Regular and reverse Midastar models: Threshold autoregression with mixed frequency data*

Kaiji Motegi† John W. Dennis‡ Seok Young Hong§

December 15, 2025

Abstract

We propose Midastar, a novel extension of the threshold autoregression (TAR) to the Mixed Data Sampling (MIDAS) framework. In the regular Midastar, the target variable is observed less frequently than the threshold variable. In the reverse Midastar, the target variable is observed more frequently. These models accurately capture threshold effects, whereas standard TAR with temporally aggregated data can point to spurious non-threshold effects. The parameters are estimated via profiling, and the no-threshold-effect hypothesis is tested via a wild bootstrap. We establish the uniform consistency and asymptotic normality under much weaker conditions than in the literature. In particular, we overcome the challenges arising from the potential lack of stationarity. Monte Carlo simulations and empirical applications demonstrate that the Midastar models are useful for modelling and predicting macroeconomic and financial indicators.

JEL codes: C22, C51, C58.

Keywords: Mixed Frequency Data, Nonlinear Time Series, Threshold Autoregression.

*We thank Luiggi Donayre, Eric Ghysels, Shigeyuki Hamori, Takamitsu Kurita, Daisuke Nagakura, Tatsuyoshi Okimoto, seminar participants at Keio University, Kobe University, Nanyang Technological University, Tokyo Metropolitan University, and conference participants at the 7th Kobe-Hawaii Conference and the WEAI Conferences for helpful comments and discussions. The first author is grateful for the financial support of JSPS KAKENHI Grant-in-Aid for Challenging Research (Exploratory) with Grant Number 23K17555, Ishii Memorial Securities Research Promotion Foundation, and Japan Securities Scholarship Foundation.

†*Corresponding author*. Kobe University, Japan. Email: motegi@econ.kobe-u.ac.jp

‡Institute for Defense Analyses (IDA), USA. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of IDA. Email: jay.dennis@alumni.unc.edu

§Nanyang Technological University, Singapore. Email: seokyoung.hong@ntu.edu.sg

1 Introduction

Time series variables often have threshold effects, a nonlinear phenomenon characterized by heterogeneous properties below versus above a certain threshold. While there is a vast literature on modelling threshold effects, the threshold autoregression (TAR) proposed by Tong (1978) is one of the earliest and most well-known models in this field. In the TAR model, a target series y follows an AR(p) process with coefficients differing across regimes, and a regime switch is triggered when a threshold variable x crosses a threshold parameter μ.[1]

The existing TAR literature is restricted to the single-frequency framework in which y and x are assumed to be sampled at the same frequency. In empirical studies, one of y and x may well be sampled at a higher frequency than the other. Aggregating the series of the higher sampling frequency into the lower level often has an adverse effect on statistical inference due to loss of information. In particular, it is well known that temporal aggregation tends to weaken nonlinearities in original series (e.g., Brännäs and Ohlsson, 1999, Granger and Lee, 1999, Paya and Peel, 2006). Hence, fitting the TAR model to aggregated data may result in spurious non-threshold effects: a failure to detect threshold effects in the underlying process.[2]

Numerous situations in economics and finance involve variables observed at different frequencies, underscoring the importance of mixed-frequency models in conducting valid statistical analyses (see, e.g., Marcellino and Sivec, 2016, Schorfheide, Song, and Yaron, 2018, Andreou, Gagliardini, Ghysels, and Rubin, 2019). This is particularly crucial in today's data-rich environment, where even ultra-high-frequency data are widely available. For instance, Li, Todorov, Tauchen, and Chen (2017) study a mixed-scale jump regression framework in which explanatory variables are sampled at an extremely fine time scale compared to the potentially less liquid dependent variables.

In the Mixed Data Sampling (MIDAS) framework that originated with Ghysels, Santa-Clara, and Valkanov (2004), a variety of methods are proposed to exploit all data available at mixed frequencies, avoiding temporal aggregation and its shortcomings. Andreou, Gagliardini, and Ghysels (2010) discuss regression models with mixed frequency data, Guay and Maurin (2015) propose a flexible disaggregation procedure, Ghysels (2016) establishes mixed frequency vector autoregressive models, and Foroni,

[1] Hansen (2011), Tong (2015), and Tsay and Chen (2019) provide extensive surveys on TAR.

[2] Silvestrini and Veredas (2008) perform a comprehensive survey on the consequences of temporal aggregation in time series models.

Marcellino, and Stevanović (2019) study mixed frequency ARMA models. Ample evidence demonstrates that the MIDAS approach improves the in-sample fit and the out-of-sample prediction compared with the single-frequency approach.[3]

Extensions of the TAR model to mixed frequency data has not been pursued in the literature, notwithstanding its considerable potential applications. To fill this research gap, we propose *Midastar* models that capture threshold effects when variables are observed at different frequencies. This paper studies both (i) *regular Midastar*, where the target variable y is observed less frequently than the threshold variable x, and (ii) *reverse Midastar*, where y is observed more frequently than x. Both cases commonly arise in various practical contexts within economics and finance. By construction, y is not directly regressed onto x: the latter merely determines the regime. Because of this indirect relationship, the Midastar models are free of parameter proliferation when the ratio of sampling frequencies between y and x takes a large value. This is a notable advantage compared with existing MIDAS regression models.

The regular Midastar is characterized by the regression parameters $\boldsymbol{\beta}$ (i.e., intercepts and AR parameters) and the nuisance parameters $\boldsymbol{\gamma}$ (i.e., the delay and threshold parameters). The former is partitioned by regimes: $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$. The nuisance parameters $\boldsymbol{\gamma}$ are unidentified *if and only if* there are no threshold effects (i.e., $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$). When threshold effects are absent, the threshold variable does not play any role in the underlying process, hence the regular Midastar and the TAR with aggregated threshold variable are equivalent to each other. When threshold effects are present (i.e., $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$), The regular Midastar is able to detect the threshold effects at its best possible precision, while TAR poses a risk of pointing to spurious non-threshold effects. The consequence of temporal aggregation can be even more serious for the reverse scenario where the target variable is aggregated, since it can lead to misspecification regardless of whether threshold effects are absent or present.

We propose to estimate $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ via profiling, a two-step procedure with a long history in statistics (Barndorff-Nielsen, 1983). For testing the null hypothesis of no threshold effects (i.e., $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$), we adopt the wild-bootstrap tests of Hansen (1996) to account for the unidentifiability of $\boldsymbol{\gamma}$ under H_0^*, analogous to those in the standard TAR literature. We establish the uniform consistency and asymptotic normality of the proposed estimator as well as the asymptotic validity of the bootstrap test for both regular and reverse Midastar models. In particular, we identify the inherent issue of lack of stationarity in the reverse Midastar, and construct the limit-

[3]See Ghysels and Marcellino (2018, Part IV) for an extensive review of the MIDAS literature.

ing theory that addresses these complications. Overall, our regularity conditions are considerably weaker than those in the literature.

Monte Carlo simulations demonstrate that our proposed methods perform well in finite samples. We present two separate empirical applications to illustrate the practical values of the regular and reverse Midastar models. For the regular scenario, the target variable is monthly realized volatility measures of the crude oil market and the threshold variable is daily CBOE Volatility Index. For the reverse scenario, the target variable is monthly employment growth and the threshold variable is quarterly real GDP growth of the United States. Both studies indicate that the Midastar models are useful for modelling and predicting macroeconomic and financial indicators, whereas the single-frequency TAR model points to spurious non-threshold effects.

The remainder of the paper is organized as follows. We specify the Midastar models in Section 2, describe the procedures of statistical inference in Section 3, and derive asymptotic properties of the proposed methods in Section 4. We present the Monte Carlo simulation in Section 5 and the empirical applications in Section 6. The conclusion is provided in Section 7. Proofs of our theorems are offered in the Appendix, and further numerical details are relegated to an online supplemental material.[4]

We use the following notation throughout: $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers, $\lfloor a \rfloor$ is the largest integer not larger than $a \in \mathbb{R}$, $\lceil a \rceil$ is the smallest integer not smaller than $a \in \mathbb{R}$, the Euclidean norm of any k-dimensional vector $\boldsymbol{c} \in \mathbb{R}^k$ is denoted as $\|\boldsymbol{c}\| = (\boldsymbol{c}^\top \boldsymbol{c})^{1/2}$, $\mathbf{1}(A)$ is the indicator function which equals 1 if event A occurs and 0 otherwise, $\text{Card}(A)$ is the cardinality of set A, $\boldsymbol{I}_k$ is the identity matrix of dimension $k \in \mathbb{N}$, convergence in probability is denoted by $\xrightarrow{p}$, convergence in distribution is denoted by $\xrightarrow{\mathcal{D}}$, weak convergence is denoted by $\Rightarrow$, and weak convergence in probability à la Giné and Zinn (1990) is denoted by $\Rightarrow_p$. The term "stationarity" is taken to mean strict stationarity. C (or C', C'') refers to some generic constant that may take different values in different places unless defined otherwise.

2 The Midastar models

This paper generalizes the threshold autoregressive (TAR) model to *mixed frequency data* by exploiting *Mixed Data Sampling (MIDAS)* methods. We refer to the proposed models as *Midastar*, in which the target variable y and the threshold variable x are

[4] `https://www2.kobe-u.ac.jp/~motegi/research/`

sampled at different frequencies.

We begin by briefly reviewing the single-frequency framework where y and x are observed at same time $t \in \mathbb{L} = \{1, 2, \ldots, n\}$, where n is the sample size. Tong's (1978) two-regime TAR model is formulated as follows.

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

where $\boldsymbol{\beta}_r = (\alpha_r, \phi_{r1}, \ldots, \phi_{rp})^\top$ are regression parameters in regime $r \in \{1, 2\}$; $d \in \mathbb{N}$ is called the delay parameter since a regime at time t is determined by the magnitude of x_{t-d}; $\mu \in \mathbb{R}$ is the threshold parameter. In many applications, d is estimated by choosing a certain optimal value from a pre-specified choice space $\mathcal{D} \subseteq \mathbb{N}$. To ensure the identification of the threshold, μ is typically estimated by choosing a certain optimal value from the choice space $\mathcal{X} = \{x_t\}_{t \in \mathbb{L}}$. Standard time series conditions apply to the error term u_t.

The Midastar models arise when (1) is generalized to mixed frequency data. In Section 2.1, we study the *"regular" Midastar model*, in which the target variable y is observed at a lower frequency than the threshold variable x. One motivating example is a study of whether the dynamics of monthly unemployment rates is influenced by the number of daily online job postings and labor market sentiment. We also explain several key features of the regular Midastar, with a special emphasis on the effects of temporal aggregation on threshold effects. In Section 2.2, we further discuss the regular Midastar specification. Section 2.3 then explores the *"reverse" Midastar model*, where y is observed more frequently than x. The reverse model is also of significant practical importance. A relevant financial application is to analyze how daily asset returns react to monthly or quarterly macro announcements. In the "reverse" scenario, some challenges emerge in the asymptotics due to the inherent lack of stationarity in x. In later sections, we elaborate on these technical difficulties and discuss how we address them.

2.1 Regular Midastar: Specification and key properties

We first consider the case where the target variable y is observed at a low frequency and the threshold variable x^* is observed at a high frequency. Assume that the ratio of sampling frequencies, $m \in \mathbb{N}$, is known and fixed across time (e.g., $m = 3$ when y is sampled quarterly and x is sampled monthly). Define the set of high frequency

time periods as $\mathbb{H} = \cup_{t\in\mathbb{L}}\mathbb{H}_t$, where each low frequency time period is divided into m evenly spaced points:

$$\mathbb{H}_t = \left\{ t-1+\frac{1}{m},\ t-1+\frac{2}{m},\ \ldots,\ t \right\}, \qquad t \in \mathbb{L}. \tag{2}$$

For each $t \in \mathbb{L}$, we observe only one realization y_t for the target variable, while we sequentially observe $\{x_j^*\}_{j\in\mathbb{H}_t}$ for the threshold variable. The asterisk is put on x to emphasize that the high frequency observations of the threshold variable are available.

Define a temporal aggregation of x^*:

$$x_t = \sum_{k=1}^{m} w_k x^*_{t-1+\frac{k}{m}}, \qquad t \in \mathbb{L}, \tag{3}$$

where $\{w_1, \ldots, w_m\} \subseteq \mathbb{R}^m$ are the pre-specified linear aggregation scheme such that $w_k \geq 0$ for all $k \in \{1, \ldots, m\}$ and $\sum_{k=1}^m w_k = 1$. Two well-known examples of the linear aggregation scheme are the following.

Stock aggregation: $w_k = \mathbf{1}(k = m)$ for all $k \in \{1, \ldots, m\}$.

Averaging: $w_k = 1/m$ for all $k \in \{1, \ldots, m\}$.

In the existing TAR literature, model (1) is fitted to $\{y_t\}_{t\in\mathbb{L}}$ and $\{x_t\}_{t\in\mathbb{L}}$ even if $\{x_t^*\}_{t\in\mathbb{H}}$ are observable. In the MIDAS literature, it is well known that such a temporal aggregation often has an adverse impact on statistical inference due to the loss of information on x^*. To fill this gap, we propose the *(regular) Midastar* model:

$$y_t = \begin{cases} \alpha_1 + \sum_{k=1}^p \phi_{1k} y_{t-k} + u_t & \text{if } \ x^*_{t-\frac{d}{m}} < \mu, \\ \alpha_2 + \sum_{k=1}^p \phi_{2k} y_{t-k} + u_t & \text{if } \ x^*_{t-\frac{d}{m}} \geq \mu, \end{cases} \qquad t \in \mathbb{L}. \tag{4}$$

Models (1) and (4) share the same AR structures for y; the difference between the two models appears in the regime-switch conditions. In (4), the delay of d high frequency periods is taken from the integer time period t, exploiting the high frequency observations of x^*. An alternative way to understand the difference between the two models is that Midastar determines in a data-driven way which of $\{w_1, \ldots, w_m\}$ should take 1, whereas TAR fixes the values of $\{w_1, \ldots, w_m\}$ subjectively.

To explain some key features of model (4), we stack the regression and nuisance

parameters as follows:

$$\boldsymbol{\beta}_1 = \begin{bmatrix} \alpha_1 \\ \phi_{11} \\ \vdots \\ \phi_{1p} \end{bmatrix}, \quad \boldsymbol{\beta}_2 = \begin{bmatrix} \alpha_2 \\ \phi_{21} \\ \vdots \\ \phi_{2p} \end{bmatrix}, \quad \underbrace{\boldsymbol{\beta}}_{K\times 1} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} d \\ \mu \end{bmatrix}, \quad \underbrace{\boldsymbol{\theta}}_{(K+2)\times 1} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix}, \quad (5)$$

where $K = 2(p+1)$. If there are no threshold effects (i.e., $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$), the threshold variable is effectively irrelevant and models (1) and (4) are essentially equivalent to each other. Thus, both models are able to detect the truth of no threshold effects.

If there are threshold effects (i.e., $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$), model (1) is generally misspecified relative to model (4), since model (1) cannot capture the high frequency delay in (4). The misspecification results in the failure to identify the true values of $\boldsymbol{\beta}$; in particular, model (1) can yield a flawed conclusion that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ (spurious non-threshold effects). This consequence is in line with the well-known fact that temporal aggregation tends to weaken nonlinearities in original series (e.g., Brännäs and Ohlsson, 1999, Granger and Lee, 1999, Paya and Peel, 2006). In this sense, Midastar is able to model and detect threshold effects more precisely than TAR.

Assuming that threshold effects are present under (4), a special case where model (1) is correctly specified relative to (4) can be constructed as follows. Suppose that the true value of the delay parameter in (4) is a multiple of m (i.e., there exists $c_0 \in \mathbb{N}$ such that $d_0 = c_0 m$). Suppose also that the aggregation scheme is the stock aggregation. Then, model (1) with the delay parameter $d = c_0$ recovers model (4). Verifying the existence of c_0, however, is impossible in practice since d_0 is unknown. Hence, model (1) is haunted by the risk of finding spurious non-threshold effects.

2.2 Further discussions on the model specification

A potential issue of our proposed specification is that the aggregated TAR model (1) is not always a special case of the Midastar model (4). To see this, suppose the true data generating process (DGP) is expressed in the form of (1). If the aggregation scheme is skipped sampling (i.e., $w_k = 1$ for some $k \in \{1, \ldots, m\}$), then model (4) with the choice space of the delay parameter containing md_0 is correctly specified relative to the DGP. Otherwise, model (4) is generally misspecified since $x^*_{t-d/m} \neq x_{t-d_0}$ for any $d \in \mathbb{N}$ in general. In this sense, model (4) is not a strict extension of (1).

