{[n]}$ be a sorted version of $\{x_t\}_{t=1}^n$. The choice space of μ is specified as

$$\mathcal{X}_{\kappa,n} = \left\{x_{[\lfloor 0.5(1-\kappa)n\rfloor]}, \ldots, x_{[\lfloor\{1-0.5(1-\kappa)\}n\rfloor]}\right\}, \tag{17}$$

where $\lfloor a \rfloor$ is the largest integer not larger than $a \in \mathbb{R}$, and $\kappa \in [0, 1)$ signifies the fraction of $\#\mathcal{X}_{\kappa,n}$ to n. 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. Following the suggestion of Andrews (1993), we pick $\kappa = 0.7$ so that each regime accounts for at least 15% of the entire sample.

To compare the forecast performance of CoTAR and a competing model, we perform rolling window out-of-sample prediction with the whole sample size $N \in \mathbb{N}$. Suppose that the window size is fixed at $n = \lfloor \tau N \rfloor$ with $\tau = 0.8$.[3] Estimate a model for $\{y_t\}_{t=1}^{n}$, and compute a one-step-ahead forecast of y_{n+1} denoted as $\hat{y}_{n+1}$. Similarly, estimate the model for $\{y_t\}_{t=2}^{n+1}$ and compute a forecast of y_{n+2} denoted as $\hat{y}_{n+2}$. Repeat until $\{\hat{y}_t\}_{t=n+1}^{n+T}$ is obtained, where $T = N - n$ is the number of windows. The forecast error is given by $\hat{e}_t = y_t - \hat{y}_t$ for $t \in \{n+1, \ldots, n+T\}$. Mean squared forecast error (MSE) is defined as $MSE = T^{-1} \sum_{t=n+1}^{n+T} \hat{e}_t^2$, and root mean squared forecast error (RMSE) is defined as $RMSE = MSE^{1/2}$. The (R)MSE is a well-known measure of prediction inaccuracy.

A simple way to compare the predictive accuracies of CoTAR and a competing model is to compare their MSEs; the smaller value of MSE implies the higher forecast accuracy. Further, Diebold and Mariano (1995, DM1995) established a novel approach to test if the spread between two MSEs is statistically significant. We adopt the asymptotic S_1-test of DM1995 to compare the MSE of CoTAR and the MSE of a benchmark model. When the benchmark is AR, a core component of the test statistic S_1 is $MSE^{ar} - MSE^{cotar}$, where MSE^{ar} and MSE^{cotar} are the MSEs of AR and CoTAR, respectively. The null hypothesis of the DM tests is $H_0^{eq} : MSE^{ar} = MSE^{cotar}$ (i.e., equal predictive accuracy). DM1995 showed that, under H_0^{eq} and some mild regularity conditions, the S_1-test statistic follows the standard normal distribution asymptotically. There are three types of alternative hypothesis, and the decision rules with nominal size $a \in (0, 1)$ are as follows.

- $H_1^{eq} : MSE^{ar} \neq MSE^{cotar}$ (two-sided test). Reject H_0^{eq} in favor of H_1^{eq} if $|S_1| > \Phi^{-1}(1 - a/2)$, where $\Phi^{-1}(\cdot)$ is the inverse distribution function of the standard normal distribution.

- $H_1^{ar} : MSE^{ar} < MSE^{cotar}$ (one-sided test). Reject H_0^{eq} in favor of H_1^{ar} if $S_1 < \Phi^{-1}(a)$.

[3] A similar value for τ, say 0.7 or 0.9, can also be used without any technical problems. Besides, an expanding window design is also possible in which all windows begin at $t = 1$.

$H_1^{cotar} : MSE^{ar} > MSE^{cotar}$ (one-sided test). Reject H_0^{eq} in favor of H_1^{cotar} if $S_1 > \Phi^{-1}(1-a)$.

When a benchmark model is TAR, a core component of the test statistic S_1 is $MSE^{tar} - MSE^{cotar}$, where MSE^{tar} is the MSE of TAR. The null hypothesis of the DM tests is $H_0^{eq} : MSE^{tar} = MSE^{cotar}$. An alternative hypothesis is either $H_1^{eq} : MSE^{tar} \neq MSE^{cotar}$, $H_1^{tar} : MSE^{tar} < MSE^{cotar}$, or $H_1^{cotar} : MSE^{tar} > MSE^{cotar}$. The testing procedure is the same as when the benchmark is AR.

4 Asymptotic theory

In this section, we derive asymptotic properties of the profiling estimator and the bootstrap tests for the no-threshold-effect hypothesis H_0^*. We set up notation and assumptions in Section 4.1, and present main theorems in Section 4.2. We use the following notation throughout: the Euclidean norm of any k-dimensional vector $\boldsymbol{c} \in \mathbb{R}^k$ is denoted as $\|\boldsymbol{c}\| = (\boldsymbol{c}^\top \boldsymbol{c})^{1/2}$, *convergence in probability* is denoted by $\xrightarrow{p}$, and *weak convergence* is denoted by $\Rightarrow$.

4.1 Notation and assumptions

First, define $\tilde{x}_t(\boldsymbol{\gamma}) = x_{t-d} - \mu_{t-d-1}(c)$ so that regime 1 arises at time t if $\tilde{x}_t(\boldsymbol{\gamma}) < 0$ and regime 2 arises if $\tilde{x}_t(\boldsymbol{\gamma}) \geq 0$. With this transformation, we can directly apply the existing asymptotic theory of TAR. Let $\bar{z}_t = \max_{\boldsymbol{\gamma}\in\Gamma}[\mathrm{tr}\{\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \boldsymbol{Z}_t(\boldsymbol{\gamma})\}]^{1/2}$ be the norm of regressors, where $\boldsymbol{Z}_t(\boldsymbol{\gamma})$ is defined in (4). Define some population quantities:

$$\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \boldsymbol{M}(\boldsymbol{\gamma}_1)^{-1} \boldsymbol{S}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \boldsymbol{M}(\boldsymbol{\gamma}_2)^{-1}, \tag{18}$$

$$\boldsymbol{S}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \mathrm{E}\left\{\boldsymbol{s}_t(\boldsymbol{\gamma}_1)\boldsymbol{s}_t(\boldsymbol{\gamma}_2)^\top\right\}, \qquad \boldsymbol{M}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \mathrm{E}\left\{\boldsymbol{Z}_t(\boldsymbol{\gamma}_1)\boldsymbol{Z}_t(\boldsymbol{\gamma}_2)^\top\right\},$$

where $\boldsymbol{s}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) u_t$ as defined in (9). We will sometimes abbreviate $\boldsymbol{V}(\boldsymbol{\gamma}) = \boldsymbol{V}(\boldsymbol{\gamma}, \boldsymbol{\gamma})$, $\boldsymbol{S}(\boldsymbol{\gamma}) = \boldsymbol{S}(\boldsymbol{\gamma}, \boldsymbol{\gamma})$, and $\boldsymbol{M}(\boldsymbol{\gamma}) = \boldsymbol{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma})$ when appropriate; they are the population counterparts of $\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma})$, $\hat{\boldsymbol{S}}_n(\boldsymbol{\gamma})$ and $\boldsymbol{M}_n(\boldsymbol{\gamma})$ defined in (11)-(12). Impose some basic assumptions which are analogous to Assumption 1 of Hansen (1996).

Assumption 1. *(i)* $\{y_t, x_t, u_t\}$ *are strictly stationary and absolutely regular with mixing coefficients* $\eta(k) = O(k^{-A})$ *for some* $A > \nu(\nu-1)^{-1}$ *and* $\nu > 1$. *(ii) For some* $\iota > \nu$, $\mathrm{E}|\bar{z}_t|^{4\iota} < \infty$ *and* $\mathrm{E}|u_t|^{4\iota} < \infty$. *(iii)* $\inf_{\boldsymbol{\gamma}\in\Gamma} det\{\boldsymbol{M}(\boldsymbol{\gamma})\} > 0$. *(iv)* $\tilde{x}_t(\boldsymbol{\gamma})$ *has*

density function $f(x)$ *such that* $\sup_x f(x) < \infty$. *(v) The true value of* $\boldsymbol{\theta}$ *is an interior point of* Θ.

