The conditional threshold autoregression (CoTAR)

Kaiji Motegi\textsuperscript{1}  John W. Dennis\textsuperscript{2}  Shigeyuki Hamori\textsuperscript{1}

\textsuperscript{1}Kobe University  \textsuperscript{2}IDA

The 6th Hawaii Conference
Zoom Webinar
November 9, 2021
Introduction

- A time series often has heterogeneous properties below versus above a certain threshold (threshold effects).
- One of the most well-known models in this field is the threshold autoregression (TAR) proposed by Tong (1978).
- In TAR, a target series $y$ follows AR($p$) with coefficients being different across regimes, and a regime switch is triggered when a threshold variable $x$ crosses a constant threshold parameter $\mu$.
- Constant-threshold models like TAR have been extended in many ways so that thresholds are time-varying or state-dependent.
We propose the **conditional threshold autoregression (CoTAR)**, where the threshold $\mu_t$ is specified as an empirical quantile of recent observations of the threshold variable $x$.

The proposed conditional threshold $\mu_t$ traces the fluctuation of $x_t$, which can enhance the fit and interpretation of the model.

In CoTAR, the existence of threshold effects can be tested by Hansen’s (1996) wild-bootstrap tests.

The estimation and hypothesis testing of CoTAR satisfy desired statistical properties in both large and small samples.

We fit CoTAR to daily new confirmed COVID-19 cases in the U.S. and Japan.

**Significant** conditional threshold effects are detected for both U.S. and Japan, indicating the practical use of CoTAR.
Table of contents

1. Introduction

2. Conditional threshold autoregression (CoTAR)
   - Motivation and specification
   - Profiling estimation of parameters
   - Bootstrap test for the no-threshold-effect hypothesis

3. Empirical application on COVID-19 cases

4. Conclusion
Let \( \{y_t\}_{t=1}^n \) be a target variable; let \( \{x_t\}_{t=1}^n \) be a threshold variable.

Consider Tong’s (1978) threshold autoregression (TAR):

\[
y_t = \begin{cases} 
\alpha_1 + \sum_{k=1}^{p} \phi_1 y_{t-k} + u_t & \text{if } x_{t-d} < \mu, \\
\alpha_2 + \sum_{k=1}^{p} \phi_2 y_{t-k} + u_t & \text{if } x_{t-d} \geq \mu.
\end{cases}
\]

\( y \) has different autocorrelation structures below vs. above the unconditional threshold \( \mu \).

“Unconditional” means that \( \mu \) is time-independent and chosen from the entire memory of \( x \):

\[
\mathcal{X}_1^n = \{x_1, \ldots, x_n\}.
\]
CoTAR: Motivation and specification

- We propose to replace $\mu$ with a conditional threshold $\mu_t$.
- $\mu_t$ is time-dependent and chosen from a local memory of size $m$:

$$\chi_{t-m+1}^t = \{x_{t-m+1}, \ldots, x_t\}.$$  

- We propose the conditional threshold autoregression (CoTAR):

$$y_t = \begin{cases} 
\alpha_1 + \sum_{k=1}^{p} \phi_1ky_{t-k} + u_t & \text{if } x_{t-d} < \mu_{t-d-1}(c), \\
\alpha_2 + \sum_{k=1}^{p} \phi_2ky_{t-k} + u_t & \text{if } x_{t-d} \geq \mu_{t-d-1}(c). 
\end{cases}$$

- $\mu_t(c)$ is the $mc$-th smallest value (the $100c\%$ point) of $\chi_{t-m+1}^t$.
- $c \in \{1/m, 2/m, \ldots, 1\}$ signifies the relevant percentile.
- When $x_t = y_t$, we have the self-exciting CoTAR (SE-CoTAR).
If $c = \frac{(m + 1)}{(2m)}$, then $\mu_{t-d-1}(c)$ (almost) coincides with the median of $X_{t-d-m}^{t-d-1}$; in this case, a regime switch is triggered by an event of $x$ crossing a “normal” level given the local memory.

If $c$ is close to the lower bound $\frac{1}{m}$ 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.

**Example:** SE-CoTAR with $m = 14$ and $c = \frac{11}{14} = 0.786$ fitted to daily changes in new confirmed COVID-19 cases.

In this example, $\mu_t$ is the 78.6% point of the recent 14-day observations of the changes in new cases; hence, regime 2 represents an extremely serious phase of the pandemic.