This issue stems from our specification that the single high frequency observation $x^*_{t-d/m}$ determines the regime. One might be tempted to generalize (4) with a weighted average of multiple observations of x^*:

$$y_t = \begin{cases} \alpha_1 + \sum_{k=1}^{p} \phi_{1k} y_{t-k} + u_t & \text{if} \quad \sum_{h=1}^{H} \omega_h x^*_{t-\frac{h}{m}} < \mu, \\ \alpha_2 + \sum_{k=1}^{p} \phi_{2k} y_{t-k} + u_t & \text{if} \quad \sum_{h=1}^{H} \omega_h x^*_{t-\frac{h}{m}} \geq \mu, \end{cases} \qquad t \in \mathbb{L} \tag{6}$$

where $\{\omega_1, \ldots, \omega_H\}$ are MIDAS weights such that $\omega_h \geq 0$ for all $h \in \{1, \ldots, H\}$ and $\sum_{h=1}^{H} \omega_h = 1$. Model (4) is a special case of model (6) if $H \geq d$ and $\omega_h = \mathbf{1}(h = d)$ for all $h \in \{1, \ldots, H\}$. Model (1) is also a special case of (6) if $H \geq md + m - 1$ and

$$\omega_h = \begin{cases} w_{md+m-h} & \text{if} \quad h \in \{md, md+1, \ldots, md+m-1\}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\{w_1, \ldots, w_m\}$ are the aggregation scheme appearing in (3). Hence, model (6) would be an ideal specification for mixed frequency data, if the nuisance parameters $\boldsymbol{\gamma} = (\omega_1, \ldots, \omega_H, \mu)^\top$ were identifiable.

Unfortunately, model (6) causes an identification problem of $\boldsymbol{\gamma}$ in practice. Given observed $\{x^*_t\}_{t \in \mathbb{H}}$, there are uncountably many values of $\{\omega_1, \ldots, \omega_H\}$ that lead to the same value of $\sum_{h=1}^{H} \omega_h x^*_{t-h/m}$. This means that optimal value of $\boldsymbol{\gamma}$ that minimizes a loss function is not unique, hence model (6) is practically intractable.

To guarantee the identification of nuisance parameters, we specify the Midastar model as in (4). This decision is admittedly a compromise, but our specification has a notable advantage: the number of parameters does not depend on m, hence a large value of m does not cause any computational problems. Parameter proliferation due to large m is a major issue in many strands of the MIDAS literature, including unrestricted MIDAS regressions (e.g., Foroni, Marcellino, and Schumacher, 2015), mixed frequency vector autoregression (e.g., Ghysels, 2016, Ghysels, Hill, and Motegi, 2016, Götz, Hecq, and Smeekes, 2016), regression-based Granger causality tests (e.g., Ghysels, Hill, and Motegi, 2020), and mixed frequency threshold regression models proposed by Yang and Zhang (2022). Model (4) makes it feasible to perform empirical applications involving large m such as yearly y and monthly x^* with $m = 12$.

2.3 Reverse Midastar: Specification and key properties

So far we assumed that y is sampled at a lower frequency than x. The opposite scenario may well arise in some applications, as aforementioned. In this section, we propose the *reverse Midastar model*, in which the target variable y^* is observed at a *higher* frequency than the threshold variable x. The high frequency observations of the target variable are expressed as $\{y_s^*\}_{s\in\mathbb{H}}$, where $\mathbb{H}$ is defined in (2). Time subscript s takes fractions, and the asterisk on y emphasizes that it is a high frequency variable. The low frequency observations of the threshold variable are expressed as $\{x_t\}_{t\in\mathbb{L}}$ with $\mathbb{L} = \{1, 2, \ldots, n\}$.

We specify the reverse Midastar model as follows.

$$y_s^* = \begin{cases} \alpha_1 + \sum_{k=1}^{p} \phi_{1k} y_{s-k/m}^* + u_s^*, & \text{if} \quad x_{\lceil s \rceil - d} < \mu, \\ \alpha_2 + \sum_{k=1}^{p} \phi_{2k} y_{s-k/m}^* + u_s^*, & \text{if} \quad x_{\lceil s \rceil - d} \geq \mu, \end{cases} \qquad s \in \mathbb{H}, \tag{7}$$

where $d \in \mathbb{N}$ is a fixed delay parameter and $\mu \in \mathbb{R}$ is a fixed threshold parameter. The target variable y^* follows high frequency AR(p) processes with the coefficients being possibly different across regimes. The regime of a high frequency time point $s \in \mathbb{H}$ is determined by the realization of the low frequency threshold variable x with $d \in \mathbb{N}$ periods of delay. The same regime keeps arising for m consecutive high frequency periods within each low frequency period $t \in \mathbb{L}$. We can define the stacking of the regression and nuisance parameters, $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta})$, as in the regular case (5).

Assuming that the true DGP is given by (7), what happens if we aggregate y^* into the low frequency y and fit the single-frequency TAR model (1)? When there are non-threshold effects (i.e., $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$), (7) reduces to the single-regime high frequency AR(p) process, implying that the aggregated y follows an infinite-order AR process in general (e.g., Silvestrini and Veredas, 2008). Hence, the TAR model with finite lag length (1) is misspecified in general, even under the absence of threshold effects.

When there are threshold effects (i.e., $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$), the implied DGP for aggregated y is analytically intractable. Nonetheless, the implied DGP should generally involve an infinite number of autoregressive terms in a nonlinear way, making the TAR model with finite lag length misspecified. Thus, temporal aggregation is expected to produce bias in estimation whether threshold effects are absent or present. These insights highlight the necessity of the reverse Midastar model when the DGP is given by (7).

To further elaborate the properties of the reverse Midastar model, we reindex $\mathbb{H}$

to $\{1, 2, \ldots, N\}$ with $N = mn$. Accordingly, each variable is redefined as follows.

$$Y_t := y^*_{t/m}, \quad X_t := x_{\lceil t/m \rceil}, \quad U_t := u^*_{t/m}; \qquad t \in \{1, 2, \ldots, N\}. \tag{8}$$

By construction, $\{X_t\}_{t=1}^N$ remains the same within each *chunk* of size m.[5] Using the redefined variables and $D = md$, model (7) is rewritten as follows.

$$Y_t = \begin{cases} \alpha_1 + \sum_{k=1}^p \phi_{1k} Y_{t-k} + U_t, & \text{if} \quad X_{t-D} < \mu, \\ \alpha_2 + \sum_{k=1}^p \phi_{2k} Y_{t-k} + U_t, & \text{if} \quad X_{t-D} \geq \mu, \end{cases} \qquad t \in \{1, 2, \ldots, N\}. \tag{9}$$

REMARK. Due to the m-fold repetition of x_t within "chunks", the process $\{X_t\}_{t=1}^N$ defined in (8) does not retain the stationarity properties inherent to x_t but is instead "stationary when indexed at the chunk level" (although the mixing condition is preserved). The following lemma provides a formal demonstration thereof. We note that this is unique to the reverse scenario because, under the regular scenario, the threshold variable is genuinely of high frequency and preserves the usual notion of stationarity. In Section 4 and the Appendix, we address the resulting challenges in establishing the consistency and limiting distribution using Bernstein's blocking technique. The objective is to develop a test for the threshold effect along with the corresponding asymptotic theory.

Lemma 1. *Suppose the process $\{x_t\}_{t=1}^n$ is stationary and α-mixing with the mixing rate $\alpha(k) = O(k^{-A})$ for some $A > 0$. Suppose further that x_t is absolutely continuous with respect to the Lebesgue measure with density that is bounded away from zero on its support. Then, the process $\{X_t\}_{t=1}^N$ defined in (8) and presented in (9) is not stationary, while it is α-mixing with the same mixing rate as $\{x_t\}$.*

The proof of Lemma 1 is provided in Appendix A.1. While the non-stationarity of X is a relatively straightforward consequence of repetition within chunks, the preservation of the α-mixing property is less immediate and plays an important role in establishing the asymptotic distribution.

[5]Note well that we refer to these as "chunks" (instead of the more formal term "blocks") to avoid any potential confusion with "(Bernstein) blocks", a different concept that we employ in the asymptotic theory and its derivations.

3 Statistical inference

In this section, we describe the statistical inference of the regular Midastar model (4). Similar arguments apply to the reverse Midastar model (7) or equivalently (9), with minor notational adjustments due to flipped sampling frequencies. The interested reader is referred to REMARK at the end of Section 3.1.

Stacking the regressors, we have that

$$\underbrace{\boldsymbol{z}_t}_{(p+1)\times 1} = \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t+1-p} \end{bmatrix}, \quad \underbrace{\boldsymbol{Z}_t(\boldsymbol{\gamma})}_{K\times 1} = \begin{bmatrix} \boldsymbol{z}_{t-1}\mathbf{1}\left(x^*_{t-\frac{d}{m}} < \mu\right) \\ \boldsymbol{z}_{t-1}\mathbf{1}\left(x^*_{t-\frac{d}{m}} \geq \mu\right) \end{bmatrix}, \qquad t \in \mathbb{L}, \tag{10}$$

where $\boldsymbol{\gamma} = (d, \mu)^\top$ as defined in (5). The notation of $\boldsymbol{Z}_t(\boldsymbol{\gamma})$ calls for some caution, since the exact time period when we observe $\boldsymbol{Z}_t(\boldsymbol{\gamma})$ is $t - \min\{1, d/m\}$, which is prior to time t for any $d \in \mathcal{D}$. Note hereafter that $\boldsymbol{Z}_t(\boldsymbol{\gamma})$ is pre-determined prior to y_t.

Using (10), model (4) can be rewritten as a single equation:

$$y_t = \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \boldsymbol{\beta} + u_t, \qquad t \in \mathbb{L}, \tag{11}$$

where $\boldsymbol{\beta}$ is defined in (5). One or both elements of $\boldsymbol{\gamma}$ could be pre-specified by the researcher, but this paper estimates both of them in order to keep sufficient generality. To focus on the estimation of the target parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, we sidestep a lag selection issue by assuming that p is known.[6]

In Section 3.1, we describe the profiling estimation of $\boldsymbol{\theta}$. In Section 3.2, we propose to use the wild bootstrap of Hansen (1996) to test the no-threshold-effect hypothesis. In Section 3.3, we describe an asymptotic loss-differential test of Diebold and Mariano (1995) to compare the out-of-sample performance of Midastar and TAR. These procedures are mostly analogous to the standard TAR models, but some extra attention is required since we need to accommodate two sampling frequencies.

[6]Data-driven lag selection of self-exciting TAR models are explored by Wong and Li (1998) and Galeano and Peña (2007), among others. Extending these studies to Midastar is left as a future task.

3.1 Profiling estimation

To implement profiling, we construct the choice spaces of delay parameter d and threshold parameter μ. The choice space of d, denoted as $\mathcal{D} = \{\underline{d}, \underline{d}+1, \ldots, \bar{d}\}$, needs to be constructed with some care about a timing issue. For model (4) to be well defined, $x^*_{t-\underline{d}/m}$ must be observed prior to y_t, hence it matters at which point of $\mathbb{H}_t$ we observe y_t. If y_t is observed at the last point of $\mathbb{H}_t$ (i.e., the integer time period t), then $\underline{d}$ can take the smallest possible value of 1. If y_t is observed at the first point of $\mathbb{H}_t$ (i.e., the non-integer time period $t - 1 + 1/m$), then $\underline{d}$ should be at least as large as m. In MIDAS, this sort of timing issue arises when analyzing ragged-edge data (e.g., Marcellino and Schumacher, 2010). Unless otherwise noted, we assume for notational simplicity that y_t is observed at the last point of $\mathbb{H}_t$. In this case, a natural specification is $\underline{d} = 1$ and hence $\mathcal{D} = \{1, 2, \ldots, \bar{d}\}$.

For the choice space of μ, let $\{x_t\}_{t\in\mathbb{L}}$ be the average of $\{x^*_t\}_{t\in\mathbb{H}}$ in accordance with (3). Let $x_{[1]} \leq \cdots \leq x_{[n]}$ be a sorted version of $\{x_t\}_{t\in\mathbb{L}}$. In general, the space of μ is specified as

$$\mathcal{X}_{\kappa,n} = \left\{x_{[\lfloor 0.5(1-\kappa)n \rfloor]}, \ldots, x_{[\lfloor \{1-0.5(1-\kappa)\}n \rfloor]}\right\}, \tag{12}$$

where $\kappa \in [0, 1)$ signifies the fraction of $\text{Card}(\mathcal{X}_{\kappa,n})$ to n. Given κ, each of the two regimes accounts for at least $50(1-\kappa)\%$ of the whole sample on average. Choosing a value for κ that is too large (e.g., $\kappa = 0.9$) may cause an identification problem in small samples. Following a well-known suggestion of Andrews (1993), we pick $\kappa = 0.7$ so that each regime accounts for at least 15% of the entire sample on average.

The choice space of $\boldsymbol{\gamma}$ is denoted as $\Gamma_{\kappa,n} = \mathcal{D} \times \mathcal{X}_{\kappa,n}$. The space of the regression parameter $\boldsymbol{\beta}$ is $B \subseteq \mathbb{R}^K$. Finally, the space of the entire parameter vector $\boldsymbol{\theta}$ is $\Theta_n = B \times \Gamma_{\kappa,n}$.

Remark. Here $\Gamma_{\kappa,n}$ is the finite, sample-dependent search grid used for computation; see for example, the argmin in (15) and the summations in (20) below. This should be distinguished from $\Gamma := \mathcal{D} \times [\underline{\mu}, \overline{\mu}]$ in Section 4, which denotes a fixed compact parameter set used to index the uniform-in-γ asymptotic arguments. As $n \to \infty$, the empirical grid $\Gamma_{\kappa,n}$ provides an increasingly fine discretization of Γ (in particular, of the threshold component). Specifically, under Assumption 1 we introduce later, we may choose fixed constants $\underline{\mu} < \overline{\mu}$ such that $\mathbb{P}(\mathcal{X}_{\kappa,n} \subseteq [\underline{\mu}, \overline{\mu}]) \to 1$, and therefore $\Gamma_{\kappa,n} \subseteq \Gamma$, w.p.a. 1 (i.e. $\mathbb{P}(\Gamma_{\kappa,n} \subseteq \Gamma) \to 1$) as $n \to \infty$. We write $\Theta := \mathcal{B} \times \Gamma$ for the

theoretical parameter space we use for asymptotic theory and work on the space in line with the existing literature, e.g. Hansen (1996).

Define the quadratic loss function $\mathcal{L}(\boldsymbol{\theta}) = \sum_{t\in\mathbb{L}}\{y_t - \mathbb{E}_{\boldsymbol{\theta}}(y_t \,|\, y_{t-1},\ldots,y_1)\}^2$, where $\mathbb{E}_{\boldsymbol{\theta}}(y_t \,|\, y_{t-1},\ldots,y_1)$ is the conditional expectation of y_t given $\boldsymbol{\theta}$ and the past observations. The least squares (LS) estimator for $\boldsymbol{\theta}$, defined as $\widehat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}\in\Theta_n}\ \mathcal{L}(\boldsymbol{\theta})$, can be computed via a two-step approach called *profiling*. Conditional on $\boldsymbol{\gamma} \in \Gamma_{\kappa,n}$, the LS estimator for $\boldsymbol{\beta}$ based on (11) is given by

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left\{\sum_{t\in\mathbb{L}} \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right\}^{-1}\left\{\sum_{t\in\mathbb{L}} \boldsymbol{Z}_t(\boldsymbol{\gamma})y_t\right\}. \tag{13}$$

The resulting residual is given by

$$\widehat{u}_t(\boldsymbol{\gamma}) = y_t - \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}). \tag{14}$$

The LS estimator for $\boldsymbol{\gamma}$ can be computed as follows:

$$\widehat{\boldsymbol{\gamma}} = \operatorname*{argmin}_{\boldsymbol{\gamma}\in\Gamma_{\kappa,n}}\ \sum_{t\in\mathbb{L}} \widehat{u}_t(\boldsymbol{\gamma})^2. \tag{15}$$

Substitute (15) into (13) to get $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}})$, resulting in $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\gamma}}^\top)^\top$.

REMARK. For completeness, we briefly outline how the estimators are defined in the case of the *reverse Midastar*, although this is fairly straightforward from the definitions. In view of (7)-(9), we have

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left\{\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right\}^{-1}\left\{\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma})Y_t\right\}, \quad \boldsymbol{Z}_t(\boldsymbol{\gamma}) = \begin{bmatrix}\mathbf{1}(X_{t-D} < \mu)\boldsymbol{z}_{t-1}\\ \\ \mathbf{1}(X_{t-D} \geq \mu)\boldsymbol{z}_{t-1}\end{bmatrix} \tag{16}$$

with $\boldsymbol{z}_t = (1, Y_t, \ldots, Y_{t+1-p})^\top$, and also

$$\widehat{\boldsymbol{\gamma}} = \operatorname*{argmin}_{\boldsymbol{\gamma}\in\Gamma_{\kappa,n}} \sum_{t=1}^{N}\left\{Y_t - \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})\right\}^2, \qquad \widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}\big(\widehat{\boldsymbol{\gamma}}\big). \tag{17}$$

3.2 Bootstrap tests for the no-threshold-effect hypothesis

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

$$H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2, \qquad H_1^* : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2, \tag{18}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the regression parameters in each regime, as defined in (5). The asterisk is put in (18) to emphasize that the no-threshold-effect hypothesis is a special case of a linear parametric restriction:

$$H_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{q}, \qquad H_1 : \boldsymbol{R}\boldsymbol{\beta} \neq \boldsymbol{q}. \tag{19}$$

When we specify $\boldsymbol{R} = (\boldsymbol{I}_{p+1},\ -\boldsymbol{I}_{p+1})$ and $\boldsymbol{q} = \boldsymbol{0}_{(p+1)\times 1}$, (19) reduces to (18). It makes the tests easier to formulate to recognize that H_0^* is a special case of H_0.