Assumption 1 is standard in the TAR literature. The mixing condition in item (i) controls the degree of serial dependence. Assumption 1 is satisfied if $\{u_t\}$ is iid with finite 4ι-th moment for some $\iota > 1$ and a well-known stability condition of univariate AR(p) holds for each regime $r \in \{1, 2\}$. Such a regime-wise stability condition is generally stronger than needed to ensure the ergodicity of TAR processes (Chan and Tong, 1985, Chen and Tsay, 1991); Chan (1993) shows that a strong form of geometric ergodicity holds for TAR processes under a faster mixing rate. Further, Hansen (1996) notes that it is likely that a martingale difference condition is sufficient in place of the iid condition. These observations apply to CoTAR as well.

To analyze the asymptotic behavior of our proposed methods, we perform a thought experiment where $\boldsymbol{\beta}_2$ depends on the sample size: $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2n}$. Consider a local alternative hypothesis:

$$H_1^*(\boldsymbol{\lambda}_n) : \boldsymbol{\beta}_{2n} = \boldsymbol{\beta}_1 + n^{-\frac{1}{2}}\boldsymbol{\lambda}_n, \qquad \boldsymbol{\lambda}_n \in \mathbb{R}^{p+1}. \tag{19}$$

The drift term $n^{-1/2}\boldsymbol{\lambda}_n$ determines distance between regimes 1 and 2. We adopt Andrews and Cheng's (2012) identification categories, which are defined as follows. First, $\boldsymbol{\gamma}$ is *unidentified* when $\boldsymbol{\lambda}_n = \mathbf{0}$, in which case $H_1^*(\boldsymbol{\lambda}_n)$ coincides with $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Second, $\boldsymbol{\gamma}$ is *weakly* identified when $\boldsymbol{\lambda}_n \neq \mathbf{0}$ and $\lim_{n\to\infty} \|\boldsymbol{\lambda}_n\| < \infty$. Third, $\boldsymbol{\gamma}$ is *semi-strongly* identified when $n^{-1/2}\boldsymbol{\lambda}_n \to \mathbf{0}$ and $\|\boldsymbol{\lambda}_n\| \to \infty$. Fourth, $\boldsymbol{\gamma}$ is *strongly* identified when $n^{-1/2}\boldsymbol{\lambda}_n \to \boldsymbol{c}$ for some $\boldsymbol{c} \neq \mathbf{0}$.

To facilitate exposition, define:

$$\underbrace{\bar{\boldsymbol{\lambda}}_n}_{K\times 1} = \begin{bmatrix} \mathbf{0}_{(p+1)\times 1} \\ \boldsymbol{\lambda}_n \end{bmatrix}, \quad \underbrace{\bar{\boldsymbol{\lambda}}}_{K\times 1} = \lim_{n\to\infty} \bar{\boldsymbol{\lambda}}_n, \quad \underbrace{\bar{\boldsymbol{\beta}}_1}_{K\times 1} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_1 \end{bmatrix}, \quad \underbrace{\boldsymbol{\beta}_n}_{K\times 1} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_{2n} \end{bmatrix},$$

where $\boldsymbol{\lambda}_n$ and $\boldsymbol{\beta}_{2n}$ are defined in (19). By construction, (19) can be rewritten as

$$H_1^*(\boldsymbol{\lambda}_n) : \boldsymbol{\beta}_n = \bar{\boldsymbol{\beta}}_1 + n^{-\frac{1}{2}}\bar{\boldsymbol{\lambda}}_n, \qquad \boldsymbol{\lambda}_n \in \mathbb{R}^{p+1}. \tag{20}$$

Finally, let $\mathcal{G}(\boldsymbol{\gamma})$ be a mean zero Gaussian process with covariance kernel $\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ defined in (18).

4.2 Main theorems

Asymptotic properties of the profiling estimator are as follows. First, distance between the conditional LS estimator for $\boldsymbol{\beta}$ and its true value, $\|\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_n\|$, converges in probability to 0 as $n \to \infty$ *uniformly* over $\boldsymbol{\gamma} \in \Gamma$. It implies consistency of the profiling estimator for $\boldsymbol{\beta}$, denoted as $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}})$, under all identification categories considered: $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n \xrightarrow{p} \mathbf{0}$. Second, when $\boldsymbol{\gamma}$ is unidentified or weakly identified, the profiling estimator $\hat{\boldsymbol{\gamma}}$ is not consistent for the true value $\boldsymbol{\gamma}$. In these cases, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}_1)$ is *not* asymptotically normal, which requires us to use a functional over Γ to test $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. These results are summarized in Theorem 1.[4]

Theorem 1. *Impose Assumption 1, then the following are true. (i)* $\sup_{\boldsymbol{\gamma}\in\Gamma} ||\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_n|| \xrightarrow{p} 0$ *as* $n \to \infty$. *(ii) Under* $H_1^*(\boldsymbol{\lambda}_n)$ *with* $\lim_{n\to\infty} \|\boldsymbol{\lambda}_n\| < \infty$, $\sqrt{n}\{\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \bar{\boldsymbol{\beta}}_1\} \Rightarrow \bar{\boldsymbol{\lambda}} + \mathcal{G}(\boldsymbol{\gamma})$ *as* $n \to \infty$.

Proof. A proof of Theorem 1 is provided in Appendix A.1. □

Now consider the wild-bootstrap test for the no-threshold-effect hypothesis H_0^*. A key condition for the asymptotic validity of this test is the uniform convergence of the conditional LS estimator, which has been established in Theorem 1.(i). Another crucial condition is that the conditional Wald or LM test statistic should converge weakly to a chi-squared process over $\boldsymbol{\gamma} \in \Gamma$ under H_0^*; this follows as an immediate implication of the weak convergence of $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ established in Theorem 1.(ii).

Under H_0^*, asymptotic distributions of the sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM statistics are non-standard. Let $\hat{\boldsymbol{v}}_n(\boldsymbol{\gamma}) = n^{-1/2} \sum_{t=1}^n \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma})\xi_t$ with $\xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$; this notation is in accordance with (15). Then, $\hat{\boldsymbol{v}}_n(\boldsymbol{\gamma})$ *converges weakly in probability* to a mean zero Gaussian process with covariance kernel $\boldsymbol{S}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ in the sense of Giné and Zinn (1990), where $\boldsymbol{S}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ is defined in (18). This implies that, under H_0^*, the wild-bootstrap p-value $\hat{p}_n^B(H_0^*)$ converges in distribution to a uniform random variable.

Under $H_1^*(\boldsymbol{\lambda}_n)$, asymptotic power properties of the test depend on identification categories. When $\boldsymbol{\gamma}$ is weakly identified, the power of the test is not guaranteed to approach 1 as $n \to \infty$. This is intuitive because the signal of threshold effects is too weak to distinguish the two regimes. When $\boldsymbol{\gamma}$ is semi-strongly or strongly identified,

[4]Different assumptions, estimators, and identification categories can result in different asymptotic distributions for $\hat{\gamma}$ under $H_1^*(\boldsymbol{\lambda}_n)$ (e.g., Chan, 1993, Hansen, 2000, Andrews and Cheng, 2012, Yang, Lee, and Chen, 2021). Deriving the asymptotic distribution of $\hat{\gamma}$ would be an interesting task, but is beyond the scope of this paper.

the test rejects H_0^* with probability approaching 1. These results are summarized in Theorem 2.

Theorem 2. *Impose Assumption 1, then the following are true. (i) Under* $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$*, the wild-bootstrap p-value* $\hat{p}_n^B(H_0^*)$ *is asymptotically uniform on* $[0,1]$*. (ii) Under* $H_1^*(\boldsymbol{\lambda}_n)$ *with* $\|\boldsymbol{\lambda}_n\| \to \infty$*,* $\hat{p}_n^B(H_0^*) \xrightarrow{p} 0$ *as* $n \to \infty$ *and* $B \to \infty$*.*

Proof. A proof of Theorem 2 is provided in Appendix A.2. □

5 Monte Carlo simulation

Up to the previous sections, we considered general CoTAR models where target variable y and threshold variable x can be distinct variables. In our numerical and empirical studies below, we focus on SE-CoTAR with $y = x$ since it is a popular class of threshold models. This section conducts Monte Carlo simulations to evaluate the finite sample performance of our proposed methods. In Section 5.1, we inspect the empirical size and power of the bootstrap tests for the no-threshold-effect hypothesis H_0^*. In Section 5.2, we report rejection frequencies of the Diebold-Mariano tests for the equal predictive accuracy hypothesis H_0^{eq}. Additional simulation results are summarized in Section 5.3.