The conditional threshold approach is intuitively reasonable, since individuals seem to evaluate the current status of the pandemic relative to the recent past, not to a constant cut-off.
CoTAR: Matrix representation

- Stack the regression parameters for each regime:
  \[ \beta_r = (\alpha_r, \phi_{r1}, \ldots, \phi_{rp})^\top, \quad r \in \{1, 2\}. \]

- Define:
  \[ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \gamma = \begin{bmatrix} d \\ c \end{bmatrix}. \]

- Define binary variables which determine the regime:
  \[ I_{1t}(c) = 1 \{x_t < \mu_{t-1}(c)\}, \quad I_{2t}(c) = 1 \{x_t \geq \mu_{t-1}(c)\}. \]

- Stack the regressors:
  \[ z_{t-1} = (1, y_{t-1}, \ldots, y_{t-p})^\top, \quad Z_{t-1}(\gamma) = \begin{bmatrix} z_{t-1} I_{1,t-d}(c) \\ z_{t-1} I_{2,t-d}(c) \end{bmatrix}. \]

- CoTAR is rewritten in matrix form as:
  \[ y_t = Z_{t-1}(\gamma)^\top \beta + u_t. \]
Profiling estimation

- The least squares estimator for the regression parameter $\beta$ conditional on the nuisance parameter $\gamma$ is given by:

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

- The profiling estimator for $\gamma$ is given by:

$$
\hat{\gamma} = \arg\min_{\gamma \in \Gamma} \sum_{t=1}^{n} \left\{ y_t - Z_{t-1}(\gamma)^\top \hat{\beta}(\gamma) \right\}^2.
$$

- The profiling estimator for $\beta$ is given by $\hat{\beta} = \beta(\hat{\gamma})$. 
Asymptotic properties of the profiling estimator depends crucially on whether conditional threshold effects are present or absent.

Conditional threshold effects are **present** if $\beta_1 \neq \beta_2$, in which case $\gamma$ is **identifiable**.

Conditional threshold effects are **absent** if $\beta_1 = \beta_2$, in which case CoTAR reduces to the single-regime AR($p$) and $\gamma$ is **unidentifiable**.

Define the no-threshold-effect hypothesis:

$$H_0^* : \beta_1 = \beta_2 \quad \text{vs.} \quad H_1^* : \beta_1 \neq \beta_2.$$
Theorem 1 (Profiling estimator)

Under standard regularity conditions, the following are true:

1. \( \sqrt{n}\{\hat{\beta}(\gamma) - \beta_0\} \Rightarrow \mathcal{N}\{0, V(\gamma)\} \) for each fixed \( \gamma \in \Gamma \).

2. \( \hat{\beta}(\gamma) \xrightarrow{p} \beta_0 \) uniformly over \( \gamma \in \Gamma \).

3. Under \( H_1^* \), \( \hat{\gamma} - \gamma_0 = O_p(n^{-1}) \) and
   \[ \sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}\{0, V(\gamma_0)\}. \]

- See the full paper for the regularity conditions, the construction of \( V(\gamma) \), and the proof of Theorem 1.

- Under \( H_0^* \), the asymptotic distribution of \( \hat{\beta} \) is non-standard.
Testing the no-threshold-effect hypothesis

Formulate the no-threshold-effect hypothesis $H_0^*$ as a linear parametric restriction:

$$H_0^* : \mathbf{R}^* \mathbf{\beta} = q^* \quad \text{vs.} \quad H_1^* : \mathbf{R}^* \mathbf{\beta} \neq q^*.$$ 

where

$$\mathbf{R}^* = (I_{p+1}, -I_{p+1}), \quad q^* = 0_{(p+1) \times 1}.$$ 

Testing $H_0^*$ requires the wild bootstrap of Hansen (1996), since $\gamma$ is unidentified and $\hat{\mathbf{\beta}}$ is not asymptotically normal under $H_0^*$.