Under H_0^*, Midastar reduces to the single-regime AR(p) where the threshold variable x^* does not play any role. Hence, the nuisance parameter $\boldsymbol{\gamma}$ is unidentified under H_0^*. This identification issue makes the test of H_0^* non-standard, and a well-known solution is the wild bootstrap of Hansen (1996). To describe the bootstrap procedure, we define some quantities conditional on $\boldsymbol{\gamma} \in \Gamma_{\kappa,n}$. The estimated regression score is given by $\widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_t(\boldsymbol{\gamma})\widehat{u}_t(\boldsymbol{\gamma})$, where $\widehat{u}_t(\boldsymbol{\gamma})$ is defined in (14). The Wald test statistic with respect to (18) is given by

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

where $\boldsymbol{R} = (\boldsymbol{I}_{p+1},\ -\boldsymbol{I}_{p+1})$, $\boldsymbol{q} = \boldsymbol{0}$, and $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ is the conditional LS estimator defined in (13).[7] The heteroscedasticity-robust covariance matrix estimator is constructed as $\widehat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\widehat{\boldsymbol{S}}_n(\boldsymbol{\gamma})\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}$, where $\widehat{\boldsymbol{S}}_n(\boldsymbol{\gamma}) = n^{-1}\sum_{t\in\mathbb{L}} \widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma})\widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma})^\top$ and $\boldsymbol{M}_n(\boldsymbol{\gamma}) = n^{-1}\sum_{t\in\mathbb{L}} \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top$.

Common methods to aggregate the functional $\mathcal{W}_n(\boldsymbol{\gamma})$ over $\Gamma_{\kappa,n}$ include supremum, average, and exponential transformations:

$$\sup\mathcal{W}_n = \sup_{\boldsymbol{\gamma}\in\Gamma_{\kappa,n}} \mathcal{W}_n(\boldsymbol{\gamma}), \qquad \text{ave}\mathcal{W}_n = \frac{1}{\text{Card}(\Gamma_{\kappa,n})}\sum_{\boldsymbol{\gamma}\in\Gamma_{\kappa,n}} \mathcal{W}_n(\boldsymbol{\gamma}), \tag{20}$$

[7] The Lagrange multiplier (LM) test can be constructed analogously, and hence omitted to save space. See Motegi, Dennis, and Hamori (2023) for the detailed construction of the LM test.

$$\exp\mathcal{W}_n = \ln\left[\frac{1}{\text{Card}(\Gamma_{\kappa,n})}\sum_{\boldsymbol{\gamma}\in\Gamma_{\kappa,n}}\exp\left\{\frac{\mathcal{W}_n(\boldsymbol{\gamma})}{2}\right\}\right].$$

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 $t \in \mathbb{L}$.

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

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

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

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:

$$\widehat{p}_n^B(H_0^*) = \frac{1}{B}\sum_{b=1}^{B}\mathbf{1}\left[g\left\{\mathcal{W}_n^{(b)}\right\} \geq g(\mathcal{W}_n)\right]. \tag{21}$$

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

REMARK. The bootstrap test described above can be applied to the reverse Midastar model with only minor adjustments, employing the quantities defined in (16)-(17).

3.3 Comparing predictive power

In the previous section, we describe the bootstrap tests for the no-threshold-effect hypothesis H_0^*. Testing H_0^* amounts to an *in-sample* analysis where we compare the relative validity of the regular Midastar and AR models. In this section, we formulate an *out-of-sample* analysis which compares the relative predictive ability of the Midastar and TAR models. For expositional simplicity, the prediction horizon is fixed at $h = 1$ low frequency period. Multi-step-ahead forecasting can be implemented analogously.

We begin by pointing out a *nowcasting* aspect of the Midastar-based prediction. In view of (4), the one-step-ahead prediction of y_{t+1} with $t \in \mathbb{L}$ requires the realizations of $\{y_{t+1-p}, y_{t+2-p},\ldots,y_t\}$ and $x_{t+1-d/m}^*$. The former is simply the past and present

realizations of the target variable, while the latter is a past realization of the threshold variable if $d > m$, a present realization if $d = m$, and a "future" realization if $d < m$. We might need one of the first $m-1$ high frequency observations of x^* within $t+1 \in \mathbb{L}$, in which sense the Midastar-based prediction is a sort of nowcasting.

We perform the rolling window out-of-sample prediction based on observed $\{y_t\}_{t\in\mathbb{L}}$ and $\{x_t^*\}_{t\in\mathbb{H}}$, where the low frequency sample size is denoted as $\bar{n} \in \mathbb{N}$. First, fit the Midastar model to $\{y_t\}_{t=1}^{\bar{n}}$ and compute the profiling estimator as described in Section 3.1. The window size is fixed at $n = \lfloor \tau\bar{n} \rfloor$ with $\tau = 0.8$.[8] Second, compute the one-step-ahead forecast of y_{n+1}, denoted as $\widehat{y}_{n+1}$. Analogously, fit the Midastar model to $\{y_t\}_{t=2}^{n+1}$ and compute the one-step-ahead forecast of y_{n+2}, denoted as $\widehat{y}_{n+2}$. Stop once $\{\widehat{y}_t\}_{t=n+1}^{n+T}$ is obtained, where $T = \bar{n} - n$ is the number of windows used for forecast evaluation. The forecast error is given by $\widehat{e}_t = y_t - \widehat{y}_t$ for $t \in \{n+1, \ldots, n+T\}$. The mean squared forecast error (MSE) is defined as $MSE_T = T^{-1}\sum_{t=n+1}^{n+T} \widehat{e}_t^2$.

A simple way to compare the predictive accuracies of Midastar and TAR is to compare their MSEs. Further, Diebold and Mariano (1995) [DM1995] established a novel approach to test if the spread between the two MSEs is statistically significant. Compute $\widehat{d}_t = (\widehat{e}_t^{midas})^2 - (\widehat{e}_t^{tar})^2$ for $t \in \{n+1, \ldots, n+T\}$, where $\widehat{e}_t^{midas}$ and $\widehat{e}_t^{tar}$ are the forecast errors associated with Midastar and TAR, respectively. DM1995 proposed multiple ways to transform $\{\widehat{d}_t\}_{t=n+1}^{n+T}$ into a single test statistic. The mean difference test statistic S_1 is most commonly used in the literature, and its core component is as follows.

$$\bar{d}_T = \frac{1}{T}\sum_{t=n+1}^{n+T} \widehat{d}_t = MSE_T^{midas} - MSE_T^{tar},$$

where MSE_T^{midas} and MSE_T^{tar} are the MSEs of Midastar and TAR, respectively.

The null hypothesis of the DM test, denoted as H_0^{eq}, is that the Midastar-based and TAR-based forecasts are equally accurate. DM1995 showed that, under H_0^{eq} and some mild regularity conditions, the test statistic follows the standard normal distribution asymptotically. We consider three alternative hypotheses separately:

H_1^{eq} : The two forecasts have different accuracies (two-sided test). Reject H_0^{eq} at the $100a\%$ level if $|S_1| > \Phi^{-1}(1 - a/2)$, where S_1 is the test statistic and $\Phi^{-1}(\cdot)$ is the inverse distribution function of the standard normal distribution.

H_1^{midas} : The Midastar forecast is more accurate than the TAR forecast (one-

[8] Alternative values such as $\tau \in \{0.7, 0.9\}$ could be used analogously. An increasing window design where all windows begin with $t = 1$ is also analogous.

sided test). Reject H_0^{eq} at the $100a\%$ level if $S_1 < \Phi^{-1}(a)$.

H_1^{tar} : The TAR forecast is more accurate than the Midastar forecast (one-sided test). Reject H_0^{eq} at the $100a\%$ level if $S_1 > \Phi^{-1}(1-a)$.

4 Asymptotic theory

This section presents asymptotic results we derive for the Midastar models. Section 4.1 examines the *regular* Midastar, where the threshold variable naturally arrives at a higher frequency than the target. In this framework, standard dynamic conditions for threshold regressions apply, and we can establish profiling consistency, asymptotic normality, and a valid bootstrap for testing the absence of a threshold effect under stationarity. Section 4.2 covers the *reverse* Midastar, where upsampling the low-frequency threshold variable necessitates repetitions in "chunks," thereby precluding stationarity. We outline the econometric methods that overcome these complications, and establish that the same profiling estimator achieves uniform consistency and asymptotic normality under suitably generalized assumptions. In constructing uniform consistency and the limiting distribution (Theorems 1 and 3), we derive the theory under the classical assumptions of α-mixing, rather than relying on the limiting theory of Hansen (1996), which assumes relatively restrictive conditions including absolute regularity (i.e., β-mixing). We retain the latter conditions only for bootstrap consistency (Theorems 2 and 4).

We begin by defining the notations for key population quantities, which will be used throughout this section:

$$\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \boldsymbol{M}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_1)^{-1} \boldsymbol{S}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \boldsymbol{M}(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_2)^{-1}, \tag{22}$$

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

where $\boldsymbol{s}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_t(\boldsymbol{\gamma}) u_t$. 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 $\widehat{\boldsymbol{V}}_n(\boldsymbol{\gamma})$, $\widehat{\boldsymbol{S}}_n(\boldsymbol{\gamma})$, and $\boldsymbol{M}_n(\boldsymbol{\gamma})$. The reader is referred to the Appendix for further details as well as the proofs for the theorems presented in this section.

4.1 Regular Midastar: Assumptions and main results

We discuss the asymptotic theory for threshold regressions in the regular Midastar framework below. In this case, we can impose the standard conditions of stationarity and asymptotic independence. Although the specification for regular Midastar allows us to adopt the asymptotic theory of Hansen (1996), we relax the absolute regularity that Hansen (1996) requires, and impose α-mixing type conditions instead in deriving the limiting distribution.

Assumption 1. *The processes $\{(y_t, u_t)\}_{t\in\mathbb{L}}$ and $\{x_t^*\}_{t\in\mathbb{H}}$ are jointly stationary and α-mixing with the mixing rate $\alpha(k) = O(k^{-A})$ for some $A > \frac{2\nu}{2\nu-1} > 0$ and the constant ν is as defined below in Assumption 2. Additionally, u_t is a martingale difference with respect to its natural filtration. Furthermore, x_t^* is absolutely continuous with respect to the Lebesgue measure with density that is bounded above and below by positive constants on its support.*

Remark. We note that stationarity and α-mixing conditions are assumed in a joint sense here. This implies that any (finite-dimensional) measurable transformation of the joint process is ergodic in the sense that Birkhoff's ergodic theorem holds (Ibragimov and Linnik, 1971, Chen, Hansen, and Carrasco, 2010). We note that our theory does not require the serial independence on the error term that the existing literature postulates. The martingale difference condition we assume can be further relaxed, in which case we can employ a heteroskedasticity and autocorrelation consistent (HAC)-type estimator. Overall, the conditions stipulated in Assumption 1 are considerably weaker than those assumed in Hansen (1996). The mixing condition controls the degree of temporal dependence, and the arithmetic rate we impose is weak enough to ensure that relevant inequalities and limit theorems for dependent processes can be applied (see Rio, 2017, for comprehensive details). The existence of a bounded density is a mild condition often assumed in the literature (see Aït-Sahalia and Park, 2012, among many others); we use this condition when constructing uniform convergence via the Vapnik-Chervonenkis (VC)-type covering argument.

Assumption 2. *The following conditions are assumed:*

(i) There exists some $\nu > 1$ such that $\mathbb{E}|\bar{z}_t|^{4\nu} < \infty$ and $\mathbb{E}|u_t|^{4\nu} < \infty$, where u_t is a martingale difference, $\bar{z}_t$ is a uniform bound on the regressor vector $\boldsymbol{Z}_t(\boldsymbol{\gamma})$ over $\boldsymbol{\gamma} \in \Gamma := \mathcal{D} \times [\underline{\mu}, \overline{\mu}]$ for some finite constants $\underline{\mu}, \overline{\mu}$, and Γ is a compact set.

(ii) $\inf_{\boldsymbol{\gamma}\in\Gamma} \det\{\boldsymbol{M}(\boldsymbol{\gamma})\} > 0$, where $\boldsymbol{M}(\boldsymbol{\gamma})$ is defined in (22).

(iii) The true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)^\top$ *is an interior point of* Θ.

REMARK. Assumption 2 is mild and general. First, the moment condition is widely adopted in nonlinear time series analysis (e.g., Hall and Yao, 2003, Hong and Lee, 2013). It is necessary for deriving the central limit theorem of the product processes; an alternative is to impose an almost sure bound, as discussed in Peligrad (1986). The compactness of Γ is also standard, Hansen (1996), and is needed to cover the set by a finite number of "balls" in order to carry out the uniform-in-γ argument. We recall the distinction between the trimmed order-statistic grid $\Gamma_{\kappa,n}$ and the compact set Γ we discussed in Remark in Section 3.1. The second condition (ii) is imposed to validate the inversion of the sample cross-product matrix and its probability limit. Furthermore, the interior-point assumption (iii) is also a standard requirement for applying the usual local-analysis arguments. We remark that since the delay index d is discrete, the interior-point condition in (iii) is to be understood with respect to the continuous components, i.e., $\beta_0 \in \text{int}(\mathcal{B})$ and $\mu_0 \in (\underline{\mu}, \overline{\mu})$, with $d_0 \in \mathcal{D}$.

We require variations in the threshold variable around the threshold so that, if H_1^* is true, identification of $\boldsymbol{\beta}$ will be possible. This restricts points of identification failure in the model to be those associated with H_0^*. Sufficient conditions for Assumption 1 and 2-(i) include iid $\{u_t\}$ with finite $4\iota^{th}$ moment for some $\iota > 1$ and the regime-wise stability condition (i.e., the roots of the characteristic equation $\lambda^p - \sum_{k=1}^{p} \lambda^{p-k}\phi_{rk} = 0$ lie strictly inside the unit circle for each $r \in \{1, 2\}$). 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).

Theorem 1 below shows that under the aforementioned conditions, we have the uniform consistency and asymptotic normality of the profiling estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$.

Theorem 1. *Suppose Assumptions 1 and 2 hold. Under the null* H_0^* *(*i.e., $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_0^*$*), the following statements hold as* $n \to \infty$*:*

(i) [Uniform Consistency] *Uniformly over* $\boldsymbol{\gamma} = (d, \mu)^\top \in \Gamma$, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ *is consistent for* $\boldsymbol{\beta}_0^*$. *That is, we have*

$$\sup_{\boldsymbol{\gamma} \in \Gamma} \left\| \widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^* \right\| \xrightarrow{p} 0.$$

(ii) [Asymptotic Normality] *For each fixed* $\boldsymbol{\gamma} = (d, \mu)^\top \in \Gamma$,

$$\sqrt{n}\big[\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^*\big] \xrightarrow{\mathcal{D}} \mathcal{G}(\boldsymbol{\gamma}),$$

where $\mathcal{G}(\boldsymbol{\gamma})$ is a mean-zero Gaussian process indexed by $\boldsymbol{\gamma}$ with covariance kernel $\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, which is defined in (22).

The proof of Theorem 1 is provided in Appendix A.3.

We now perform a thought experiment where $\boldsymbol{\beta}_2$ depends on the sample size: $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2n}$. Consider a local alternative hypothesis:

$$H_1^*(\boldsymbol{\lambda}_n) : \boldsymbol{\beta}_{2n} = \boldsymbol{\beta}_1 + n^{-\frac{1}{2}}\boldsymbol{\lambda}_n, \qquad \boldsymbol{\lambda}_n \in \mathbb{R}^{p+1}. \tag{23}$$

The drift term $n^{-1/2}\boldsymbol{\lambda}_n$ determines distance between regimes 1 and 2. We utilize Andrews and Cheng's (2012) characterization of identification categories. 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 (23). By construction, (23) can be rewritten as

$$H_1^*(\boldsymbol{\lambda}_n) : \boldsymbol{\beta}_n = \bar{\boldsymbol{\beta}}_1 + n^{-\frac{1}{2}}\bar{\boldsymbol{\lambda}}_n. \tag{24}$$

The asymptotic properties of the profiling estimator under the local alternatives H_1^* are as follows. First, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_n$ converges to $\mathbf{0}$ uniformly over any $\boldsymbol{\gamma} \in \Gamma$ as $n \to \infty$, implying consistency of the profiling estimator $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}})$ under all identification categories. In particular, the convergence holds even when $\widehat{\boldsymbol{\gamma}}$ is not consistent. Second, a consistent estimator for $\boldsymbol{\gamma}$ is not available under non-identification and weak identification, but is available under semi-strong and strong identification. Third, under non-identification and weak identification, $\sqrt{n}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}}) - \boldsymbol{\beta}_n)$ may not be asymptotically normal because $\widehat{\boldsymbol{\gamma}}$ does not converge; consequently, $\boldsymbol{\gamma}$ is treated as a nuisance parameter and we employ functionals over Γ to test H_0^*. These results are summarized below, and follow immediately from Theorem 1:

Corollary 1. *Under Assumptions 1 and 2, the following statements hold as $n \to \infty$:*

(i) $\sup_{\boldsymbol{\gamma}\in\Gamma}||\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})-\boldsymbol{\beta}_n||\xrightarrow{p}0$. *(ii) Under* $H_1^*(\boldsymbol{\lambda}_n)$ *with* $\lim_{n\to\infty}\|\boldsymbol{\lambda}_n\|<\infty$, $\sqrt{n}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})-\bar{\boldsymbol{\beta}}_1\}\xrightarrow{\mathcal{D}}\bar{\boldsymbol{\lambda}}+\mathcal{G}(\boldsymbol{\gamma})$.