5.1 Testing the no-threshold-effect hypothesis

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

$$y_t = \begin{cases} \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t & \text{if } \ y_{t-d_0} < \mu_{t-d_0-1}(c_0), \\ \alpha_{20} + \phi_{20} y_{t-1} + \epsilon_t & \text{if } \ y_{t-d_0} \geq \mu_{t-d_0-1}(c_0), \qquad t \in \{1, \ldots, n\}, \end{cases} \tag{21}$$

where $\boldsymbol{\beta}_{10} = (\alpha_{10}, \phi_{10})^\top = (0, 0.2)^\top$, $d_0 = 1$, $c_0 = 0.5$, 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 $\{y_t, y_{t-1}, \ldots, y_{t-m+1}\}$. Memory size is fixed at $m = 6$, and it is assumed to be known. Consider two specifications for $\boldsymbol{\beta}_{20} = (\alpha_{20}, \phi_{20})^\top$:

Case 1: $\boldsymbol{\beta}_{20} = (0, 0.2)^\top$, in which case (21) reduces to the single-regime AR(1) process: $y_t = 0.2 y_{t-1} + \epsilon_t$. In this case, threshold effects are absent and the nuisance parameters $\boldsymbol{\gamma}_0 = (d_0, c_0)^\top$ are unidentified.

Case 2: $\boldsymbol{\beta}_{20} = (0.35, 0.55)^\top$, in which case (21) does not degenerate. In this case, threshold effects are present and $\boldsymbol{\gamma}_0$ is strongly identified.

We generate $J = 1000$ Monte Carlo samples of size $n \in \{125, 250, 500, 1000\}$ indepedently from DGP (21).

We fit a SE-CoTAR model with $p = 1$ to each Monte Carlo sample:

$$y_t = \begin{cases} \alpha_1 + \phi_1 y_{t-1} + u_t & \text{if } \; y_{t-d} < \mu_{t-d-1}(c), \\ \alpha_2 + \phi_2 y_{t-1} + u_t & \text{if } \; y_{t-d} \geq \mu_{t-d-1}(c), \end{cases} \tag{22}$$

where the conditional threshold $\mu_t(c)$ is specified as in (21). The choice space of the delay parameter d is $D = \{1, 2, 3\}$. The choice space of the percentile parameter c is given by (6) so that each regime accounts for at least 15% of the entire sample.

Based on model (22), we test the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, where $\boldsymbol{\beta}_1 = (\alpha_1, \phi_1)^\top$ and $\boldsymbol{\beta}_2 = (\alpha_2, \phi_2)^\top$. As described in Section 3.2, we adopt the wild-bootstrap tests of Hansen (1996) since $\boldsymbol{\gamma}_0$ is unidentified under H_0^*. The sup-Wald, ave-Wald, exp-Wald, sup-LM, ave-LM, and exp-LM test statistics are all compared. We use the heteroscedasticity-robust covariance matrix estimator, although the true error ϵ is homoscedastic. We compute rejection frequencies of each test, where the nominal size is $a = 0.05$ and the number of bootstrap samples is $B = 500$. The rejection frequencies correspond to empirical size for Case 1 (non-identification) and empirical power for Case 2 (strong identification).

The resulting rejection frequencies are reported in Table 1. Under Case 1, the empirical size of the six tests all converges to the nominal size as n grows, confirming Theorem 2.(i). The Wald tests tend to over-reject the correct null hypothesis when $n \leq 250$, although the size distortions diminish as n grows. The empirical size of the ave-Wald test, for instance, is $\{0.120, 0.085, 0.064, 0.058\}$ for $n \in \{125, 250, 500, 1000\}$, respectively. The sup-Wald and exp-Wald tests lead to even worse over-rejections than the ave-Wald test when $n = 125$. By contrast, the LM tests achieve accurate size. The empirical size of the ave-LM test is $\{0.040, 0.044, 0.046, 0.049\}$. The sup-LM and exp-LM tests perform as well as the ave-LM test. Thus, the LM tests are preferred to the Wald tests in terms of controlling the type-I error rate in small samples.

As expected from Theorem 2.(ii), the empirical power of each test approaches 1 as n grows in Case 2. The empirical power of the ave-LM test is $\{0.245, 0.558, 0.922, 1.000\}$ for $n \in \{125, 250, 500, 1000\}$, respectively (Table 1). The sup-LM and exp-LM tests are roughly as powerful as the ave-LM test. The Wald tests are more powerful than

Table 1: Rejection frequencies of the wild-bootstrap tests for the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$

	Case 1: H_0^* is true				Case 2: H_0^* is false			
test \ n	125	250	500	1000	125	250	500	1000
sup-Wald	0.197	0.100	0.070	0.068	0.550	0.805	0.986	1.000
ave-Wald	0.120	0.085	0.064	0.058	0.484	0.730	0.949	1.000
exp-Wald	0.191	0.099	0.071	0.066	0.548	0.811	0.988	1.000
sup-LM	0.045	0.026	0.041	0.054	0.208	0.630	0.968	1.000
ave-LM	0.040	0.044	0.046	0.049	0.245	0.558	0.922	1.000
exp-LM	0.046	0.028	0.047	0.052	0.224	0.648	0.973	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 $y_t = \alpha_{20} + \phi_{20} y_{t-1} + \epsilon_t$ if $y_{t-d_0} \geq \mu_{t-d_0-1}(c_0)$, where $\alpha_{10} = 0$, $\phi_{10} = 0.2$, $d_0 = 1$, $c_0 = 0.5$, $m = 6$, 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: $\alpha_{20} = 0$ and $\phi_{20} = 0.2$, hence the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ is true. Case 2: $\alpha_{20} = 0.35$ and $\phi_{20} = 0.55$, hence H_0^* is false. Sample size: $n \in \{125, 250, 500, 1000\}$. 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 $y_{t-d} \geq \mu_{t-d-1}(c)$. This table reports rejection frequencies of the wild-bootstrap tests for H_0^* across $J = 1000$ Monte Carlo samples, where the nominal size is $\alpha = 0.05$.

the LM tests for $n \leq 250$, but it should be interpreted with caution since the Wald tests have size distortions under Case 1. Overall, the simulation results in Table 1 are all reasonable, and we can conclude that our proposed LM tests achieve desired size and power properties in small and large samples.

5.2 Testing the equal predictive accuracy hypothesis

In this section, we compare the out-of-sample performance of the SE-CoTAR model (22) 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 in (1) with $y = x$. We consider three scenarios to make fair comparison. First, the DGP is SE-CoTAR (21) 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 (21).

We perform rolling window one-step-ahead prediction with the whole sample size $N \in \{500, 1000, 2000, 4000\}$. 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 DM1995 with nominal size $a = 0.05$, as described in Section 3.3.

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 2), 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 2.

We expect the following results for 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 frequencies of H_0^{eq} : $MSE^{setar} = MSE^{secotar}$ should be close to 0 whichever the alternative hypothesis is. Under Case 2, 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 2 that rejection frequencies 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 frequencies of H_0^{eq} should be close to 0 whichever the alternative hypothesis is. Under Case 2, 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 2, rejection frequencies 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.

Resulting rejection frequencies of the DM test are reported in Table 2. We begin with the first scenario, where the DGP is SE-CoTAR and the benchmark model is AR. Under Case 1, the DM test chooses H_1^{ar} more often than $H_1^{secotar}$, reflecting the correct and parsimonious specification of the AR model. This result is consistent with our con-

jecture 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. Under Case 2, the DM test chooses $H_1^{secotar}$ with probability approaching 1 as N grows. This is not surprising since SE-CoTAR is correctly specified and AR is misspecified under Case 2. Fixing $N = 1000$, the rejection frequency of H_0^{eq} is $\{0.329, 0.474, 0.000\}$ for $\{H_1^{eq}, H_1^{secotar}, H_1^{ar}\}$. When we raise the sample size from 1000 to 4000, the rejection frequency becomes $\{0.925, 0.958, 0.000\}$.