The Wald test statistic conditional on $\gamma$ is given by:

$$W_n^*(\gamma) = n \left\{ \mathbf{R}^* \hat{\mathbf{\beta}}(\gamma) - q^* \right\}^\top \left\{ \mathbf{R}^* \hat{\mathbf{V}}_n(\gamma)(\mathbf{R})^\top \right\}^{-1} \left\{ \mathbf{R}^* \hat{\mathbf{\beta}}(\gamma) - q^* \right\}.$$ 

See the full paper for the construction of $\hat{\mathbf{V}}_n(\gamma)$. 
Incorporate all possible values of $\gamma$ as in:

$$\sup_{\gamma \in \Gamma} \mathcal{W}_n^* = \sup_{\gamma \in \Gamma} \mathcal{W}_n^*(\gamma).$$

Let $g(\mathcal{W}_n^*)$ be either $\sup \mathcal{W}_n^*$, $\text{ave} \mathcal{W}_n^*$, or $\exp \mathcal{W}_n^*$.

Let $\{g\{\mathcal{W}_n^*\}^{(b)}\}_{b=1}^B$ be the set of wild-bootstrap test statistics. (See the full paper for the bootstrap procedure.)

The bootstrap p-value is defined as:

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

Reject $H_0^*$ if $\hat{p}_n^B(H_0^*) < a$, where $a \in (0, 1)$ is the nominal size.
Testing the no-threshold-effect hypothesis

Theorem 2 (Bootstrap test for $H_0^*$)

Under standard regularity conditions, the following are true:

1. Under $H_0^*$, $\hat{p}_n^B(H_0^*)$ is asymptotically uniform on $[0, 1]$.
2. Under $H_1^*$, $\hat{p}_n^B(H_0^*) \xrightarrow{p} 0$ as $n \to \infty$ and $B \to \infty$.

See the full paper for the regularity conditions and the proof.

The bootstrap test for $H_0^*$ is asymptotically valid; the test has size approaching the nominal size $a$ under $H_0^*$, and power approaching 1 under $H_1^*$. 
Empirical application: Set-up

- We analyze the number of daily new confirmed cases per million people in the U.S. and Japan, denoted as \( \{w_t\}_{t=1}^n \).

- Sample period: April 4, 2020 – June 23, 2021 (\( n = 446 \) days).

- We fit the SE-CoTAR model with \( p = 3 \) and \( m = 14 \) to

\[
y_t = \Delta \ln w_t = \ln w_t - \ln w_{t-1} \quad \text{(i.e., the log-difference of the number of daily new confirmed cases per million people)}:
\]

\[
y_t = \begin{cases} 
\alpha_1 + \sum_{k=1}^{3} \phi_1 k y_{t-k} + u_t & \text{if } y_{t-d} < \mu_{t-d-1}(c), \\
\alpha_2 + \sum_{k=1}^{3} \phi_2 k y_{t-k} + u_t & \text{if } y_{t-d} \geq \mu_{t-d-1}(c). 
\end{cases}
\]

- Regime 1 represents a **deceleration** phase where the change in new confirmed cases is small relative to the local memory.

- Regime 2 represents an **acceleration** phase where the change is relatively large.
The log series move smoothly due to a seasonal adjustment.

Several waves of the pandemic are observed clearly.

The number of new confirmed cases per million people in the U.S. is larger than that in Japan.

\[ w_n = 34.1 \] for the U.S. and \[ w_n = 11.4 \] for Japan, almost triple.
The log-difference series is the target of the SE-CoTAR model.
The log-difference series exhibits rather complex fluctuations with persistent swings and temporary noise being combined, which suggests the presence of nonlinear effects.
The log-difference of Japan is more volatile than that of the U.S.
Empirical application: Results

The estimated conditional threshold $\mu_t(\hat{c})$ traces the persistent swing of $y_t$ strikingly well, highlighting the feature of CoTAR.

The p-value of the bootstrap test for $H_0^*$ is 0.049 for the U.S. and 0.002 for Japan.

Hence, we conclude at the 5% level that conditional threshold effects are present for both countries.
Conclusion

- We have proposed the conditional threshold autoregression (CoTAR), where the threshold is specified as an empirical quantile of the local memory of a threshold variable $x$.

- The resulting conditional threshold traces the fluctuation of $x$, which can enhance the fit and interpretation of the model.

- The parameters of CoTAR can be estimated via profiling.

- The bootstrap test for the no-threshold-effect hypothesis $H_0^*$ is asymptotically valid.

- We have analyzed the daily new confirmed COVID-19 cases per million people in the U.S. and Japan by fitting SE-CoTAR.

- Significant conditional threshold effects have been detected for both countries, indicating the practical use of CoTAR.
References