Next, we consider the bootstrap test for the no-threshold-effect hypothesis H_0^*, under a set of stronger conditions. The uniform convergence of the conditional LS estimator is a key condition for the asymptotic validity of this test. While Corollary 1(ii) above offers a pointwise (fixed-γ) convergence in distribution result under the mild α-mixing conditions, the boostrap test requires a functional weak convergence result for the relevant processes indexed by γ. Under a stronger set of conditions in the style of Hansen (1996), the conditional Wald and LM test statistics weakly converge to a chi-squared process over $\boldsymbol{\gamma}\in\Gamma$ under H_0^*.

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 $\widehat{\boldsymbol{v}}_n(\boldsymbol{\gamma})=n^{-1/2}\sum_{t\in\mathbb{L}}\widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma})\xi_t$ with $\xi_t\overset{i.i.d.}{\sim}\mathcal{N}(0,1)$. It follows that $\widehat{\boldsymbol{v}}_n(\boldsymbol{\gamma})$ 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 (22). This implies convergence in distribution of the bootstrap p-value $\widehat{p}_n^B(H_0^*)$ in (21) to a uniform random variable under H_0^*.

The asymptotic power of the test depends on the identification category. 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, the test rejects H_0^* with probability approaching 1. These results are summarized in Theorem 2.

Theorem 2. *If Assumptions 1 and 2 hold with the mixing condition strengthened to absolute regularity with rate* $\eta(k)=O(k^{-B})$ *for some* $B>\varphi(\varphi-1)^{-1}$ *and* $\varphi>1$*then the following are true. (i) Under* $H_0^*:\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$*, the wild-bootstrap p-value* $\widehat{p}_n^B(H_0^*)$ *is asymptotically Uniform on* $[0,1]$. *(ii) Under* $H_1^*(\boldsymbol{\lambda}_n)$ *with* $\|\boldsymbol{\lambda}_n\|\to\infty$, $\widehat{p}_n^B(H_0^*)\xrightarrow{p}0$ *as* $n\to\infty$ *and* $B\to\infty$.

The proof of Theorem 2 is provided in Appendix A.4.

4.2 Reverse Midastar: Assumptions and main results

As aforementioned in Section 2.3, the m-fold repetition of x results in a repetitive structure within each "chunk" and therefore yields non-stationarity in its dynamics. We show that these difficulties can be effectively handled, and the limiting theories can be established; see the Appendix for full details.

One noteworthy econometric contribution of our paper is the establishment of the asymptotic theory under considerably more general conditions than what is assumed in the literature. To the best of our knowledge, this is the first threshold-regression framework to derive the uniform consistency and asymptotic normality without requiring stationarity of the threshold variable. We believe our results are applicable to broad contexts including non-stationarity due to mixed frequency data.

Below we construct limiting theories for threshold regressions in the situations we face in the reverse model case. What follows is the set of assumptions we impose on the joint dynamics of the process (Y_t, X_t, U_t), cf. (8).

Assumption 1′. *The joint process $\{(Y_t, X_t, U_t)\}_{t=1}^N$ is ergodic as well as α-mixing with the mixing rate $\alpha(k) = O(k^{-A})$ for some $A > \frac{2\nu}{2\nu-1} > 0$, where U_t is a martingale difference, $X_t := x_{\lceil t/m \rceil}$ with $\{x_t\}_{t=1}^n$ being stationary, and ν is as defined below in Assumption 2′. Furthermore, X_t is absolutely continuous with respect to the Lebesgue measure with density that is bounded above and below by positive constants on its support.*

Remark. These conditions broadly follow Assumption 1, except stationarity. We reiterate that $\{X_t\}_{t=1}^N$ is not stationary, due to the repetition of some same values, see Lemma 1 in Section 2.3. Furthermore, ergodicity and mixing are imposed in a joint sense. We are not assuming (Y_t, U_t) to be stationary, and nonetheless, the theory to follow continues to hold. In deriving the asymptotics, we employ Bernstein's blocking method, which sets the dynamics to mimic approximate independence as the gap between the blocks grows. Lastly, ergodicity and α-mixing are *both* assumed, unlike the case where we have stationarity before. It is well known that α-mixing alone is not sufficient for ergodicity without stationarity (see Ibragimov and Linnik, 1971, Corollary 17.1.1).

Assumption 2′. *The conditions assumed in Assumption 2 continue to hold, except that the error process $\{u_t\}_{t\in\mathbb{L}}$ is now replaced with $\{U_t\}_{t=1}^N$.*

Theorem 3 below shows that under the aforementioned conditions, we continue to have the uniform consistency and asymptotic normality of the profiling estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ in the *reverse* framework.

Theorem 3. *Suppose Assumptions 1′ and 2′ hold. Then, under the null H_0^* (*i.e.*, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_0^*$), the following statements hold as $n \to \infty$ (i.e., as $N = nm \to \infty$):*

(i) [Uniform Consistency] *Uniformly over* $\boldsymbol{\gamma} = (d, \mu)^\top \in \Gamma$, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ *is consistent for* $\boldsymbol{\beta}_0^*$. *That is, we have*

$$\sup_{\boldsymbol{\gamma} \in \Gamma} \left\| \widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^* \right\| \xrightarrow{p} 0.$$

(ii) [Asymptotic Normality] *For each fixed* $\boldsymbol{\gamma} = (d, \mu)^\top \in \Gamma$,

$$\sqrt{N}\big[\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^*\big] \xrightarrow{\mathcal{D}} \mathcal{G}(\boldsymbol{\gamma}), \tag{25}$$

where $\mathcal{G}(\boldsymbol{\gamma})$ *is a mean-zero Gaussian process indexed by* $\boldsymbol{\gamma}$ *with covariance kernel* $\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, *which is defined in* (22).

The proof of Theorem 3 is provided in Appendix A.2.

A brief sketch of the main steps in the proof is as follows. First, we apply Bernstein's blocking approach to separate the sample into blocks and gaps so as to control dependence. Second, we employ Davydov-type inequalities to bound covariances across blocks and thus ensure pointwise consistency. Third, the standard Vapnik-Chervonenkis covering argument is utilized to achieve uniform convergence. Finally, we establish the asymptotic normality in a manner similar to how an α-mixing CLT for the partial sums is constructed via an asymptotic independence argument.

We now establish the following Theorem 2 analogue of reverse Midastar. The reverse Midastar quantities are meant to be computed using reverse-sample estimator and residuals, summing over $t = 1, \ldots, N$, and using the $\sqrt{N}$-scaling implied by the reverse asymptotics.

Theorem 4. *Let* $\widehat{p}_N^{\mathcal{B}}(H_0^*)$ *denote the wild-bootstrap p-value constructed exactly as in (21), but computed using the reverse Midastar quantities. If Assumptions* $1'$ *and* $2'$ *hold with the mixing condition strengthened to absolute regularity with rate* $\eta(k) = O(k^{-B})$ *for some* $B > \varphi(\varphi - 1)^{-1}$ *and* $\varphi > 1$, *then the following are true: (i)Under* $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, *the wild-bootstrap p-value* $\widehat{p}_N^{\mathcal{B}}(H_0^*)$ *is asymptotically Uniform on* $[0, 1]$. *(ii) Under the local alternative* $H_1^*(\boldsymbol{\lambda}_N)$ *defined by* $\boldsymbol{\beta}_{2,N} = \boldsymbol{\beta}_1 + N^{-1/2}\boldsymbol{\lambda}_N$ *(equivalently,* $\boldsymbol{\beta}_N = \bar{\boldsymbol{\beta}}_1 + N^{-1/2}\bar{\boldsymbol{\lambda}}_N$*) with* $\|\boldsymbol{\lambda}_N\| \to \infty$, *we have* $\widehat{p}_N^{\mathcal{B}}(H_0^*) \xrightarrow{p} 0$ *as* $n \to \infty$ ($N \to \infty$) *and* $B \to \infty$.

The proof of Theorem 4 is provided in Appendix A.5.

5 Monte Carlo simulation

This section conducts Monte Carlo simulations to evaluate the finite sample performance of our proposed methods. The regular and reverse Midastar models are covered in Sections 5.1 and 5.2, respectively.

5.1 Regular Midastar

Focusing on the regular Midastar, we present our simulation design in Section 5.1.1 and results in Section 5.1.2. Additional simulation results are summarized in Section 5.1.3.

5.1.1 Simulation design

Suppose that the DGP of the low frequency target variable y is the regular Midastar process with lag length $p = 1$:

$$y_t = \begin{cases} \alpha_{10} + \phi_{10} y_{t-1} + \epsilon_t & \text{if } x^*_{t-\frac{d_0}{m}} < \mu_0, \\ \alpha_{20} + \phi_{20} y_{t-1} + \epsilon_t & \text{if } x^*_{t-\frac{d_0}{m}} \geq \mu_0, \end{cases} \qquad t \in \mathbb{L}, \tag{26}$$

where the regression parameters in regime 1 are $\boldsymbol{\beta}_{10} = (\alpha_{10}, \phi_{10})^\top = (0, 0.2)^\top$, the delay parameter is $d_0 = 1$, the threshold parameter is $\mu_0 = 0$, the ratio of sampling frequencies is $m \in \{3, 12\}$, and the error term is $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. We consider two cases for $\boldsymbol{\beta}_{20} = (\alpha_{20}, \phi_{20})^\top$:

Case 1 (non-identification): $\boldsymbol{\beta}_{20} = (0, 0.2)^\top$, in which case the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20}$ is true and $\boldsymbol{\gamma}_0 = (d_0, \mu_0)^\top$ is not identified.

Case 2 (strong identification): $\boldsymbol{\beta}_{20} = (0.35, 0.55)^\top$, in which case H_0^* is false and $\boldsymbol{\gamma}_0$ is strongly identified in the sense of Andrews and Cheng (2012).

Suppose that the DGP of the high frequency threshold variable x^* is AR(1):

$$x^*_t = \psi_0 x^*_{t-\frac{1}{m}} + \nu^*_t, \qquad \nu^*_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1), \qquad t \in \mathbb{H}.$$

Assume that $\{\epsilon_t\}_{t \in \mathbb{L}}$ and $\{\nu^*_t\}_{t \in \mathbb{H}}$ are mutually independent sequences. The AR coefficient of x^* is chosen from $\psi_0 \in \{0.3, 0.9\}$.

We fit the regular Midastar model with lag length $p = 1$ to generated samples:

$$y_t = \begin{cases} \alpha_1 + \phi_1 y_{t-1} + u_t & \text{if } x^*_{t-\frac{d}{m}} < \mu, \\ \alpha_2 + \phi_2 y_{t-1} + u_t & \text{if } x^*_{t-\frac{d}{m}} \geq \mu, \qquad t \in \mathbb{L}. \end{cases} \tag{27}$$

The choice space of the delay parameter d is $\mathcal{D} = \{1, 2, 3\}$. Note that the value of d is in terms of high frequency time periods. The choice space of μ is given by (12) with $\kappa = 0.7$, which is in accordance with the suggestion of Andrews (1993).

We perform the wild-bootstrap exp-Wald and exp-LM tests for the no-threshold-effect hypothesis H_0^*, as described in Section 3.2. When computing actual and bootstrap test statistics, we use the heteroscedasticity-robust covariance matrix estimator, although the true error term ϵ is homoscedastic. Rejection frequencies of each test are computed across $J = 1000$ Monte Carlo samples of size $n \in \{125, 250, 500, 1000\}$, where the nominal size is $a = 0.05$ and the number of bootstrap samples is $B = 500$.

To gauge the effects of temporal aggregation, we average the high frequency threshold variable x^*. We then fit a single-frequency TAR model to $\{y_t, x_t\}_{t \in \mathbb{L}}$:

$$y_t = \begin{cases} \alpha_1 + \phi_1 y_{t-1} + u_t & \text{if } x_{t-d} < \mu, \\ \alpha_2 + \phi_2 y_{t-1} + u_t & \text{if } x_{t-d} \geq \mu, \qquad t \in \mathbb{L}. \end{cases} \tag{28}$$

The choice space of the delay parameter is $\mathcal{D} = \{1, 2, 3\}$. Note that the value of d is in terms of low frequency time periods. Other configurations such as (κ, J, n, a, B) are all the same as in the Midastar simulation. Based on model (28), we perform the bootstrap exponential tests for the no-threshold-effect hypothesis H_0^*.

In Case 1 (non-identification), the threshold variable x^* is irrelevant under DGP (26), hence model (28) is correctly specified although its two-regime structure is redundant. In Case 2 (strong identification), a regime change at time $t \in \mathbb{H}$ is triggered by $x^*_{t-1/m}$, and it cannot be captured by model (28) with averaged x. Model (28) is therefore misspecified relative to DGP (26). Thus, we expect a typical symptom of spurious non-threshold effects that the TAR-based tests for H_0^* should have lower power than the Midastar-based tests.

How much power we lose after aggregating x^* should depend on two quantities: the persistence of the threshold variable ψ_0 and the ratio of sampling frequencies m. When ψ_0 is larger or m is smaller, temporal aggregation should incur the smaller loss of information. Hence, the power loss due to temporal aggregation is expected to be

smallest for $(\psi_0, m) = (0.9, 3)$ and largest for $(\psi_0, m) = (0.3, 12)$.

Finally, we compare the out-of-sample forecast performance of the Midastar model (27) and the aggregated TAR model (28). The DGP is kept the same, but we consider the larger sample sizes $\bar{n} \in \{250, 500, 1000, 2000\}$ so that the difference between the two models is sufficiently salient. For each model, we perform the rolling window one-step ahead prediction. The window size is fixed at $n = 0.8\bar{n}$, and the number of windows is $T = \bar{n} - n = 0.2\bar{n}$. Note that the prediction horizon $h = 1$ is in terms of low frequency time periods.

We perform the asymptotic Diebold-Mariano test as described in Section 3.3. The null hypothesis H_0^{eq} is that the Midastar-based and TAR-based forecasts are equally accurate. We consider two alternative hypotheses: H_1^{eq} states that the two forecasts have different accuracies (two-sided test); H_1^{midas} states that the Midastar forecast is more accurate than the TAR forecast (one-sided test).

In Case 1, there are no threshold effects and H_0^{eq} is true. In Case 2, there are threshold effects and H_1^{eq} and H_1^{midas} are true. Hence, our conjecture is that the probability of rejecting H_0^{eq} in favor of H_1^{eq} and H_1^{midas} should be sufficiently low in Case 1 and high in Case 2. Similar to the in-sample analysis, the superiority of Midastar relative to TAR is expected to be least salient for $(\psi_0, m) = (0.9, 3)$ and most salient for $(\psi_0, m) = (0.3, 12)$.

5.1.2 Simulation results

In Table 1, we report rejection frequencies of the Midastar-based bootstrap tests for the no-threshold-effect hypothesis H_0^*. In Case 1, the empirical size of the exp-Wald and exp-LM tests converges to the nominal size $a = 0.05$ as n grows, confirming Theorem 2.(i). The exp-Wald test rejects the correct null hypothesis too often when $n \leq 250$, although the over-rejections vanish as n grows. Taking $(\psi_0, m) = (0.3, 3)$ as an example, the empirical size of the exp-Wald test is $\{0.132, 0.073, 0.058, 0.047\}$ for $n \in \{125, 250, 500, 1000\}$, respectively. The exp-LM test, by contrast, achieves accurate empirical size for all sample sizes considered. For $(\psi_0, m) = (0.3, 3)$, the empirical size of the exp-LM test is $\{0.036, 0.036, 0.042, 0.039\}$. Hence, it is advised to use the exp-LM test instead of the exp-Wald test to better control the type-I error rate in small samples.

In Case 2, the empirical power of the exp-Wald and exp-LM tests converges to 1 as n grows, confirming Theorem 2.(ii). Focusing on $(\psi_0, m) = (0.3, 3)$, the empirical power of the exp-LM test is $\{0.501, 0.924, 1.000\}$ for $n \in \{125, 250, 500\}$, respectively

Table 1: Rejection frequencies of the bootstrap exponential tests for the no-threshold-effect hypothesis H_0^* (regular Midastar model)

			$\psi_0 = 0.3$				$\psi_0 = 0.9$			
			$m = 3$		$m = 12$		$m = 3$		$m = 12$	
Case	H_0^*	n	Wald	LM	Wald	LM	Wald	LM	Wald	LM
1	true	125	0.132	0.036	0.093	0.035	0.168	0.035	0.129	0.042
1	true	250	0.073	0.036	0.080	0.046	0.098	0.046	0.074	0.042
1	true	500	0.058	0.042	0.067	0.056	0.064	0.039	0.066	0.052
1	true	1000	0.047	0.039	0.056	0.048	0.048	0.035	0.068	0.053
2	false	125	0.711	0.501	0.718	0.590	0.761	0.558	0.801	0.625
2	false	250	0.949	0.924	0.960	0.937	0.944	0.916	0.973	0.956
2	false	500	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
2	false	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The DGP for y is regular Midastar, the DGP for x^* is AR(1), and the model is regular Midastar. The no-threshold-effect hypothesis H_0^* is true in Case 1 and false in Case 2. We report the rejection frequencies of the bootstrap exponential tests for H_0^*, where the nominal size is $a = 0.05$.

(Table 1). The empirical size and power of the exp-LM test are almost unchanged when the ratio of sampling frequencies increases from 3 to 12. Robustness to m is a desirable feature of Midastar, as we emphasized in Section 2.2. Overall, the simulation results in Table 1 are all reasonable and we can conclude that the exp-LM test achieves desired size and power properties in small and large samples.