Table 2: Rejection frequencies of the Diebold-Mariano test for the equal predictive accuracy hypothesis H_0^{eq} between the SE-CoTAR and benchmark models

		Case 1: $\boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20}$				Case 2: $\boldsymbol{\beta}_{10} \neq \boldsymbol{\beta}_{20}$			
		Sample size N:				Sample size N:			
scenario	H_1	500	1000	2000	4000	500	1000	2000	4000
1	H_1^{eq}	0.103	0.108	0.083	0.105	0.134	0.329	0.595	0.925
1	$H_1^{secotar}$	0.010	0.012	0.011	0.007	0.206	0.474	0.713	0.958
1	H_1^{ar}	0.166	0.148	0.136	0.159	0.010	0.000	0.000	0.000
2	H_1^{eq}	0.040	0.041	0.037	0.048	0.160	0.344	0.535	0.811
2	$H_1^{secotar}$	0.075	0.062	0.058	0.069	0.254	0.469	0.663	0.892
2	H_1^{setar}	0.023	0.030	0.025	0.024	0.003	0.000	0.000	0.000
3	H_1^{eq}	0.046	0.046	0.049	0.046	0.071	0.072	0.180	0.279
3	$H_1^{secotar}$	0.068	0.056	0.059	0.067	0.046	0.011	0.006	0.001
3	H_1^{setar}	0.029	0.022	0.031	0.026	0.075	0.122	0.260	0.413

Scenario 1: DGP is SE-CoTAR and the benchmark model is AR. Scenario 2: DGP is SE-CoTAR and the benchmark model is SETAR. Scenario 3: DGP is SETAR and the benchmark model is SETAR. Case 1: $\alpha_{20} = 0$ and $\phi_{20} = 0.2$, hence there are not threshold effects ($\boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20}$). Case 2: $\alpha_{20} = 0.35$ and $\phi_{20} = 0.55$, hence there are threshold effects ($\boldsymbol{\beta}_{10} \neq \boldsymbol{\beta}_{20}$). Sample size: $N \in \{500, 1000, 2000, 4000\}$. The rolling window out-of-sample prediction is performed, where the window size is fixed at $n = 0.8N$. For each scenario, we perform the asymptotic Diebold-Mariano test with nominal size $a = 0.05$. H_0^{eq}: The MSEs of the SE-CoTAR and benchmark models are equal. H_1^{eq}: The two MSEs are not equal. $H_1^{secotar}$: The MSE of SE-CoTAR is smaller. H_1^{ar}: The MSE of AR is smaller. H_1^{setar}: The MSE of SETAR is smaller. This table reports the rejection frequencies across $J = 1000$ Monte Carlo samples.

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 2). 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 Case 2, the DM test selects $H_1^{secotar}$ with probability approaching 1. The rejection frequency in favor of $H_1^{secotar}$ is $\{0.254, 0.469, 0.663, 0.892\}$ for $N \in \{500, 1000, 2000, 4000\}$, respectively.

We now discuss the third scenario, where the DGP is SETAR and the benchmark model is SETAR. For Case 1, rejection frequencies are all below 10%, indicating that $MSE^{secotar}$ and MSE^{setar} are sufficiently close to each other under non-threshold effects (Table 2). Curiously, we observe low rejection rates even for Case 2. 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 2, the rejection frequencies of H_0^{eq} in favor of H_1^{setar} are $\{0.075, 0.122, 0.260, 0.413\}$ for $N \in \{500, 1000, 2000, 4000\}$, respectively. These are much lower than the corresponding rejection frequencies in favor of $H_1^{secotar}$ under the second scenario.

In summary, the simulation results in Table 2 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.

5.3 Further discussions

We perform further simulations to augment the main simulation of the previous sections. We overview these additional results in this section, and relegate complete results to the supplemental material. First, we consider local alternative hypotheses (19) with the nuisance parameters being weakly or semi-strongly identified in the sense of Andrews and Cheng (2012). These are intermediate cases between non-identification (Case 1) and strong identification (Case 2). We find that the empirical power of the bootstrap tests for the no-threshold-effect hypothesis H_0^* approaches 1 under semi-strong identification but not under weak identification, verifying Theorem 2.(ii). Besides, a rejection of the equal predictive accuracy hypothesis H_0^{eq} does not occur frequently under weak or semi-strong identification. This result suggests that asymptotically vanishing threshold effects are too weak for the DM test to detect the subtle gap between the SE-CoTAR model and the benchmark AR or SETAR model.

Second, we examine finite sample properties of the profiling estimator described in Section 3.1. Simulation results are all consistent with Theorem 1. The profiling estimator for $\boldsymbol{\beta}$ converges to its true value for all identification categories. The profiling estimator for $\boldsymbol{\gamma}$ neither converges nor diverges under non- and weak identification,

while it converges to its true value under semi-strong and strong identification.

Third, we raise the memory size m from 6 to 18. The small sample performance of the SE-CoTAR model becomes slightly worse, since more initial observations need to be secured for practical implementation. In terms of the relative predictive ability between the SE-CoTAR and SETAR models, the larger memory size makes the two MSEs closer to each other. 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 $m \to \infty$.

6 Empirical application

Measuring, modelling, and forecasting the volatility of stock returns are a vital area of research. Volatility Index (VIX) calculated by Chicago Board Options Exchange (CBOE) is a well-known proxy of stock return volatility. In particular, VIX of S&P 500 Index (SPX) is a representative measure of the U.S. stock market or more broadly the global financial market.[5] Starting in June 2010, CBOE also issues VIX of five major U.S. shares: Apple, Amazon, Google, Goldman Sachs (GS), and IBM. Analyzing the VIX of SPX and these individual shares are of direct interest for investors and policymakers, as efficient portfolio management and effective economic policies hinge on an accurate evaluation of financial risk. Based on this background, we analyze the VIX of SPX and the five shares using the SE-CoTAR model.

Numerous approaches are proposed for modelling and predicting VIX. Fernandes, Medeiros, and Scharth (2014) show that the strong persistence of VIX is adequately captured by HAR models. Psaradellis and Sermpinis (2016) combine HAR and a genetic algorithm-support vector regression model, which improves the statistical significance of HAR. Liu, Guo, and Qiao (2015) use GARCH, GJR, and Heston-Nandi models, whereas Qiao, Yang, and Li (2020) add observable dynamic jump intensity to GARCH models. We are not aware of any existing work that analyzes VIX with TAR-type models, and fill this gap by employing SE-CoTAR.

In Section 6.1, we present the VIX data and some preliminary analyses. In Section 6.2, we perform an in-sample analysis with a special emphasis on testing the no-threshold-effect hypothesis H_0^*. In Section 6.3, we perform an out-of-sample analysis using the Diebold-Mariano test for the equal predictive accuracy hypothesis H_0^{eq}.

[5]See Whaley (2009) for a general overview of VIX.

6.1 Data and preliminary analysis

Let VIX_0 be Volatility Index of SPX. Similarly, let $\text{VIX}_1, \ldots, \text{VIX}_5$ be Volatility Indices of Apple, Amazon, Google, GS, and IBM, respectively. Let VIX_{it} denote a realization of VIX_i with $i \in \{0, 1, \ldots, 5\}$ at month $t \in \{1, \ldots, n\}$. We analyze $y_{it} = \ln \text{VIX}_{it}$ for each separate $i \in \{0, 1, \ldots, 5\}$. Subscript i is omitted below unless there is risk of confusion. The VIX data are downloaded from Federal Reserve Economic Data (FRED) maintained by the Federal Reserve Bank of St. Louis. The sample period is January 1990 – May 2024 ($n = 413$ months) for SPX and June 2010 – May 2024 ($n = 168$ months) for the five individual shares. These are the longest possible sample periods available at FRED as of the analysis.[6]

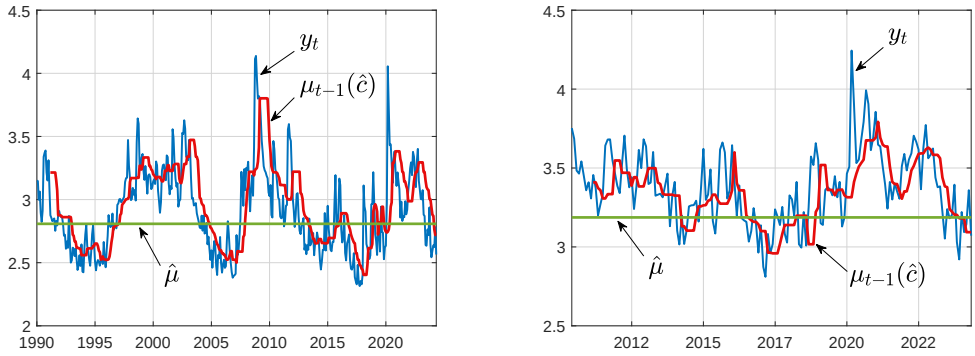
We plot the monthly log-VIX of SPX in Figure 1. SPX experienced two periods of excessive volatility: the global financial crisis (GFC) in 2007–2008 and the pandemic of novel coronavirus disease in 2020 (COVID-19). These two shocks had roughly equal impacts on the SPX volatility. The dot-com bubble and burst happened around 2000, and S&P lowered the credit rating of long-term U.S. government debt from AAA to AA+ in August 2011. These episodes also had substantial impacts on the SPX volatility, but their magnitudes are smaller than GFC and COVID-19.

Figure 1 also depicts the monthly log-VIX of the five individual shares. The dot-com bubble and GFC are not included in the sample period of the individual shares, as it begins in June 2010. Apple, Google, GS, and IBM experienced a surge in volatility during COVID-19. Interestingly, Amazon is less affected by COVID-19 than the other shares. The S&P downgrade of long-term U.S. government debt in 2011 seems to have had larger impacts on GS and IBM than on the other three shares.

In Table 3, we report sample statistics of the monthly log-VIX. The individual shares seem to be similar to each other in terms of mean, median, minimum, maximum, and standard deviation, but differ from each other in skewness and kurtosis. Amazon is the only share whose log-VIX is negatively skewed, with the skewness being -0.131. This fact is consistent with our previous observation that Amazon is least affect by COVID-19 of the five shares (Figure 1). GS and IBM have large positive skewness and kurtosis, reflecting their volatility spikes in response to the S&P downgrade of long-term U.S. government debt. For GS, the skewness is 1.072 and the kurtosis is 4.810. For IBM, the skewness is 0.952 and the kurtosis is 4.913.

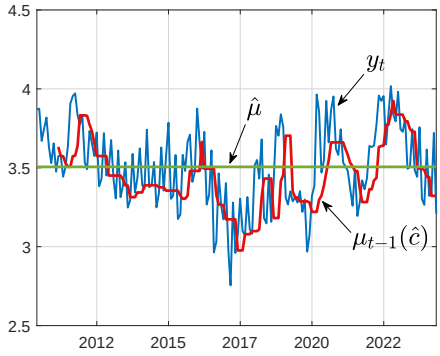
[6]The series ID is `VIXCLS` for SPX, `VXAPLCLS` for Apple, `VXAZNCLS` for Amazon, `VXGOGCLS` for Google, `VXGSCLS` for GS, and `VXIBMCLS` for IBM.

Figure 1: The monthly log-VIX and estimated conditional and constant thresholds

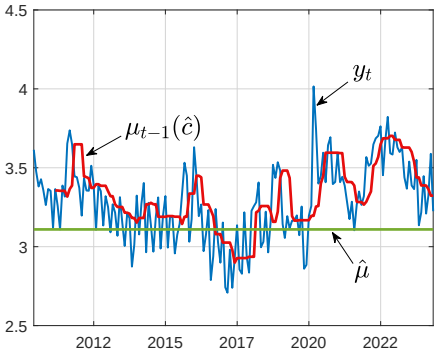


S&P 500

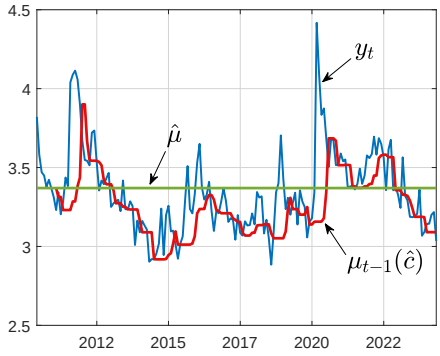
Apple



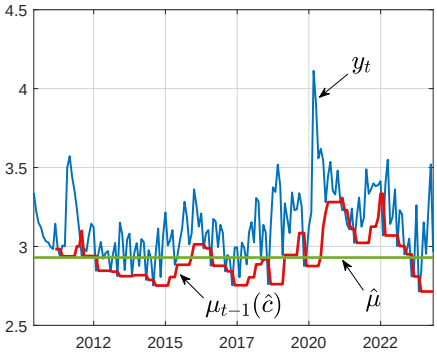
Amazon



Google



Goldman Sachs



IBM

y_t = the log-VIX of S&P 500 (SPX) and the five individual shares. $\mu_t(\hat{c})$ = 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. The sample period is January 1990 – May 2024 (413 months) for SPX and June 2010 – May 2024 (168 months) for the individual shares.

Table 3: Sample statistics of the monthly log-VIX of SPX and the individual shares

variable	n	mean	median	min	max	stdev	skew	kurt
S&P 500	413	2.912	2.873	2.315	4.138	0.335	0.642	3.358
Apple	168	3.375	3.368	2.811	4.244	0.235	0.339	3.375
Amazon	168	3.505	3.492	2.755	4.017	0.253	−0.131	2.580
Google	168	3.287	3.273	2.709	4.016	0.243	0.041	2.805
Goldman Sachs	168	3.350	3.312	2.885	4.417	0.259	1.072	4.810
IBM	168	3.129	3.101	2.714	4.113	0.221	0.952	4.913

In this table, we report summary statistics of the monthly log-VIX of S&P 500 (SPX) and the five individual shares. The sample period is January 1990 – May 2024 (413 months) for SPX and June 2010 – May 2024 (168 months) for the individual shares.

6.2 Testing the no-threshold-effect hypothesis

For each of the six target variables, we fit a SE-CoTAR model with lag length $p = 2$:

$$y_t = \begin{cases} \alpha_1 + \phi_{11} y_{t-1} + \phi_{12} y_{t-2} + u_t & \text{if } \; y_{t-d} < \mu_{t-d-1}(c), \\ \alpha_2 + \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + u_t & \text{if } \; y_{t-d} \geq \mu_{t-d-1}(c), \end{cases} \tag{23}$$

where y is the monthly log-VIX of SPX or an individual share. 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 $m = 12$ months for SPX and $m = 6$ months for the individual shares. We use the larger memory size for SPX than for the individual shares since SPX has the longer sample period. The choice space of the delay parameter d is $D = \{1, 2, 3\}$. The choice space of the percentile parameter c is given by (6) so that each regime accounts for at least 15% of the entire sample.

Based on model (23), 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, 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 5.1). To address potential heteroscedasticity, we use the robust covariance matrix estimator appearing in (16). 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 (23) with the constant threshold μ. The choice space for μ is given by (17). The remaining procedure is the same as SE-CoTAR.

Resulting bootstrap p-values are reported in Table 4. Focusing on SPX, H_0^* is rejected at the 5% level for SE-CoTAR with the p-value being 0.018, whereas H_0^* is not rejected at any conventional level for SETAR with the p-value being 0.317. This contrast highlights the validity of conditional threshold effects relative to constant threshold effects. The two models lead to the same conclusion for individual shares: H_0^* is rejected at the 10% level for Apple, Amazon, Google, and IBM, whereas H_0^* is not rejected at any conventional level for GS. These results suggest the presence of threshold effects, whether they are conditional or constant, in the log-VIX of all individual shares except for GS.

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 log-VIX sufficiently well, which is a key feature of the SE-CoTAR model. For IBM, regime 1 accounts for only about 15% of the whole sample for both models, almost hitting the prior restrictions (6) and (17). A correct interpretation of such a case is that regime 2 is a "normal" phase and regime 1 arises when the IBM volatility is exceptionally low, relative to either recent $m = 6$ months (SE-CoTAR) or the entire memory (SETAR).