In Table 2, we report rejection frequencies of the TAR-based bootstrap exp-LM test for H_0^*. The exp-Wald test is omitted since we know from the Midastar scenario that it has size distortions in small samples. In Case 1, the empirical size of the exp-LM test is sufficiently close to 5% for all cases considered. The empirical size with $(\psi_0, m, n) = (0.3, 3, 250)$ is 0.037, for instance.

In Case 2, aggregating x^* makes the exp-LM test much less powerful. The bootstrap tests with the aggregated threshold variable often fail to detect threshold effects in the underlying Midastar process, a common symptom of spurious non-threshold effects. When $(\psi_0, m, n) = (0.9, 3, 250)$, the empirical power of the exp-LM test is 0.916 for Midastar and 0.464 for TAR (Tables 1-2). When we raise m from 3 to 12 or lower ψ_0 from 0.9 to 0.3, the exp-LM test with the aggregated x has almost no power (Table 2). These results confirm our conjecture that the adverse impacts of tempo-

Table 2: Rejection frequencies of the bootstrap exponential LM tests for the no-threshold-effect hypothesis H_0^* (aggregated TAR model)

	Case 1: H_0^* is true				Case 2: H_0^* is false			
	$\psi_0 = 0.3$		$\psi_0 = 0.9$		$\psi_0 = 0.3$		$\psi_0 = 0.9$	
n	$m = 3$	$m = 12$	$m = 3$	$m = 12$	$m = 3$	$m = 12$	$m = 3$	$m = 12$
125	0.026	0.020	0.018	0.032	0.026	0.033	0.181	0.023
250	0.037	0.047	0.044	0.044	0.043	0.036	0.464	0.053
500	0.039	0.048	0.037	0.047	0.043	0.037	0.831	0.062
1000	0.033	0.040	0.052	0.053	0.055	0.035	0.993	0.137

The DGP for y is regular Midastar, the DGP for x^* is AR(1), and the model is TAR with x being averaged. The no-threshold-effect hypothesis H_0^* is true in Case 1 and false in Case 2. We report the rejection frequencies of the bootstrap exp-LM test for H_0^*, where the nominal size is $a = 0.05$.

ral aggregation on the test power are particularly large when the ratio of sampling frequencies is large or the original threshold variable x^* is transitory.

In Table 3, we report rejection frequencies of the asymptotic DM test for the equal predictive accuracy hypothesis H_0^{eq}. For Case 1, the empirical size of the DM test is sufficiently close to 5% for all cases considered, indicating that the type-I error rate is well controlled. Taking $(\psi_0, m, \bar{n}) = (0.3, 3, 250)$ as an example, the empirical probability of rejecting H_0^{eq} in favor of H_1^{eq} is 0.042.

For Case 2, the DM test exhibits power approaching 1 for all scenarios considered. It requires thousands of observations for the power to reach 1, however. When $(\psi_0, m, \bar{n}) = (0.9, 3, 2000)$, the empirical rejection probability in favor of H_1^{eq} is 0.701 (Table 3). The slow convergence of the empirical power to 1 suggests that it is quite hard to detect the subtle difference in the predictive ability between Midastar and TAR. When we lower ψ_0 from 0.9 to 0.3, the empirical power increases to 0.828. Alternatively, when we raise m from 3 to 12, the power increases to 0.833. These power improvements are consistent with our previous findings that Midastar is robust to (ψ_0, m) whereas TAR is negatively affected by small ψ_0 or large m.

Not surprisingly, the one-sided test (H_1^{midas}) has higher power than the two-sided test (H_1^{eq}) for each combination of $(\psi_0, m, \bar{n})$. In Case 2 with $(\psi_0, m, \bar{n}) = (0.3, 3, 1000)$, the rejection probability of H_0^{eq} is 0.505 for H_1^{eq} and 0.647 for H_1^{midas} (Table 3). In summary, the size and power properties of the DM test are all reasonable, although the power increases rather slowly as the sample size grows.

Table 3: Rejection frequencies of the Diebold-Mariano test for the equal predictive accuracy hypothesis H_0^{eq} (regular Midastar vs. aggregated TAR)

		Case 1: H_0^{eq} is true				Case 2: H_0^{eq} is false			
		$\psi_0 = 0.3$		$\psi_0 = 0.9$		$\psi_0 = 0.3$		$\psi_0 = 0.9$	
H_1	$\bar{n}$	$m = 3$	$m = 12$	$m = 3$	$m = 12$	$m = 3$	$m = 12$	$m = 3$	$m = 12$
H_1^{eq}	250	0.042	0.048	0.054	0.047	0.139	0.142	0.093	0.109
H_1^{eq}	500	0.043	0.039	0.048	0.041	0.276	0.267	0.177	0.230
H_1^{eq}	1000	0.037	0.041	0.040	0.037	0.505	0.520	0.368	0.523
H_1^{eq}	2000	0.033	0.048	0.036	0.043	0.828	0.851	0.701	0.833
H_1^{midas}	250	0.041	0.053	0.063	0.054	0.211	0.234	0.158	0.172
H_1^{midas}	500	0.048	0.051	0.046	0.048	0.395	0.404	0.286	0.366
H_1^{midas}	1000	0.050	0.051	0.048	0.044	0.647	0.644	0.501	0.644
H_1^{midas}	2000	0.043	0.052	0.049	0.045	0.907	0.922	0.791	0.890

The DGP for y is regular Midastar, the DGP for x^* is AR(1), and the model is either regular Midastar or aggregated TAR. The equal predictive accuracy hypothesis H_0^{eq} is true in Case 1 and false in Case 2. The one-step ahead forecasts are compared by the asymptotic Diebold-Mariano test. H_1^{eq}: The two forecasts have different accuracies. H_1^{midas}: The Midastar forecast is more accurate than the TAR forecast. Rejection frequencies of the DM test are reported, where the nominal size is $a = 0.05$.

5.1.3 Further discussions

We overview additional simulations in this section, and collect complete results and discussions in the separate supplemental material. The extended simulation is more general than the main simulation in many directions. First, we consider local alternative hypotheses (23) 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 and strong identification elaborated in the main simulation. We find that the Midastar-based bootstrap tests for the no-threshold-effect hypothesis H_0^* have power approaching 1 under semi-strong identification but not under weak identification, verifying Theorem 2.(ii).

Second, we examine the performance of the profiling estimator $\widehat{\boldsymbol{\theta}}$. For the Midastar model, the performance of $\widehat{\boldsymbol{\theta}}$ is all consistent with Theorem 1. For the TAR model with the aggregated threshold variable, $\widehat{\boldsymbol{\beta}}_1$ is positively biased and $\widehat{\boldsymbol{\beta}}_2$ is negatively biased. The probability limits of $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ are nearly identical to each other, which

explains the low power of the TAR-based bootstrap tests for H_0^* (Table 2).

Third, when testing the equal predictive accuracy hypothesis H_0^{eq}, we add an alternative hypothesis H_1^{tar} that the TAR forecast is more accurate than the Midastar forecast. Given our DGP, H_1^{tar} is always false and we expect that the Diebold-Mariano test should reject H_0^{eq} in favor of H_1^{tar} with negligibly small probability. In fact, the rejection frequencies are roughly 5% or lower for all cases considered. We found in the main simulation that the probability for the DM test to correctly choose Midastar over TAR is not necessarily high. It is a desired result, however, that the probability of incorrectly choosing TAR over Midastar is almost zero.

5.2 Reverse Midastar

In this section, we perform Monte Carlo simulations for the reverse Midastar model. Suppose that the DGP of a high frequency target variable y^* is the reverse Midastar with $p = 1$:

$$y_t^* = \begin{cases} \alpha_{10} + \phi_{10} y^*_{t-\frac{1}{m}} + \epsilon_t^* & \text{if } \ x_{\lceil t \rceil - d_0} < \mu_0, \\ \alpha_{20} + \phi_{20} y^*_{t-\frac{1}{m}} + \epsilon_t^* & \text{if } \ x_{\lceil t \rceil - d_0} \geq \mu_0, \qquad t \in \mathbb{H}, \end{cases} \tag{29}$$

where $\alpha_{10} = \alpha_{20} = 0$, $d_0 = 1$, $\mu_0 = 0$, and $\epsilon_t^* \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. The AR(1) parameters are $\phi_{10} = 0.2$ and $\phi_{20} \in \{0.2, 0.8\}$. Threshold effects are absent when $\phi_{20} = 0.2$ and present when $\phi_{20} = 0.8$. The DGP of the low frequency threshold variable x is AR(1):

$$x_t = 0.4 x_{t-1} + \nu_t, \qquad \nu_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \qquad t \in \mathbb{L}.$$

Assume that $\{\epsilon_t^*\}_{t \in \mathbb{H}}$ and $\{\nu_t\}_{t \in \mathbb{L}}$ are mutually independent sequences. The low frequency sample size is $n \in \{40, 80, 160\}$, and the ratio of sampling frequencies is $m \in \{3, 12\}$. Note that the sample size of the target variable y^* is $N = mn$. These values mimic a variety of empirical applications in economics and finance. The case with $(n, m) = (40, 3)$, for instance, matches monthly y^* and quarterly x of 10 years; the case with $(n, m) = (80, 12)$ approximates weekly y^* and quarterly x of 20 years.

The reverse Midastar model with $p = 1$ is specified as

$$y_t^* = \begin{cases} \alpha_1 + \phi_1 y^*_{t-\frac{1}{m}} + u_t^* & \text{if } \ x_{\lceil t \rceil - d} < \mu, \\ \alpha_2 + \phi_2 y^*_{t-\frac{1}{m}} + u_t^* & \text{if } \ x_{\lceil t \rceil - d} \geq \mu, \qquad t \in \mathbb{H}. \end{cases} \tag{30}$$

The space of d is $\mathcal{D} = \{1, 2, 3\}$, and the space of μ is $\mathcal{X} = \{x_t\}_{t \in \mathbb{L}}$. The space of $\boldsymbol{\gamma} = (d, \mu)^\top$ is restricted so that each regime accounts for at least 15% of the entire sample. We fit (30) to each of $J = 1000$ Monte Carlo samples generated from (29), and estimate $\boldsymbol{\beta} = (\alpha_1, \phi_1, \alpha_2, \phi_2)^\top$ and $\boldsymbol{\gamma}$ via profiling. Further, the bootstrap LM tests with $B = 500$ iterations are performed to test the no-threshold-effect hypothesis $H_0^* : (\alpha_{10}, \phi_{10}) = (\alpha_{20}, \phi_{20})$ at the 5% level. The LM test statistics conditional on the nuisance parameter $\boldsymbol{\gamma}$ are summarized into a single test statistic using the supremum, average, and exponential functions. (The Wald tests are omitted since, as in the regular Midastar simulation, they produce size distortions in small samples.)

We also fit the single-frequency TAR model after implementing the stock aggregation: $y_t = y_t^*$ for each $t \in \mathbb{L}$. (The averaging leads to similar results, and hence omitted.) The space of d is $\mathcal{D} = \{1, 2, 3\}$, and the space of μ is $\mathcal{X}$. We restrict the space of $\boldsymbol{\gamma}$ so that each regime accounts for at least 15% of the entire sample.

In Table 4, the rejection frequencies of the bootstrap LM tests for H_0^* are reported. The LM tests based on the reverse Midastar achieve accurate size and high power for both $m \in \{3, 12\}$. The empirical size of the exp-LM test with $m = 3$, for instance, is $\{0.019, 0.054, 0.046\}$ for $n \in \{40, 80, 160\}$, respectively. The empirical power of the exp-LM test is $\{0.553, 0.967, 1.000\}$ for $m = 3$ and $\{1.000, 1.000, 1.000\}$ for $m = 12$. The larger value of m results in the higher power for each n, due to the fact that the sample size of y^* (i.e., $N = mn$) becomes larger.

The aggregated TAR model leads to sharp size but low power (Table 4). The empirical power of the exp-LM test, for example, is $\{0.022, 0.137, 0.512\}$ for $m = 3$ and $\{0.011, 0.014, 0.039\}$ for $m = 12$. Temporal aggregation causes the greater amount of information loss as m increases. It is therefore reasonable that the empirical power decreases when m changes from 3 to 12. Indeed, the LM tests have no power for $m = 12$, an evidence of spurious non-threshold effects. In summary, the reverse Midastar model can accurately detect both the absence and presence of threshold effects, while the aggregated TAR model more frequently fails to detect threshold effects.

6 Empirical application

In this section, we perform two separate empirical applications to show practical values of the regular and reverse Midastar models. The regular Midastar is applied to crude oil market volatility in Section 6.1, and the reverse Midastar is applied to U.S. labor

Table 4: Rejection frequencies of the bootstrap LM tests for H_0^* (reverse Midastar and aggregated TAR)

		Reverse Midastar model				Aggregated TAR model			
		$\phi_{20} = 0.2$ (empirical size)		$\phi_{20} = 0.8$ (empirical power)		$\phi_{20} = 0.2$ (empirical size)		$\phi_{20} = 0.8$ (empirical power)	
Test	n	$m = 3$	$m = 12$	$m = 3$	$m = 12$	$m = 3$	$m = 12$	$m = 3$	$m = 12$
sup-LM	40	0.018	0.046	0.493	1.000	0.010	0.011	0.023	0.009
sup-LM	80	0.046	0.035	0.956	1.000	0.012	0.019	0.118	0.015
sup-LM	160	0.044	0.051	1.000	1.000	0.026	0.035	0.457	0.037
ave-LM	40	0.049	0.064	0.601	0.998	0.035	0.032	0.053	0.026
ave-LM	80	0.061	0.060	0.947	1.000	0.047	0.042	0.209	0.047
ave-LM	160	0.064	0.067	1.000	1.000	0.047	0.048	0.515	0.061
exp-LM	40	0.019	0.044	0.553	1.000	0.011	0.013	0.022	0.011
exp-LM	80	0.054	0.038	0.967	1.000	0.016	0.016	0.137	0.014
exp-LM	160	0.046	0.057	1.000	1.000	0.024	0.040	0.512	0.039

The DGP for y^* is reverse Midastar, the DGP for x is AR(1), and the model is either reverse Midastar or aggregated TAR. The no-threshold-effect hypothesis H_0^* is true when $\phi_{20} = 0.2$ and false when $\phi_{20} = 0.8$. We report the rejection frequencies of the bootstrap LM tests for H_0^*, where the nominal size is $a = 0.05$.

market indicators in Section 6.2.

6.1 Regular Midastar

In this section, we compare the in-sample and out-of-sample performance of the regular Midastar and aggregated TAR models. We analyze monthly realized volatility measures of the crude oil market, where the threshold variable is daily or monthly VIX. We explain the background and motivation in Section 6.1.1, present data and preliminary analysis in Section 6.1.2, and perform main analysis in Section 6.1.3.

6.1.1 Background and motivation

Crude oil is one of the most actively traded commodities in the world and a major source of the global energy supply. Modelling and predicting the crude oil market volatility are therefore of direct relevance for market participants' risk management

and central bankers' inflation targeting. *Realized volatility* (RV) is a common volatility measure of financial assets which was originally put forward by Andersen and Bollerslev (1998). Due to its theoretical validity and computational simplicity, RV has been analyzed and extended in various directions.[9]

Chen, Qiao, and Zhang (2022) [CQZ2022] fitted the TAR model to monthly RVs of the West Texas Intermediate (WTI) spot returns, where the threshold variable is monthly RVs of S&P 500 Index (SPX) returns. CQZ2022 detected significant threshold effects, implying that the crude oil volatility evolves asymmetrically when the overall financial market risk is low versus high. They also showed that TAR achieves significantly high out-of-sample performance than benchmark AR models.

While the SPX-RVs are certainly a compelling measure of the financial market risk, there are some alternative measures including CBOE Volatility Index (VIX). In fact, Chen, Liang, and Umar (2021) found significantly positive impacts of VIX on the crude oil market volatility in a linear predictive regression. VIX data are publicly available at the daily frequency, whereas direct calculation of daily RVs would require intraday returns. This motivates the usage of the regular Midastar model where the target variable is the monthly WTI-RVs as in CQZ2022 but the threshold variable is the daily VIX in contrast to CQZ2022.

6.1.2 Data and preliminary analysis

The sample period begins in January 1990 and ends in either December 2021 or 2023. The shorter sample period contains $n = 384$ months or equivalently 32 years, and it matches the sample period of CQZ2022. We also consider the longer sample period, which contains $n = 408$ months or 34 years, for a robustness check.

Crude oil prices are proxied by daily data of "Cushing, OK WTI Spot Price FOB (Dollars per Barrel)", which are publicly available at the website of the U.S. Energy Information Administration (ID: `RWTC`). To formulate monthly RVs from daily prices, we should deal with a time-varying ratio of sampling frequencies as the number of business days differs across months. Let $m_t \in \mathbb{N}$ be the number of business days in month $t \in \mathbb{L} = \{1, 2, \ldots, n\}$. Let $\mathbb{H}_t = \{t - 1 + 1/m_t, t - 1 + 2/m_t, \ldots, t\}$ be the set of days in month $t \in \mathbb{L}$. Let P_t be the WTI price on day $t \in \mathbb{H}$, where $\mathbb{H} = \cup_{t \in \mathbb{L}} \mathbb{H}_t$. The WTI return on day $t \in \mathbb{H}$ is defined as $r_t = (P_t - P_{t-1})/P_{t-1}$.