6.3 Testing the equal predictive accuracy hypothesis

In this section, we conduct out-of-sample analyses in accordance with Section 3.3. 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 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 4. For SPX, RMSEs of the four models are as follows.

$$\begin{aligned} RMSE^{const} &= 0.334, \qquad RMSE^{ar} = 0.204, \\ RMSE^{setar} &= 0.210, \qquad RMSE^{secotar} = 0.203. \end{aligned} \tag{24}$$

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. $RMSE^{secotar}$

Table 4: Empirical results of the in-sample and out-of-sample analyses on the log-VIX of SPX and the individual shares

	SPX	Apple	Amazon	Google	GS	IBM
$\hat{p}^{secotar}(H_0^*)$	0.018**	0.064*	0.000***	0.016**	0.518	0.067*
$\hat{p}^{setar}(H_0^*)$	0.317	0.096*	0.000***	0.008***	0.938	0.049**
$RMSE^{const}$	0.334	0.212	0.278	0.294	0.189	0.221
$RMSE^{ar}$	0.204	0.149	0.200	0.174	0.128	0.210
$RMSE^{setar}$	0.210	0.148	0.171	0.157	0.133	0.225
$RMSE^{secotar}$	0.203	0.136	0.164	0.167	0.144	0.210
$\hat{p}^{ar}(H_1^{eq})$	0.800	0.052*	0.001***	0.280	0.047**	0.926
$\hat{p}^{ar}(H_1^{ar})$	0.600	0.974	1.000	0.860	0.023**	0.537
$\hat{p}^{ar}(H_1^{secotar})$	0.400	0.026**	0.000***	0.140	0.977	0.463
$\hat{p}^{setar}(H_1^{eq})$	0.163	0.077*	0.377	0.295	0.150	0.245
$\hat{p}^{setar}(H_1^{setar})$	0.918	0.962	0.812	0.147	0.075*	0.878
$\hat{p}^{setar}(H_1^{secotar})$	0.082*	0.038**	0.189	0.853	0.925	0.122

The target variable y is the monthly log-VIX of S&P 500 (SPX) and the five individual shares. The sample period is January 1990 – May 2024 (413 months) for SPX and June 2010 – May 2024 (168 months) for the individual shares. 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.

attains the minimum value out of four, highlighting 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 $\{0.800, 0.600, 0.400\}$ for $\{H_1^{eq}, H_1^{ar}, H_1^{secotar}\}$, respectively (Table 4). We do not find a significant evidence against H_0^{eq}, which is consistent with our previous observation from (24) that $RMSE^{ar}$ and $RMSE^{secotar}$ are almost equal to each other. We learn from this result that the in-sample and out-of-sample findings do not always agree with each other; significant conditional threshold effects are detected in the in-sample analysis, but significant improvement

in forecast accuracy is not detected in the out-of-sample analysis.

The p-value of the DM test between SETAR and SE-CoTAR is $\{0.163, 0.918, 0.082\}$ for $\{H_1^{eq}, H_1^{setar}, H_1^{secotar}\}$, respectively (Table 4). Hence, we reject H_0^{eq} in favor of $H_1^{secotar}$ at the 10% level, suggesting that the spread between $RMSE^{setar} = 0.210$ and $RMSE^{secotar} = 0.203$ is significant. This result is in line with the in-sample result that the no-threshold-effect hypothesis H_0^* is rejected for SE-CoTAR but not for SETAR. Overall, SE-CoTAR better fits and predicts the log-VIX of SPX than SETAR.

Focusing on the individual shares, the smallest RMSE is attained by the AR model for GS, by the SETAR model for Google, and by the SE-CoTAR model for Apple, Amazon, and IBM (Table 4). These results are consistent with our in-sample finding that H_0^* is rejected for all firms but for GS. For the non-GS firms, the improved RMSEs relative to AR suggest that the log-VIX can be better predicted by taking threshold effects into account. Conditional threshold effects seem more relevant than constant threshold effects, judging from the fact that RMSE is minimized by SE-CoTAR for three shares and by SETAR for one share (i.e., Google).

We observe the following results from the DM test for AR versus SE-CoTAR with nominal size 5%. First, H_0^{eq} is rejected in favor of $H_1^{secotar}$ for Apple and Amazon, confirming the usefulness of SE-CoTAR (Table 4). Second, H_0^{eq} is rejected in favor of H_1^{ar} for GS, which reconfirms the absence of threshold effects in GS. Third, H_0^{eq} is not rejected for Google and IBM, whichever the alternative hypothesis is. We did detect significant conditional threshold effects for Google and IBM in the in-sample analysis, but it does not necessarily deliver an incremental gain in out-of-sample prediction.

The following results are obtained from the DM test for SETAR versus SE-CoTAR with nominal size 5%. First, H_0^{eq} is rejected in favor of $H_1^{secotar}$ for Apple, suggesting that conditional threshold effects are more relevant than constant threshold effects (Table 4). Second, H_0^{eq} is not rejected for the other shares, whichever the alternative hypothesis is.[7] These results imply that SE-CoTAR is at least on par with SETAR for all shares, and significantly better for Apple, in terms of out-of-sample predictive ability. This implication is consistent with our simulation evidence that the relative superiority of SE-CoTAR under the truth of conditional threshold effects is larger than that of SETAR under the truth of constant threshold effects (Section 5.2).

In summary, the SE-CoTAR model achieves desired in-sample and out-of-sample performance for the log-VIX of SPX and the five major U.S. shares. This is a useful

[7]For GS, H_0^{eq} is rejected in favor of H_1^{setar} at the 10% level but not at the 5% level, with the p-value being 0.075.

finding which helps market participants better monitor the volatility of stock returns.[8]

7 Conclusion

There is a growing attention to time-varying threshold models in nonlinear time series analysis. Time-varying thresholds are compatible with a common perception in applied economics and finance that recent information matters more than past information. We have added to the literature by proposing the Conditional Threshold Autoregressive (CoTAR) model, in which the threshold is an empirical quantile of recent observations of x. The conditional threshold $\mu_t(c)$ represents a "normal" level of recent x if c is around 0.5 and an "abnormal" level if c is away from 0.5.

Statistical inference of CoTAR is analogous to TAR. First, we adopt the two-step method called profiling to estimate the regression parameters $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and the nuisance parameters $\boldsymbol{\gamma}$. Second, the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ can be tested with the wild-bootstrap tests of Hansen (1996). Third, we test the equal predictive accuracy hypothesis H_0^{eq} between CoTAR and a benchmark model (e.g., AR or SETAR) via the asymptotic test of Diebold and Mariano (1995).

Identification categories of $\boldsymbol{\gamma}$ play a key role in characterizing the asymptotic properties of our proposed methods. By considering the local alternative hypothesis $H_1^*(\boldsymbol{\lambda}_n) : \boldsymbol{\beta}_{2n} = \boldsymbol{\beta}_1 + n^{-1/2}\boldsymbol{\lambda}_n$, we demonstrate that the profiling estimator and the bootstrap tests for H_0^* satisfy desired asymptotic properties under standard regularity conditions. Further, our simulation study indicates that the proposed methods operate well in small samples.

We analyze the monthly log-VIX of SPX and the five major U.S. shares, using the SE-CoTAR model. The bootstrap test rejects the no-threshold-effect hypothesis H_0^* for all cases but for Goldman Sachs, indicating the relevance of conditional threshold effects. The SE-CoTAR model leads to the smaller mean squared forecast error than the constant-only, AR, and SETAR models for SPX, Apple, Amazon, and IBM. In particular, the superior out-of-sample performance of SE-CoTAR is statistically significant according to the Diebold-Mariano test for SPX, Apple, and Amazon. Overall, CoTAR is a promising model for tracing and forecasting VIX and other time series variables in which threshold effects are possible.

[8]In the supplemental material, we present another empirical application of SE-CoTAR. We analyze daily new confirmed COVID-19 cases of Japan and the U.S., finding the relevance of conditional threshold effects. Empirical results are especially striking for the U.S., in which the SE-CoTAR model attains significantly higher predictive power than the benchmark AR and SETAR models.

Acknowledgements

We thank Takero Doi, Yuzo Hosoya, Yasumasa Matsuda, seminar participants at IDA and Kobe University, and participants at the 3rd Hosoya Prize Lecture at Tohoku University, the 6th Annual International Conference on Applied Econometrics in Hawaii, EcoSta 2022, SETA 2022, and 2022 AMES in East and South-East Asia for helpful comments and discussions. The first author, Kaiji Motegi, is grateful for the financial support of Ishii Memorial Securities Research Promotion Foundation, Nomura Foundation, Japan Securities Scholarship Foundation, and JSPS KAKENHI Grant-in-Aid for Challenging Research (Exploratory) with Grant Number 23K17555.

References

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.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62(6), 1383–1414.

Chan, K. S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *The Annals of Statistics*, 21, 520–533.

Chan, K. S., and H. Tong (1985): "On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations," *Advances in Applied Probability*, 17, 666–678.

Chen, R., and R. S. Tsay (1991): "On the ergodicity of TAR(1) processes," *The Annals of Applied Probability*, 1, 613–634.

Davies, R. B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64(2), 247–254.

——— (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74(1), 33–43.

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

Dueker, M. J., Z. Psaradakis, M. Sola, and F. Spagnolo (2013): "State-Dependent Threshold Smooth Transition Autoregressive Models," *Oxford Bulletin of Economics and Statistics*, 75, 835–854.

Fernandes, M., M. C. Medeiros, and M. Scharth (2014): "Modeling and predicting the CBOE market volatility index," *Journal of Banking & Finance*, 40, 1–10.

Giné, E., and J. Zinn (1990): "Bootstrapping general empirical measures," *The Annals of Probability*, 18, 851–869.

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

——— (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603.

——— (2011): "Threshold autoregression in economics," *Statistics and Its Interface*, 4, 123–127.

——— (2017): "Regression Kink With an Unknown Threshold," *Journal of Business & Economic Statistics*, 35, 228–240.

Liu, Q., S. Guo, and G. Qiao (2015): "VIX forecasting and variance risk premium: A new GARCH approach," *North American Journal of Economics and Finance*, 34, 314–322.

Motegi, K., X. Cai, S. Hamori, and H. Xu (2020): "Moving average threshold heterogeneous autoregressive (MAT-HAR) models," *Journal of Forecasting*, 39, 1035–1042.

Psaradellis, I., and G. Sermpinis (2016): "Modelling and trading the U.S. implied volatility indices: Evidence from the VIX, VXN and VXD indices," *International Journal of Forecasting*, 32, 1268–1283.

Qiao, G., J. Yang, and W. Li (2020): "VIX forecasting based on GARCH-type model with observable dynamic jumps: A new perspective," *North American Journal of Economics and Finance*, 53, #101186.

Seo, M. H., and O. Linton (2007): "A smoothed least squares estimator for threshold regression models," *Journal of Econometrics*, 141, 704–735.

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

——— (2011): "Threshold models in time series analysis — 30 years on," *Statistics and Its Interface*, 4, 107–118.

——— (2015): "Threshold models in time series analysis—Some reflections," *Journal of Econometrics*, 189, 485–491.

Whaley, R. E. (2009): "Understanding the VIX," *Journal of Portfolio Management*, 35, 98–105.

Yang, L., C. Lee, and I. Chen (2021): "Threshold model with a time-varying threshold based on Fourier approximation," *Journal of Time Series Analysis*, 42, 406–430.

Yang, L., and J.-J. Su (2018): "Debt and growth: Is there a constant tipping point?," *Journal of International Money and Finance*, 87, 133–143.

Yu, P., and X. Fan (2021): "Threshold Regression With a Threshold Boundary," *Journal of Business & Economic Statistics*, 39, 953–971.

Zhu, Y., H. Chen, and M. Lin (2019): "Threshold models with time-varying threshold values and their application in estimating regime-sensitive Taylor rules," *Studies in Nonlinear Dynamics & Econometrics*, 23, #20170114.

Appendices

We prove Theorems 1-2 in this appendix. For arbitrary $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \Gamma$, define some matrices conditional on the sample:

$$\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \boldsymbol{M}_n(\boldsymbol{\gamma}_1)^{-1} \hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \boldsymbol{M}_n(\boldsymbol{\gamma}_2)^{-1},$$

$$\hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{1}{n} \sum_{t=1}^{n} \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma}_1) \hat{\boldsymbol{s}}_t(\boldsymbol{\gamma}_2)^\top, \qquad \boldsymbol{M}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}_1) \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}_2)^\top,$$

where $\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})$ is the vector of regressors defined in (4); $\hat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) \hat{u}_t(\boldsymbol{\gamma})$ is the estimated regression score under $H_1^* : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ conditional on $\boldsymbol{\gamma}$, as defined in (10). We sometimes abbreviate $\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}, \boldsymbol{\gamma})$, $\hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}) = \hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}, \boldsymbol{\gamma})$, and $\boldsymbol{M}_n(\boldsymbol{\gamma}) = \boldsymbol{M}_n(\boldsymbol{\gamma}, \boldsymbol{\gamma})$ when appropriate, recovering (11)-(12). The population versions of these matrices, denoted as $\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, $\boldsymbol{S}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, and $\boldsymbol{M}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, are defined in (18). Recall that $\mathcal{G}(\boldsymbol{\gamma})$ is a mean zero Gaussian process with covariance kernel $\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. Finally, recall from (9) that the regression score conditional on $\boldsymbol{\gamma}$ is written as $\boldsymbol{s}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_{t-1}(\boldsymbol{\gamma}) u_t$.

The following lemma is useful for proving Theorems 1-2.

Lemma A.1. *Impose Assumption 1, then the following are true. (i)* $\hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \xrightarrow{p} \boldsymbol{S}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ *as* $n \to \infty$ *uniformly over* $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \Gamma$. *(ii)* $\boldsymbol{M}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \xrightarrow{p} \boldsymbol{M}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ *uniformly over* $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \Gamma$. *(iii)* $\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \xrightarrow{p} \boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ *uniformly over* $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \Gamma$. *(iv)* $n^{-1/2}\, \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1} \sum_{t=1}^{n} \boldsymbol{s}_t(\boldsymbol{\gamma}) \Rightarrow \mathcal{G}(\boldsymbol{\gamma})$.

Proof. A proof follows directly from Assumption 1 by application of the law of large numbers and the central limit theorem paired with the finite support of Γ.[9] □

[9] Assumption 1 implies the uniform convergence even if the support of Γ is infinite (e.g., Hansen, 1996, Theorems 1 and 3). The uniform convergence is required for our Theorems 1-2 to hold.

A.1 Proof of Theorem 1

Recall that the CoTAR model is given in (5) and the conditional least squares estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ is given in (7). Since we are considering local asymptotics as in (20), the regression parameter $\boldsymbol{\beta}$ in model (5) should be replaced with $\boldsymbol{\beta}_n$. We have that

$$\begin{aligned}
\sqrt{n}\left\{\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})-\boldsymbol{\beta}_n\right\} &= \sqrt{n}\left[\frac{1}{n}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\sum_{t=1}^{n}\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})\left\{\boldsymbol{Z}_{t-1}(\boldsymbol{\gamma})^{\top}\boldsymbol{\beta}_n+u_t\right\}-\boldsymbol{\beta}_n\right] \\
&= \sqrt{n}\left\{\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\boldsymbol{M}_n(\boldsymbol{\gamma})\boldsymbol{\beta}_n+\frac{1}{n}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\sum_{t=1}^{n}\boldsymbol{s}_t(\boldsymbol{\gamma})-\boldsymbol{\beta}_n\right\} \\
&= \frac{1}{\sqrt{n}}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\sum_{t=1}^{n}\boldsymbol{s}_t(\boldsymbol{\gamma}) \\
&\Rightarrow \mathcal{G}(\boldsymbol{\gamma}). \qquad (\because \text{ Lemma A.1.}(iv)) \qquad \text{(A.1)}
\end{aligned}$$

This weak convergence implies $\sup_{\boldsymbol{\gamma}\in\Gamma}||\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})-\boldsymbol{\beta}_n|| \xrightarrow{p} 0$, establishing Theorem 1.(i). Substitute $H_1^*(\boldsymbol{\lambda}_n):\boldsymbol{\beta}_n=\bar{\boldsymbol{\beta}}_1+n^{-1/2}\bar{\boldsymbol{\lambda}}_n$ into (A.1) to get

$$\sqrt{n}\left\{\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})-\bar{\boldsymbol{\beta}}_1\right\}=\bar{\boldsymbol{\lambda}}_n+\frac{1}{\sqrt{n}}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\sum_{t=1}^{n}\boldsymbol{s}_t(\boldsymbol{\gamma}). \qquad \text{(A.2)}$$

Equation (A.2) implies that, if $\lim_{n\to\infty}\|\boldsymbol{\lambda}_n\|<\infty$, then $\sqrt{n}\{\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})-\bar{\boldsymbol{\beta}}_1\}\Rightarrow\bar{\boldsymbol{\lambda}}+\mathcal{G}(\boldsymbol{\gamma})$. Thus, Theorem 1.$(ii)$ has been established.