[9]Recent developments of the RV literature are overviewed in Ghysels and Marcellino (2018, Ch.14), among others.

There are three common measures of volatility at month $t \in \mathbb{L}$:

$$RV_t^{\pm} = \sum_{j \in \mathbb{H}_t} r_j^2, \quad RV_t^{+} = \sum_{j \in \mathbb{H}_t} r_j^2 \mathbf{1}(r_j > 0), \quad RV_t^{-} = \sum_{j \in \mathbb{H}_t} r_j^2 \mathbf{1}(r_j < 0).$$

The *(total) realized variance*, $RV^{\pm}$, takes into account positive and negative returns equally. The *realized semi-variances*, RV^{+} and RV^{-}, were originally put forward by Barndorff-Nielsen, Kinnebrock, and Shephard (2010) as proxies of "good volatility" and "bad volatility". By distinguishing good and bad volatilities, portfolio managers can handle upside and downside risks separately. In CQZ2022's out-of-sample analysis, the superiority of TAR relative to AR is more pronounced when predicting the bad volatility in crude oil markets than when predicting the good volatility.

For VIX, we retrieve "CBOE Volatility Index: VIX, Index, Daily, Not Seasonally Adjusted" (ID: `VIXCLS`) at Federal Reserve Economic Data (FRED). The regular Midastar model with a time-varying ratio of sampling frequencies is conceptually feasible, but it would add considerable burden to notation and coding. Following Ghysels, Hill, and Motegi (2020) and some other MIDAS applications, we average a few daily observations at the end of each month to make m constant. Let $\{\tilde{x}_t^*\}_{t \in \tilde{\mathbb{H}}}$ be an original high frequency variable with $\tilde{\mathbb{H}} = \cup_{t \in \mathbb{L}} \tilde{\mathbb{H}}_t$ and $\tilde{\mathbb{H}}_t = \{t - 1 + 1/m_t, t - 1 + 2/m_t, \ldots, t\}$. The tilde symbols are used to emphasize that the ratio of sampling frequencies is time-varying. Define $\mathbb{H} = \cup_{t \in \mathbb{L}} \mathbb{H}_t$, and $\mathbb{H}_t = \{t - 1 + 1/m, t - 1 + 2/m, \ldots, t\}$ for $t \in \mathbb{L}$, where $m = \min\{m_1, m_2, \ldots, m_n\}$. Note that m is kept constant at its minimum value. *Wrapping* operates as follows.

$$x_t^* = \begin{cases} \tilde{x}_t^* & \text{if } \ (t - \lfloor t \rfloor) m_{\lceil t \rceil} < m, \\ \frac{1}{m_{\lceil t \rceil} - m + 1} \sum_{j=0}^{m_{\lceil t \rceil} - m} \tilde{x}^*_{\lfloor t \rfloor + \frac{m+j}{m_{\lceil t \rceil}}} & \text{if } \ (t - \lfloor t \rfloor) m_{\lceil t \rceil} = m, \end{cases} \qquad t \in \mathbb{H}.$$

This operation produces a *wrapped* series $\{x_t^*\}_{t \in \mathbb{H}}$.

A special attention is required for September 2001, in which month the number of business days was unusually small ($m_t = 15$) due to four consecutive days of market closure following the September 11 attacks. If we simply picked $m = 15$, many high frequency observations would be averaged every month because of only one irregular month. We take an alternative approach of imputing the four missing values of VIX on September 11–14 (Tuesday–Friday) with 41.76, an observed value on Monday, September 17. This is an ad-hoc interpolation, but it raises the minimum value m from 15 to 18; November 1997 has 5 Saturdays, 5 Sundays, and 2 holidays hence

$m_t = 30 - 5 - 5 - 2 = 18$. Thus, we pick $m = 18$ and average the 18th through 23rd high frequency observations for each month (if available).

In Figure 1, we plot the monthly log-RVs of WTI and the daily $\ln \mathrm{VIX}$ from January 1990 through December 2023, where VIX is wrapped with $m = 18$. Several facts are salient from Figure 1. First, the time series paths of $\ln RV^{\pm}$, $\ln RV^{+}$, and $\ln RV^{-}$ are roughly similar to each other, if not identical. Second, the crude oil market volatility surged during the Gulf War in 1990–1991, the global financial crisis (GFC) in 2007–2009, the Chinese stock market turbulence in 2015–2016, the COVID-19 pandemic in 2020, and a few other periods of emergency. The COVID-19 pandemic had by far the largest impact on the crude oil volatility. Third, the impacts of GFC and COVID-19 on VIX are roughly as large as each other, which implies the crude oil market is more affected by the coronavirus pandemic than other financial markets.

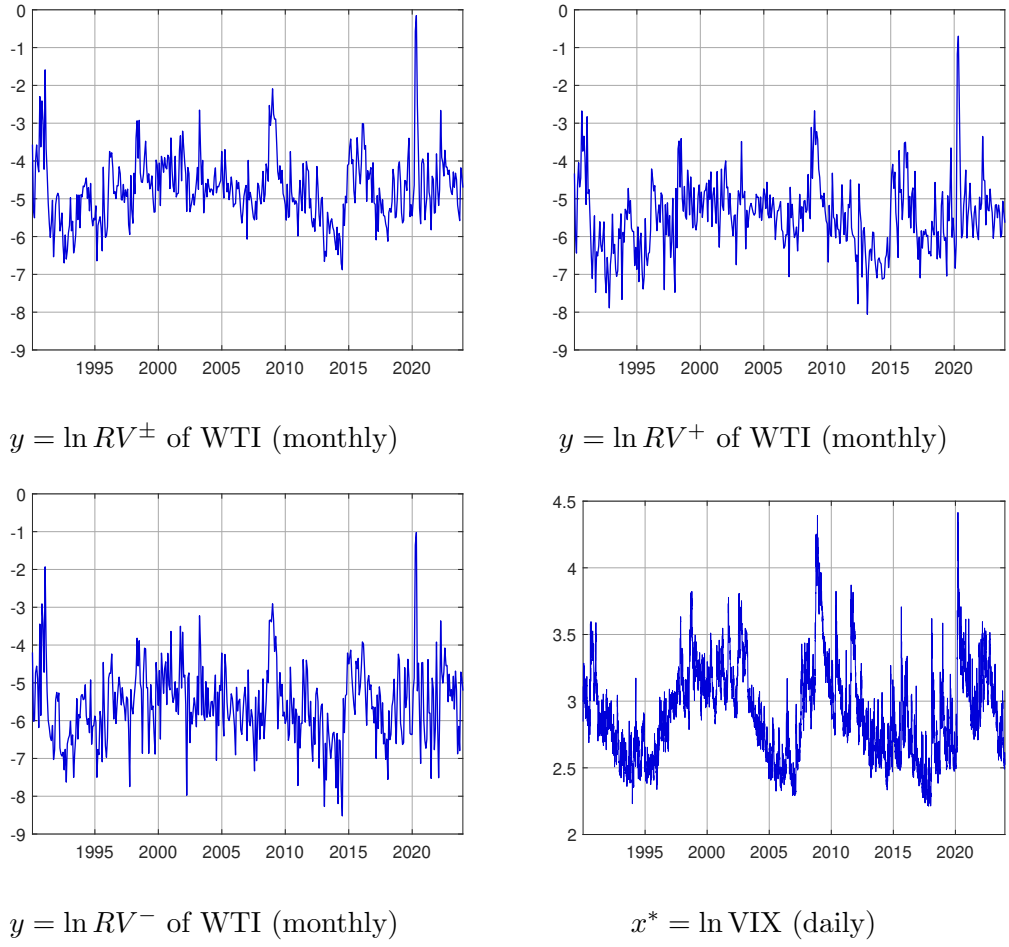
In Table 5, we report sample statistics of the target and threshold variables. Some insights can be drawn from Table 5. First, comparing the monthly $\ln RV^{\pm}$, $\ln RV^{+}$, and $\ln RV^{-}$ of WTI, the total volatility has the smallest standard deviation and the bad volatility has the largest standard deviation, while the order is reversed for kurtosis. These imply that, although the three RV measures look similar to each other in Figure 1, there are some differences in their statistical properties. Second, the daily $\ln \mathrm{VIX}$ and its monthly average have similar sample statistics, which indicates that averaging does not alter basic characteristics of VIX. Third, all five variables considered in Table 5 are positively skewed, suggesting that positive spikes are more dominant than negative spikes for the crude oil market volatility and VIX.

Table 5: Sample statistics of the target and threshold variables

variable	freq	mean	median	min	max	stdev	skew	kurt
$y = \ln RV^{\pm}$ (WTI)	month	−4.749	−4.817	−6.879	−0.153	0.896	0.841	5.676
$y = \ln RV^{+}$ (WTI)	month	−5.475	−5.499	−8.057	−0.698	0.968	0.649	5.185
$y = \ln RV^{-}$ (WTI)	month	−5.571	−5.591	−8.521	−1.022	1.024	0.432	4.475
$x^* = \ln \mathrm{VIX}$	day	2.910	2.878	2.213	4.415	0.348	0.632	3.386
$x = \mathrm{ave}(\ln \mathrm{VIX})$	month	2.910	2.869	2.303	4.136	0.335	0.631	3.315

WTI: West Texas Intermediate crude oil spot price. VIX: CBOE Volatility Index. The target variable y is the monthly log-RVs of WTI. The high frequency threshold variable x^* is the daily $\ln \mathrm{VIX}$, where VIX is wrapped with $m = 18$. The low frequency threshold variable x is the monthly average of x^*. Sample period: January 1990 – December 2023 ($n = 408$ months or $mn = 7344$ days).

Figure 1: Time series plots of the monthly log-RVs of WTI and the daily VIX



$y = \ln RV^{\pm}$ of WTI (monthly)

$y = \ln RV^{+}$ of WTI (monthly)

$y = \ln RV^{-}$ of WTI (monthly)

$x^* = \ln \text{VIX}$ (daily)

WTI: West Texas Intermediate crude oil spot price. VIX: CBOE Volatility Index. We plot the monthly $\ln RV^{\pm}$, $\ln RV^{+}$, and $\ln RV^{-}$ of WTI as well as the daily $\ln$ VIX, where VIX is wrapped with $m = 18$. Sample period: January 1990 – December 2023 ($n = 408$ months or $mn = 7344$ days).

6.1.3 Main empirical analysis

We fit the regular Midastar model (4) with lag length $p = 2$ to our data. The target variable y is the monthly log-RVs of WTI, and the high frequency threshold variable x^* is the daily $\ln$ VIX. We perform the wild-bootstrap exponential LM test for the no-threshold-effect hypothesis $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, where $\boldsymbol{\beta}_r = (\alpha_1, \phi_{r1}, \phi_{r2})^\top$ with $r \in \{1, 2\}$. We use the heteroscedasticity-robust covariance matrix estimator to address potential heteroscedasticity in the error term. The choice space for delay parameter is $\mathcal{D} = \{1, 2, \ldots, 9\}$ measured in terms of days. The choice space for the threshold parameter is given by (12) with $\kappa = 0.7$ so that each regime accounts for at least 15% of the entire sample on average. The number of bootstrap iterations is $B = 5000$. To

gauge the impacts of temporal aggregation, we also fit the TAR model (1) with $p = 2$ after averaging x^*. We then test H_0^* analogously, where the choice space for delay parameter is $\mathcal{D} = \{1, 2, 3\}$ measured in terms of months.

Resulting p-values of the bootstrap exp-LM test for H_0^* are reported in Table 6. For Midastar, H_0^* is rejected at the 1% level for both the shorter and longer sample periods, and for all three RV measures. This is overwhelming evidence for the existence of threshold effects. Statistical significance becomes weaker once x^* is averaged. H_0^* cannot be rejected at the 1% level for any of the six cases considered. In particular, we fail to reject H_0^* at the 10% level for $\ln RV^-$ in the shorter sample, a signal of spurious non-threshold effects.

Table 6: Bootstrap p-values of the exp-LM test for the no-threshold-effect hypothesis H_0^* (regular Midastar and aggregated TAR)

Sample ends in:	December 2021 ($n = 384$)			December 2023 ($n = 408$)		
Target variable y is:	$\ln RV^{\pm}$	$\ln RV^{+}$	$\ln RV^{-}$	$\ln RV^{\pm}$	$\ln RV^{+}$	$\ln RV^{-}$
Regular Midastar (x^*)	0.000^{***}	0.002^{***}	0.000^{***}	0.000^{***}	0.001^{***}	0.000^{***}
Aggregated TAR (x)	0.051^{*}	0.024^{**}	0.252	0.029^{**}	0.015^{**}	0.098^{*}

This table reports the bootstrap p-values of the exp-LM test for the no-threshold-effect hypothesis H_0^*, where the target variable y is the monthly log-RVs of WTI; the high frequency threshold variable x^* is the daily ln VIX; the low frequency threshold variable x is the monthly average of x^* ($m = 18$). The model is either regular Midastar or aggregated TAR with lag length $p = 2$. The sample period begins in January 1990 and ends in either December 2021 or 2023. The asterisks ***, **, * are put on the p-values when H_0^* is rejected at the 1%, 5%, 10% levels, respectively.

We next conduct the rolling window out-of-sample analysis. The entire sample size is either $\bar{n} = 384$ (January 1990 – December 2021) or $\bar{n} = 408$ (January 1990 – December 2023), as in the in-sample analysis. The window size is fixed at $n = \lfloor 0.8\bar{n} \rfloor$, which is 307 months for the shorter sample and 326 for the longer sample. We perform one-month-ahead prediction with either the constant-only, single-regime AR(2), TAR, or regular Midastar model. The constant-only and AR(2) models are added as benchmarks, and the TAR and Midastar models are specified in the same way as the in-sample analysis. We compute the root mean squared forecast error (RMSE) for each model, and implement the asymptotic Diebold-Mariano test to compare the predictive power of Midastar and TAR.

Empirical results are summarized in Table 7. For all sample periods and RV measures considered, the regular Midastar attains the smallest RMSE of the four

models. This is strong evidence that the out-of-sample performance has successfully improved by taking mixed frequency data into account. Focusing on the shorter sample with $y = \ln RV^{\pm}$, the RMSE is 1.040 for the constant-only model, 0.871 for AR, 0.865 for TAR, and 0.844 for Midastar.

Table 7: RMSEs of the out-of-sample prediction and the p-values of the Diebold-Mariano test for the equal predictive accuracy hypothesis H_0^{eq} (regular Midastar)

Sample ends in:	December 2021 ($N = 384$)			December 2023 ($N = 408$)		
Target variable y	$\ln RV^{\pm}$	$\ln RV^{+}$	$\ln RV^{-}$	$\ln RV^{\pm}$	$\ln RV^{+}$	$\ln RV^{-}$
$RMSE^{const}$	1.040	1.137	1.125	0.970	1.026	1.132
$RMSE^{ar}$	0.871	1.020	1.049	0.843	0.933	1.089
$RMSE^{tar}$	0.865	1.027	1.072	0.831	0.922	1.088
$RMSE^{midas}$	0.844	0.990	1.025	0.800	0.891	1.039
$\widehat{p}(H_1^{eq})$	0.328	0.268	0.080*	0.143	0.388	0.062*
$\widehat{p}(H_1^{midas})$	0.164	0.134	0.040**	0.072*	0.194	0.031**
$\widehat{p}(H_1^{tar})$	0.836	0.866	0.960	0.929	0.806	0.969

The target variable y is the monthly log-RVs of WTI; the high frequency threshold variable x^* is the daily ln VIX; the low frequency threshold variable x is the monthly average of x^* ($m = 18$). We report the root mean squared forecast errors of the constant-only, AR, TAR, and regular Midastar models. We also report the asymptotic p-values of the Diebold-Mariano test for Midastar vs. TAR. H_0^{eq}: The two models have an equal accuracy. H_1^{eq}: The two models have different accuracies. H_1^{midas}: Midastar is more accurate than TAR. H_1^{tar}: TAR is more accurate than Midastar. The asterisks ***, **, * are put on the p-values when H_0^{eq} is rejected at the 1%, 5%, 10% levels, respectively.

For the bad volatility RV^{-}, the two-sided DM test rejects the equal predictive accuracy hypothesis H_0^{eq} in favor of the regular Midastar at the 10% level. The corresponding p-values are 0.080 for the shorter sample and 0.062 for the longer sample (Table 7). When we replace the two-sided test with the one-sided test, the corresponding p-values improve to 0.040 and 0.031, rejecting H_0^{eq} more strongly at the 5% level. In addition, the one-sided test rejects H_0^{eq} at the 10% level for the total volatility $RV^{\pm}$ in the longer sample with the p-value being 0.072.

These empirical findings are relevant for market participants since predicting the crude oil market volatility plays a crucial role in financial risk management. It also matters for policymakers since a drastic change in crude oil prices is often associated with business cycle fluctuations and inflation. In summary, this empirical application demonstrates that the regular Midastar model is a promising tool for modelling and

forecasting financial indicators.