A.2 Proof of Theorem 2

We focus on the Wald tests in this proof, as the LM tests are analogous. Define:

$$\mathcal{W}^{\lambda}(\boldsymbol{\gamma})=\left\{\mathcal{G}(\boldsymbol{\gamma})+\boldsymbol{Q}(\boldsymbol{\gamma})\bar{\boldsymbol{\lambda}}_n\right\}^{\top}\boldsymbol{R}^{\top}\{\boldsymbol{R}\boldsymbol{V}(\boldsymbol{\gamma})\boldsymbol{R}^{\top}\}^{-1}\boldsymbol{R}\left\{\mathcal{G}(\boldsymbol{\gamma})+\boldsymbol{Q}(\boldsymbol{\gamma})\bar{\boldsymbol{\lambda}}_n\right\},$$

where $\boldsymbol{R}$ is the selection matrix associated with $H_0^*:\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$, as defined in (8); $\boldsymbol{Q}(\boldsymbol{\gamma})=\boldsymbol{M}(\boldsymbol{\gamma})^{-1}\boldsymbol{M}(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)$; $\boldsymbol{M}(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)=\mathrm{E}\{\boldsymbol{Z}_t(\boldsymbol{\gamma}_1)\boldsymbol{Z}_t(\boldsymbol{\gamma}_2)^{\top}\}$; $\boldsymbol{\gamma}_0$ is the fixed true value of $\boldsymbol{\gamma}$ when $\boldsymbol{\lambda}_n\neq\mathbf{0}$. Incorporate all possible values of $\boldsymbol{\gamma}$ with

$$\begin{aligned}
&\sup\mathcal{W}^{\lambda}=\sup_{\boldsymbol{\gamma}\in\Gamma}\mathcal{W}^{\lambda}(\boldsymbol{\gamma}), \qquad \mathrm{ave}\mathcal{W}^{\lambda}=\int_{\Gamma}\mathcal{W}^{\lambda}(\boldsymbol{\gamma})d\mu^*(\boldsymbol{\gamma}), \qquad \text{(A.3)} \\
&\exp\mathcal{W}^{\lambda}=\ln\left[\int_{\Gamma}\exp\left\{\frac{\mathcal{W}^{\lambda}(\boldsymbol{\gamma})}{2}\right\}d\mu^*(\boldsymbol{\gamma})\right],
\end{aligned}$$

where some subset of Γ has positive measure with respect to μ^* (e.g., Davies, 1977, 1987, Andrews and Ploberger, 1994). Equation (A.3) is an asymptotic counterpart to (13).

Let $g^\lambda = g(\mathcal{W}^\lambda)$ denote either $\sup\mathcal{W}^\lambda$, $\text{ave}\mathcal{W}^\lambda$, or $\exp\mathcal{W}^\lambda$ as appropriate. Observe that $g(\cdot)$ is a continuous functional of the Gaussian process $\mathcal{G}(\boldsymbol{\gamma})$. Application of Theorems 1 and 3 of Hansen (1996) implies that $\mathcal{W}_n \Rightarrow \mathcal{W}^\lambda$ and $g(\mathcal{W}_n) \Rightarrow g^\lambda$; the asymptotic distribution of $\mathcal{W}_n$ under H_0^* is $\mathcal{W}^0(\boldsymbol{\gamma}) = \mathcal{G}(\boldsymbol{\gamma})^\top \boldsymbol{R}^\top \{\boldsymbol{R}\boldsymbol{V}(\boldsymbol{\gamma})\boldsymbol{R}^\top\}^{-1}\boldsymbol{R}\mathcal{G}(\boldsymbol{\gamma})$; the asymptotic distribution of $\mathcal{W}_n$ under $H_1^*(\boldsymbol{\lambda}_n)$ is a noncentral chi-squared process on Γ. Let $F^0(\cdot)$ denote the distribution function of $g^0 = g(\mathcal{W}^0)$, and define $p_n = 1 - F^0(g_n)$. By monotonicity and continuity of F^0, $p_n \Rightarrow p^\lambda = 1 - F^0(g^\lambda)$; hence, the asymptotic distribution of p_n under H_0^* is Uniform on $[0,1]$.

Let $\{\xi_t\}_{t=1}^n$ be iid standard normal random variables. Define:

$$\widehat{\mathcal{W}}_n(\boldsymbol{\gamma}) = \hat{\boldsymbol{v}}_n(\boldsymbol{\gamma})^\top \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\boldsymbol{R}^\top \left\{\boldsymbol{R}\hat{\boldsymbol{V}}_n(\boldsymbol{\gamma})\boldsymbol{R}^\top\right\}^{-1} \boldsymbol{R}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\hat{\boldsymbol{v}}_n(\boldsymbol{\gamma}),$$

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

$\hat{\boldsymbol{v}}_n(\boldsymbol{\gamma})$ is a mean zero Gaussian process with covariance kernel $\hat{\boldsymbol{S}}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ conditional on the sample. Let $\hat{F}_n$ denote the distribution function of $\hat{g}_n = g(\widehat{\mathcal{W}}_n)$ conditional on the sample. Let $\hat{p}_n = 1 - \hat{F}_n(g_n)$, then Theorem 2 of Hansen (1996) implies that $\hat{p}_n \Rightarrow_p p^\lambda$, where $\Rightarrow_p$ signifies *weak convergence in probability* of Giné and Zinn (1990).

To prove Theorem 2.(i), impose H_0^*. Under H_0^*, the asymptotic distribution of $\hat{p}_n$ is Uniform on $[0,1]$. Application of the Glivenko-Cantelli Theorem implies that $\hat{p}_n^B(H_0^*) - \hat{p}_n \xrightarrow{p} 0$ as $n \to \infty$ and $B \to \infty$, establishing Theorem 2.(i).

To prove Theorem 2.(ii), impose $H_1^*(\boldsymbol{\lambda}_n)$ with $\|\boldsymbol{\lambda}_n\| \to \infty$. We have that $\mathcal{W}_n \Rightarrow \mathcal{W}^\lambda$, a noncentral chi-squared process on Γ, and that $\hat{p}_n \Rightarrow_p p^\lambda$. Hence, it is sufficient to show that $g^\lambda = g(\mathcal{W}^\lambda)$ diverges almost surely as $\|\boldsymbol{\lambda}_n\| \to \infty$. Recall that $g(\mathcal{W}_n) \Rightarrow g^\lambda$. By Assumption 1.$(ii)$-$(iii)$, $\inf_{\boldsymbol{\gamma}\in\Gamma}\det\{\boldsymbol{M}(\boldsymbol{\gamma})\} > 0$ and $\inf_{\boldsymbol{\gamma}\in\Gamma}\det\{\boldsymbol{V}(\boldsymbol{\gamma})\} > 0$. Hence,

$$\left\{\boldsymbol{R}\boldsymbol{Q}(\boldsymbol{\gamma})\bar{\boldsymbol{\lambda}}_n\right\}^\top \{\boldsymbol{R}\boldsymbol{V}(\boldsymbol{\gamma})\boldsymbol{R}^\top\}^{-1}\boldsymbol{R}\boldsymbol{Q}(\boldsymbol{\gamma})\bar{\boldsymbol{\lambda}}_n \to \infty$$

almost surely as $\|\boldsymbol{\lambda}_n\| \to \infty$ for every $\boldsymbol{\gamma} \in \Gamma$. Hence, $g^\lambda \to \infty$ almost surely as $\|\boldsymbol{\lambda}_n\| \to \infty$, establishing Theorem 2.(ii).