6.2 Reverse Midastar

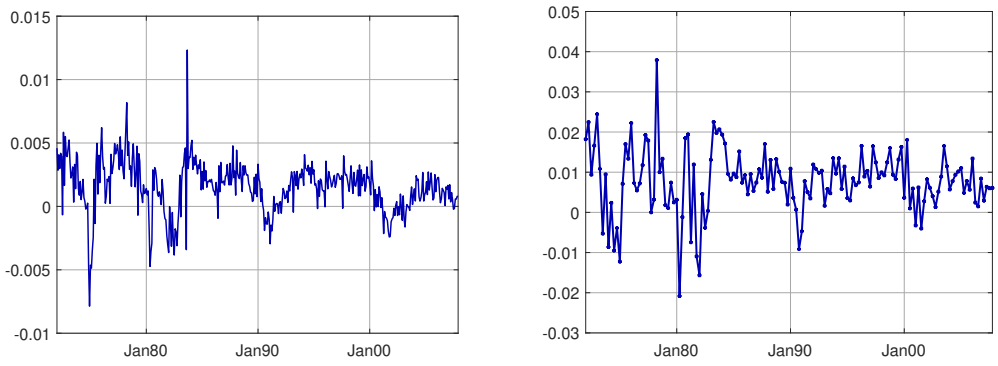
In this section, we perform an empirical application of the reverse Midastar model. It is well known that some macroeconomic indicators move asymmetrically over the business cycle. In particular, U.S. labor market indicators tend to contract deeply and sharply during recessions, and recover slowly and gradually during expansions; see Neftçi (1984), Hamilton (1989), McQueen and Thorley (1993), and Sichel (1993) for early contributions; see Ferraro (2018) and Pizzinelli, Theodoridis, and Zanetti (2020) for more recent studies from both theoretical and empirical perspectives.

To our best knowledge, threshold effects of the business cycle on the labor market fluctuations have always been inspected in the single-frequency framework. One of the most prominent proxies of the business cycle is real GDP, which is sampled quarterly. Major indicators of the labor market condition, such as the number of total nonfarm employees and the unemployment rate, are sampled monthly. Under the single-frequency framework, the labor market indicator needs to be aggregated into the quarterly level, which might hide the underlying threshold effects. This issue motivates the use of the reverse Midastar with $m = 3$, where the target variable is the monthly labor market indicator and the threshold variable is the quarterly real GDP.

Let Emp_t be the number of total nonfarm employees in the U.S. at month $t \in \mathbb{H}$ (seasonally adjusted). Emp is proxied with the series `PAYEMS` of FRED. Let GDP_t be the seasonally adjusted real GDP of the U.S. at quarter $t \in \mathbb{L}$. GDP is proxied with the series `GDPC1` of FRED. The sample period is January 1972 through December 2007 ($n = 144$ quarters and $mn = 432$ months). The sample period is terminated at the end of 2007 due to the extreme turbulence caused by the global financial crisis in 2008 and the COVID-19 pandemic in 2020.

Let $\Delta \ln Emp_t = \ln Emp_t - \ln Emp_{t-1/m}$ be the monthly employment growth. Let $\Delta \ln GDP_t = \ln GDP_t - \ln GDP_{t-1}$ be the quarterly GDP growth. In Figure 2, we plot $\{\Delta \ln Emp_t\}_{t\in\mathbb{H}}$ and $\{\Delta \ln GDP_t\}_{t\in\mathbb{L}}$. On the left panel, $\Delta \ln Emp$ marked -0.003 in August 1983 and then 0.012 in the next month; these large values reflect the sharp economic recovery from the prolonged recession in 1981-82. In Table 8, the sample statistics of the growth series are reported. The employment growth is negatively skewed, which is consistent with the existing literature (e.g., Ferraro, 2018).

Figure 2: Time series plots of the U.S. macroeconomic indicators



$y^* = \Delta \ln Emp$ (month) $\qquad$ $x = \Delta \ln GDP$ (quarter)

Emp = the number of total nonfarm employees. GDP = the real GDP. This figure plots the log-difference of Emp and GDP from January 1972 through December 2007.

Table 8: Sample statistics of the U.S. macroeconomic indicators

series	freq	size	mean	median	min	max	stdev	skew	kurt
$\Delta \ln Emp$	month	432	0.002	0.002	−0.008	0.012	0.002	−0.263	6.017
$\Delta \ln GDP$	quarter	144	0.008	0.008	−0.021	0.038	0.008	−0.281	5.017
ave($\Delta \ln Emp$)	quarter	144	0.002	0.002	−0.004	0.006	0.002	−0.716	4.019

$\Delta \ln Emp$ = monthly employment growth. $\Delta \ln GDP$ = quarterly GDP growth. ave($\Delta \ln Emp$) = quarterly average of $\Delta \ln Emp$. This table reports the sample size, mean, median, minimum, maximum, standard deviation, skewness, and kurtosis of each series.

The reverse Midastar model with $p = 4$ and $m = 3$ is specified as follows:

$$y_t^* = \begin{cases} \alpha_1 + \sum_{k=1}^4 \phi_{1k} y_{t-k/3}^* + u_t^* & \text{if } \; x_{\lceil t \rceil - d} < \mu, \\ \alpha_2 + \sum_{k=1}^4 \phi_{2k} y_{t-k/3}^* + u_t^* & \text{if } \; x_{\lceil t \rceil - d} \geq \mu, \qquad t \in \mathbb{H}, \end{cases}$$

where $y_t^* = \Delta \ln Emp_t$ is the monthly employment growth and $x_t = \Delta \ln GDP_t$ is the quarterly GDP growth. Regime 1 represents recession, while regime 2 represents expansion. The space of d is $\mathcal{D} = \{1, \ldots, 8\}$, allowing for at most two years of delay in the threshold effects of x on y^*. The space of μ is given by $\mathcal{X} = \{x_t\}_{t \in \mathbb{L}}$. We restrict the space of $\boldsymbol{\gamma} = (d, \mu)^\top$ so that each regime accounts for at least 15% of the entire sample.

Empirical results of the reverse Midastar are summarized as follows. The estimated threshold parameter is $\hat{\mu} = 0.006$. The share of the recession phase to the whole sample is 40.1%, indicating that the expansion phase is the majority. The estimated delay parameter is $\hat{d} = 2$ quarters, an intuitively plausible value. The bootstrap tests with $B = 5000$ iterations reject the no-threshold-effect hypothesis H_0^* at the 5% level; the bootstrap p-value is 0.036 for the sup-LM test and 0.029 for the exp-LM test. These results indicate that the persistence structures of the employment growth differ significantly between the recession and expansion.

For comparison, we fit the TAR model with $p = 4$ after taking the quarterly average of the monthly employment growth. The estimated delay parameter is $\hat{d} = 2$ quarters, which coincides with the reverse Midastar scenario. The bootstrap tests based on the aggregated TAR model cannot reject H_0^* at any conventional levels; the bootstrap p-value is 0.507 for the sup-LM test and 0.371 for the exp-LM test. Averaging y^* seems to have weaken the underlying threshold effects; another possibility is that the test has become less powerful as the sample size of the target variable decreased from $mn = 432$ to $n = 144$. In summary, the threshold effects of the real GDP on the total nonfarm employment can be captured by the reverse Midastar model but not by the aggregated TAR model.

7 Conclusion

In this paper, we have proposed the Midastar models by expanding the TAR model to the MIDAS context. Despite its considerable potential applications, this extension is not examined in the literature, and this paper initiates its development. Our framework allows the target variable y and the threshold variable x to be sampled at mixed frequencies. Specifically, y is observed less frequently than x in the regular Midastar model, while y is more frequently observed than x in the reverse Midastar model.

Midastar captures threshold effects accurately, while the standard TAR model with temporally aggregated data can lead to a biased conclusion, in particular spurious non-threshold effects. Besides, Midastar is free of parameter proliferation when the ratio of sampling frequencies between y and x takes a large value, since y is not directly regressed onto x. This is a notable advantage compared with existing MIDAS regression models. The parameters of interest are estimated via profiling, and the no-threshold-effect hypothesis H_0^* is tested via the wild bootstrap. We have established the uniform consistency and asymptotic normality of the proposed estimator as well

as the asymptotic validity of the bootstrap test, imposing much weaker regularity conditions than in the literature. In particular, we have addressed the non-stationarity problem in the reverse Midastar scenario by exploiting the blocking argument.

Our Monte Carlo simulations indicate that the proposed methods perform well in finite samples. We apply the regular Midastar model to monthly RV measures of WTI, in which the threshold variable is daily VIX. We also apply the reverse Midastar model to monthly U.S. employment growth, in which the threshold variable is quarterly U.S. real GDP growth. For both empirical examples, the Midastar approaches detect significant threshold effects, and the statistical significance weakens or disappears after temporal aggregation. These results highlight that Midastar is useful for modelling and forecasting financial and macroeconomic indicators.

References

Aït-Sahalia, Y., and J. Y. Park (2012): "Stationarity-Based Specification Tests for Diffusions When the Process is Nonstationary," *Journal of Econometrics*, 169, 279–292.

Andersen, T. G., and T. Bollerslev (1998): "Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies," *The Journal of Finance*, 53, 219–265.

Andreou, E., P. Gagliardini, and E. Ghysels (2010): "Regression models with mixed sampling frequencies," *Journal of Econometrics*, 158, 246–261.

Andreou, E., P. Gagliardini, E. Ghysels, and M. Rubin (2019): "Inference in Group Factor Models With an Application to Mixed-Frequency Data," *Econometrica*, 87, 1267–1305.

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.

Barndorff-Nielsen, O. E. (1983): "On a formula for the distribution of the maximum likelihood estimator," *Biometrika*, 70, 343–365.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shephard (2010): "Measuring Downside Risk—Realized Semivariance," in *Volatility and Time Series*

Econometrics: Essays in Honor of Robert Engle, ed. by T. Bollerslev, J. Russell, and M. Watson. Oxford University Press.

Bosq, B. (1998): "Nonparametric Statistics for Stochastic Processes: Estimation and Prediction," New York: Springer-Verlag.

Brännäs, K., and H. Ohlsson (1999): "Asymmetric time series and temporal aggregation," *The Review of Economics and Statistics*, 81, 341–344.

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.

Chen, X., L. P. Hansen, and M. Carrasco (2010): "Nonlinearity and Temporal Dependence," *Journal of Econometrics*, 155, 155–169.

Chen, Y., G. Qiao, and F. Zhang (2022): "Oil price volatility forecasting: Threshold effect from stock market volatility," *Technological Forecasting & Social Change*, 180, #121704.

Chen, Z., C. Liang, and M. Umar (2021): "Is investor sentiment stronger than VIX and uncertainty indices in predicting energy volatility?," *Resources Policy*, 74, #102391.

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.

Doukhan, P., and S. Louhichi (1999): "A new weak dependence condition and applications to moment inequalities," *Stochastic Processes and Their Applications*, 84(2), 313–342.

Ferraro, D. (2018): "The asymmetric cyclical behavior of the U.S. labor market," *Review of Economic Dynamics*, 30, 145–162.

Foroni, C., M. Marcellino, and C. Schumacher (2015): "Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials," *Journal of Royal Statistical Society, Series A*, 178, 57–82.

Foroni, C., M. Marcellino, and Stevanović (2019): "Mixed-frequency models with moving-average components," *Journal of Applied Econometrics*, 34, 688–706.

Galeano, P., and D. Peña (2007): "Improved model selection criteria for SETAR time series models," *Journal of Statistical Planning and Inference*, 137, 2802–2814.

Ghysels, E. (2016): "Macroeconomics and the reality of mixed frequency data," *Journal of Econometrics*, 193, 294–314.

Ghysels, E., J. B. Hill, and K. Motegi (2016): "Testing for Granger causality with mixed frequency data," *Journal of Econometrics*, 192, 207–230.

——— (2020): "Testing a large set of zero restrictions in regression models, with an application to mixed frequency Granger causality," *Journal of Econometrics*, 218, 633–654.

Ghysels, E., and M. Marcellino (2018): *Applied Economic Forecasting Using Time Series Methods*. Oxford University Press.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004): "The MIDAS Touch: Mixed Data Sampling Regression Models," Working paper at UCLA.

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

Götz, T. B., A. Hecq, and S. Smeekes (2016): "Testing for Granger causality in large mixed-frequency VARs," *Journal of Econometrics*, 193, 418–432.

Granger, C. W. J., and T.-H. Lee (1999): "The effect of aggregation on nonlinearity," *Econometric Reviews*, 18, 259–269.

Guay, A., and A. Maurin (2015): "Disaggregation methods based on MIDAS regression," *Economic Modelling*, 50, 123–129.

Hall, P., and Q. Yao (2003): "Inference in ARCH and GARCH Models with Heavy-Tailed Errors," *Econometrica*, 71, 285–317.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

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

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

Hong, S. Y., and O. B. Linton (2020): "Nonparametric estimation of infinite order regression and its application to the risk-return tradeoff," *Journal of Econometrics*, 219, 389–424.

Hong, Y., and Y.-J. Lee (2013): "A loss function approach to model specification testing and its relative efficiency," *Annals of Statistics*, 41, 1166–1203.

Ibragimov, I., and I. V. Linnik (1971): "Independent and Stationary Sequences of Random variables," Groningen: Wolters-Noordhoff.

Li, J., V. Todorov, G. Tauchen, and R. Chen (2017): "Mixed-scale jump regressions with bootstrap inference," *Journal of Econometrics*, 201, 417–432.

Marcellino, M., and C. Schumacher (2010): "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP," *Oxford Bulletin of Economics and Statistics*, 72, 518–550.

Marcellino, M., and V. Sivec (2016): "Monetary, fiscal and oil shocks: Evidence based on mixed frequency structural FAVARs," *Journal of Econometrics*, 193, 335–348.

McQueen, G., and S. Thorley (1993): "Asymmetric business cycle turning points," *Journal of Monetary Economics*, 31, 341–362.

Motegi, K., J. W. Dennis, and S. Hamori (2023): "Conditional Threshold Autoregression (CoTAR)," SSRN Working Paper No. 3960058.

Neftçi, S. N. (1984): "Are Economic Time Series Asymmetric over the Business Cycle?," *Journal of Political Economy*, 92, 307–328.

Paya, I., and D. A. Peel (2006): "Temporal aggregation of an ESTAR process: Some implications for purchasing power parity adjustment," *Journal of Applied Econometrics*, 21, 655–668.

Peligrad, M. (1986): "Recent advances in the central limit theorems and its weak invariance principle for mixing sequences of random variables (a survey)," in *Contemporary Mathematics: Statistical Inference from Stochastic Processes*, ed. by E. Eberlein, and e. M. S. Taqqu, pp. 193–223. Boston: Birkhauser.

Pizzinelli, C., K. Theodoridis, and F. Zanetti (2020): "State dependence in labor market fluctuations," *International Economic Review*, 61, 1027–1072.

Rio, E. (2017): "Asymptotic Theory of Weakly Dependent Random Processes," Springer, Heidelberg.

Schorfheide, F., D. Song, and A. Yaron (2018): "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach," *Econometrica*, 86, 617–654.

Sichel, D. E. (1993): "Business cycle asymmetry: A deeper look," *Economic Inquiry*, 31, 224–236.

Silvestrini, A., and D. Veredas (2008): "Temporal aggregation of univariate and multivariate time series models: A survey," *Journal of Economic Surveys*, 22, 458–497.

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

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

Tsay, R. S., and R. Chen (2019): *Nonlinear Time Series Analysis*. John Wiley & Sons, Inc.

Wong, C. S., and W. K. Li (1998): "A note on the corrected Akaike information criterion for threshold autoregressive models," *Journal of Time Series Analysis*, 19, 113–124.

Yang, L., and C. Zhang (2022): "Threshold mixed data sampling models with a covariate-dependent threshold," *Applied Economics Letters*, https://doi.org/10.1080/13504851.2022.2081657.

Appendix: Proofs of theorems

In this appendix, we prove Lemma 1 and Theorems 1-3, which are the key results that underpin the Midastar models. We begin by reviewing and adding some notations; as before, unless stated otherwise, we present notations for the regular Midastar, for example using u_t. Extensions to the reverse Midastar framework are straightforward, for example, by employing U_t instead of u_t for the error process.

Recall from (18) that the no-threshold-effect hypothesis is expressed as H_0^* : $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. The regression score conditional on the nuisance parameter $\boldsymbol{\gamma}$ is given by $\boldsymbol{s}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_t(\boldsymbol{\gamma})u_t$, and the estimated regression score conditional on $\boldsymbol{\gamma}$ is given by $\widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_t(\boldsymbol{\gamma})\widehat{u}_t(\boldsymbol{\gamma})$, where $\boldsymbol{Z}_t(\boldsymbol{\gamma})$ is the regressor vector defined in (10) and $\widehat{u}_t(\boldsymbol{\gamma})$ is the residual defined in (14). For arbitrary $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \Gamma$, define some matrices conditional on the sample:

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

$$\widehat{\boldsymbol{S}}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{1}{n}\sum_{t\in\mathbb{L}} \widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma}_1)\widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma}_2)^\top, \qquad \boldsymbol{M}_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{1}{n}\sum_{t\in\mathbb{L}} \boldsymbol{Z}_t(\boldsymbol{\gamma}_1)\boldsymbol{Z}_t(\boldsymbol{\gamma}_2)^\top.$$

We will sometimes abbreviate $\widehat{\boldsymbol{V}}_n(\boldsymbol{\gamma}) = \widehat{\boldsymbol{V}}_n(\boldsymbol{\gamma}, \boldsymbol{\gamma})$, $\widehat{\boldsymbol{S}}_n(\boldsymbol{\gamma}) = \widehat{\boldsymbol{S}}_n(\boldsymbol{\gamma}, \boldsymbol{\gamma})$, and $\boldsymbol{M}_n(\boldsymbol{\gamma}) = \boldsymbol{M}_n(\boldsymbol{\gamma}, \boldsymbol{\gamma})$ when appropriate. 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 (22). Recall that $\mathcal{G}(\boldsymbol{\gamma})$ is a mean zero Gaussian process with covariance kernel $\boldsymbol{V}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$.

A.1 Proof of Lemma 1 (m-fold repetition of x)

Proof. We note that $X_1 = x_{\lceil 1/m \rceil} = x_1,\ X_2 = x_{\lceil 2/m \rceil} = x_1, \ldots, X_m = x_{\lceil m/m \rceil} = x_1$ by definition, whereas $X_{m+1} = x_{\lceil (m+1)/m \rceil} = x_2$. Hence, we have $(X_1, X_2, \ldots, X_m) = (x_1, x_1, \ldots, x_1)$ and $(X_2, X_3, \ldots, X_{m+1}) = (x_1, x_1, \ldots, x_2)$, almost surely. Since x_t is not degenerate by the bounded density assumption, the two aforementioned vectors cannot have the same distribution, showing that the distributional equivalence does not hold under a shift in time. Thus, $\{X_t\}_{t=1}^N$ is non-stationary.

On the other hand, since $X_t = x_{\lceil t/m \rceil}$, we have that

$$\begin{aligned}
\sigma(X_u : u \le s) \ &\subseteq\ \sigma\big(x_v : v \le \lceil s/m \rceil\big), \\
\sigma(X_u : u \ge s+n) \ &\subseteq\ \sigma\big(x_v : v \ge \lceil (s+n)/m \rceil\big),
\end{aligned}$$

where $\sigma(Z)$ denotes the σ-field of events generated by Z. Since m is fixed and finite, there exists some constant $C > 0$ and integer n_0 such that for all $n > n_0$,

$$\alpha_X(n) \ \le\ \alpha_x\big(\lceil (s+n)/m \rceil - \lceil s/m \rceil\big) \ \le\ \alpha_x\big(\lfloor n/m \rfloor\big) \ \le\ C\ \big(\lfloor n/m \rfloor\big)^{-A},$$

The upper bound tends to 0 as $n \to \infty$ because $(\lfloor n/m \rfloor)^{-A} \to 0$, from which $\alpha_X(n) \to 0$ as $n \to \infty$. Thus, we conclude that $\{X_t\}_{t=1}^N$ is α-mixing with the same mixing rate as $\{x_t\}$. □

A.2 Proof of Theorem 3 (Reverse Midastar)

Proof. We first prove Theorem 3 ahead of other theorems because the proof of Theorem 1 can be viewed as a special case of it, apart from minor notational changes due to operating in the reverse framework.

Recall that we have

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left(\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right)^{-1} \left(\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) Y_t \right), \quad \boldsymbol{Z}_t(\boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{1}(X_{t-D} < \mu) \boldsymbol{z}_t \\ \mathbf{1}(X_{t-D} \ge \mu) \boldsymbol{z}_t \end{bmatrix},$$

with $\boldsymbol{z}_t = (1, Y_{t-1}, \ldots, Y_{t-p})^\top$ and $N = nm$, and under $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_0^*$,

$$Y_t \ =\ \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \boldsymbol{\beta}_0^* + U_t, \quad \forall \boldsymbol{\gamma} \in \Gamma,$$

from which we see that

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) \;=\; \boldsymbol{\beta}_0^* + \left(\frac{1}{N}\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right)^{-1} \left(\frac{1}{N}\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma})U_t\right).$$

We first show that

$$\sup_{\boldsymbol{\gamma}\in\Gamma} \left\| \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big] \right\| \xrightarrow{\;p\;} 0, \tag{A.1}$$

where $N = N_n = mn \to \infty$ as $n \to \infty$, and $\boldsymbol{\gamma} = (d, \mu)^\top$.

Consider the sequences of numbers r_n and d_n, which denote the "block size" and "gap size", satisfying $r_n \to \infty$, $d_n \to \infty$, $r_n + d_n \le N$, and $r_n + d_n = o(n)$. We then define the blocks $B_j := \{t; t = (j-1)(r_n+d_n)+1, \ldots, (j-1)(r_n+d_n)+r_n\}$, and the gaps $G_j := \{t; t = (j-1)(r_n+d_n)+r_n+1, \ldots, j(r_n+d_n)\}$. Therefore each block B_j has length r_n, and is followed by a gap G_j of length d_n.

Define the total number of blocks $J_n = \lfloor N/(r_n+d_n)\rfloor$, so that the total sample $\{1,\ldots,N\}$ is covered by $\cup_{j=1}^{J_n} B_j \cup R$, where R is the set that contains the "remainders" after the last block B_{J_n}. Then,

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big] \\
&\qquad = \frac{1}{N}\sum_{j=1}^{J_n}\sum_{t\in B_j} \Big\{\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big]\Big\} \\
&\qquad\qquad + \frac{1}{N}\sum_{t\in R} \Big\{\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big]\Big\} \\
&\qquad =: \frac{1}{N}\sum_{j=1}^{J_n}\sum_{t\in B_j} \Delta_t(\boldsymbol{\gamma}) + \frac{1}{N}\sum_{t\in R}\Delta_t(\boldsymbol{\gamma}) \\
&\qquad =: \frac{1}{N}\sum_{j=1}^{J_n} W_j + \frac{1}{N}\sum_{t\in R}\Delta_t(\boldsymbol{\gamma}) \\
&\qquad =: \mathcal{A}_1 + \mathcal{A}_2.
\end{aligned} \tag{A.2}$$

We note that since each pair of consecutive blocks B_j and B_{j+1} is separated by a gap G_j of length d_n, any two points $t \in B_j$ and $t' \in B_{j+1}$ are at least d_n steps apart. Therefore, between events measurable with respect to the σ-algebra for B_j and that for B_{j+1}, the degree of dependence as measured by the mixing coefficient is $\alpha_X(d_n)$,

where

$$\alpha_X(d_n) = \sup_s \sup \left\{ |P(A \cap B) - P(A)P(B)| \,;\; A \in \sigma(X_u;\; u \le s), B \in \sigma(X_u;\; u \ge s + l) \right\}.$$

Since $d_n \to \infty$ so that $\alpha_X(d_n) \to 0$ as $n \to \infty$, the blocks of x, i.e. X_t over the set B_j, is α-mixing and is approximately independent for large n.

Given $W_j = \sum_{t \in B_j} \{ \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top] \} = \sum_{t \in B_j} \Delta_t(\boldsymbol{\gamma})$, we see that $\mathbb{E}\|\Delta_t(\boldsymbol{\gamma})\|^2 \le C_1 < \infty$, since by the assumption $\sup_{t,\boldsymbol{\gamma}} \mathbb{E}\|\boldsymbol{Z}_t(\boldsymbol{\gamma})\|^{2+\nu} < \infty$. Therefore, using the Cauchy-Schwarz inequality, we see that

$$\begin{aligned}
&\sum_{j=1}^{J_n} \mathbb{E}\left[W_j(\boldsymbol{\gamma})^2\right] = \sum_{j=1}^{J_n} \sum_{t,s \in B_j} \mathbb{E}\left[\Delta_t(\boldsymbol{\gamma}) \Delta_s(\boldsymbol{\gamma})\right] \\
&= \sum_{j=1}^{J_n} \mathbb{E}\left(\sum_{t \in B_j} \left\{ \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right] \right\} \right) \left(\sum_{s \in B_j} \left\{ \boldsymbol{Z}_s(\boldsymbol{\gamma}) \boldsymbol{Z}_s(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_s(\boldsymbol{\gamma}) \boldsymbol{Z}_s(\boldsymbol{\gamma})^\top \right] \right\} \right) \\
&= \sum_{j=1}^{J_n} \sum_{t \in B_j} \sum_{s \in B_j} \mathbb{E}\left\{ \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right] \right\} \left\{ \boldsymbol{Z}_s(\boldsymbol{\gamma}) \boldsymbol{Z}_s(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_s(\boldsymbol{\gamma}) \boldsymbol{Z}_s(\boldsymbol{\gamma})^\top \right] \right\} \\
&= \sum_{j=1}^{J_n} \sum_{t \in B_j} \sum_{s \in B_j} \left(\mathbb{E}\left\| \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right] \right\|^2 \mathbb{E}\left\| \boldsymbol{Z}_s(\boldsymbol{\gamma}) \boldsymbol{Z}_s(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_s(\boldsymbol{\gamma}) \boldsymbol{Z}_s(\boldsymbol{\gamma})^\top \right] \right\|^2 \right)^{1/2} \\
&\le J_n \cdot r_n^2 \cdot C_1 \;\le\; C_1 \frac{N}{r_n + d_n} \cdot r_n \;\le\; C_1 \frac{N}{r_n + d_n} \cdot (r_n + d_n)^2 \;=\; C_1 N (r_n + d_n) \\
&= o(N^2)
\end{aligned} \tag{A.3}$$

in view of the fact that each block B_j has length r_n, and that $r_n + d_n = o(n) = o(N)$.

We also consider the "off-diagonal terms". Applying Davydov's inequality (see, e.g., Bosq, 1998, Eq. (1.10)), we have

$$\begin{aligned}
&\sum_{1 \le i < j \le J_n} \mathbb{E}\left[W_i(\boldsymbol{\gamma}) W_j(\boldsymbol{\gamma})\right] \;\le\; \binom{J_n}{2} \sum_{t \in B_i} \sum_{s \in B_j} \left| \mathrm{Cov}(\Delta_t(\boldsymbol{\gamma}), \Delta_s(\boldsymbol{\gamma})) \right| \\
&\le\; C J_n^2 r_n^2 \cdot 2 \frac{2+\nu}{\nu} 2^{\nu/(2+\nu)} \left\{ \alpha_X(d_n) \right\}^{\nu/(2+\nu)} \cdot \left(\mathbb{E}|\Delta_t(\boldsymbol{\gamma})|^{2+\nu} + \mathbb{E}|\Delta_s(\boldsymbol{\gamma})|^{2+\nu} \right)^{2/(2+\nu)} \\
&\le\; C' J_n^2 r_n^2 \left\{ \alpha_X(d_n) \right\}^{\nu/(2+\nu)} = C' \left(\frac{N}{r_n + d_n} \right)^2 r_n^2 \left\{ \alpha_X(d_n) \right\}^{\nu/(2+\nu)} \\
&\le\; C' \frac{N^2 (r_n + d_n)^2}{(r_n + d_n)^2} \left\{ \alpha_X(d_n) \right\}^{\nu/(2+\nu)} \;=\; C' N^2 \left\{ \alpha_X(d_n) \right\}^{\nu/(2+\nu)} \;=\; o(N^2)
\end{aligned} \tag{A.4}$$

since $d_n \to \infty$ as $n \to \infty$.

Consequently, for the leading term $\mathcal{A}_1$ in (A.2), for each fixed $\boldsymbol{\gamma}$ and any $\varepsilon > 0$,

by Chebyshev's inequality we have

$$\begin{aligned} P\left(\left|\frac{1}{N}\sum_{j=1}^{J_n} W_j\right| > \varepsilon\right) &= P\left(\left|\frac{1}{N}\sum_{j=1}^{J_n}\sum_{t\in B_j}\left\{\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right]\right\}\right| > \varepsilon\right) \\ &\leq \frac{\mathbb{E}\left[\sum_{j=1}^{J_n}\sum_{i=1}^{J_n} W_j(\boldsymbol{\gamma})W_i(\boldsymbol{\gamma})\right]}{\varepsilon^2 N^2} \\ &\leq \frac{\sum_{j=1}^{J_n}\mathbb{E}\left[W_j(\boldsymbol{\gamma})^2\right] + 2\sum_{1\leq i<j\leq J_n}\mathbb{E}\left[W_i(\boldsymbol{\gamma})W_j(\boldsymbol{\gamma})\right]]}{\varepsilon^2 N^2} \\ &\to 0 \end{aligned}$$

in view of (A.3) and (A.4). This proves

$$\frac{1}{N}\sum_{j=1}^{J_n} W_j \xrightarrow{p} 0. \tag{A.5}$$

We now move on to the "remainder term" $\mathcal{A}_2$ in (A.2). We recall that the sample is blocked into sets $B_1, \ldots, B_{J_n}$ (each of length r_n) plus gaps $G_1, \ldots, G_{J_n}$ (each of length d_n), and the remainder R is the set of any leftover indices so that

$$\{1, 2, \ldots, N\} = \left(\bigcup_{j=1}^{J_n} B_j\right) \cup \left(\bigcup_{j=1}^{J_n} G_j\right) \cup R.$$

Therefore, by construction, $\mathrm{Card}(R) \leq r_n + d_n \leq cN$ for some constant $c > 0$ because the number of elements in R is at most $O(r_n + d_n)$, where $r_n + d_n = O(N)$. We note that

$$\begin{aligned} &\left\|\frac{1}{N}\sum_{t\in R}\left\{\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right]\right\}\right\| \\ &\qquad \leq \frac{1}{N}\sum_{t\in R}\left\|\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right]\right\| \\ &\qquad \leq \frac{\mathrm{Card}(R)}{N}\cdot\max_{t\in R}\left\|\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right]\right\|. \end{aligned}$$

But we have

$$P\left(\max_{t\in R}\left\|\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\left[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right]\right\| > M\right)$$

$$
\begin{aligned}
&\leq \sum_{t\in R} P\left(\left\|\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big]\right\| > M\right) \\
&\leq \mathrm{Card}(R)\frac{C}{M^{2+\nu}} \;\leq\; cN\frac{C}{M^{2+\nu}} = \frac{C'm\cdot n}{M^{2+\nu}}.
\end{aligned}
$$

Letting $n \to \infty$, for each n we can choose $M \to \infty$ to make $C'/M^{2+\delta}$ arbitrarily small, and therefore $\max_{t\in R}\|\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top]\| = \max_{t\in R}\|\Delta_t(\boldsymbol{\gamma})\|$ is bounded in probability.

Now, since $\mathrm{Card}(R)/N \to 0$, it follows that

$$
\frac{1}{N}\sum_{t\in R}\Delta_t(\boldsymbol{\gamma}) \xrightarrow{\;p\;} 0. \tag{A.6}
$$

Therefore, in view of (A.5) and (A.6) we finally have

$$
\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big] = \frac{1}{N}\sum_{j=1}^{J_n}W_j + \frac{1}{N}\sum_{t\in R}\Delta_t(\boldsymbol{\gamma}) \;\xrightarrow{p}\; 0 \tag{A.7}
$$

for each fixed $\boldsymbol{\gamma}\in\Gamma$.

We now establish the uniform convergence via the standard Vapnik-Chervonenkis (VC)-type covering argument on the compact set $\Gamma \subset \mathbb{R}^d$ (see, *inter alia*, Hong and Linton, 2020). For each $\delta > 0$, we construct a finite δ-net $\{\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_m\} \subset \boldsymbol{\gamma}$ such that for any $\boldsymbol{\gamma}\in\Gamma$, there exists some $\boldsymbol{\gamma}_j$ with $\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_j\| \leq \delta$. Specifically, we partition Γ by $M = M(\boldsymbol{\gamma}_j,\delta)$ number of balls of radius δ, each of which has center $\boldsymbol{\gamma}_j$, for $j = 1,\ldots,M$, i.e.

$$
\boldsymbol{\gamma} \subset \bigcup_{j=1}^{M} B(\boldsymbol{\gamma}_j,\delta).
$$

The covering number $M = M(\delta)$ grows as $\delta \to 0$, but remains finite for each fixed δ.

By Minkowski's inequality, we have

$$
\begin{aligned}
&\sup_{\boldsymbol{\gamma}\in\Gamma}\left\|\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big]\right\| \\
&\qquad\leq \left\|\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)^\top\Big]\right\| \\
&\qquad\quad + \sup_{\boldsymbol{\gamma}\in\Gamma}\left\|\frac{1}{N}\sum_{t=1}^{N}\left(\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)^\top\right)\right\|
\end{aligned}
$$

$$+ \sup_{\boldsymbol{\gamma}\in\Gamma} \left\| \frac{1}{N}\sum_{t=1}^{N} \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)^\top\Big] - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\Big] \right\|$$

$$=: \ \mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3.$$

First, we note that due to (A.7), which holds for each fixed point $\boldsymbol{\gamma}_j$, $\mathcal{B}_1 = o_p(1)$. Moving on to the second term, we first remark that $\boldsymbol{Z}_t$ can be rewritten compactly as follows:

$$\begin{aligned} \boldsymbol{Z}_t(\boldsymbol{\gamma}) &= \begin{cases} (\boldsymbol{Z}_t, 0)^\top & \text{if } X_{t-D} < \mu, \\ (0, \boldsymbol{Z}_t)^\top & \text{if } X_{t-D} \geq \mu, \end{cases} \\ &\equiv \Big(\mathbf{1}(X_{t-D} < \mu)\boldsymbol{Z}_t,\ \mathbf{1}(X_{t-D} \geq \mu)\boldsymbol{Z}_t\Big)^\top. \end{aligned}$$

It is worth noting that d is fixed (and so is $D = md$), and μ is what changes the indicator. In fact, when μ is moved by a small amount, the indicator functions only change if X_{t-D} is in the vicinity of the threshold. In other words, $\mathbf{1}(X_{t-D} < \mu)$ and $\mathbf{1}(X_{t-D} \geq \mu)$ can only differ on a boundary set.

Specifically, a time index t is said to be in the boundary set $\mathcal{U}$ if X_{t-D} lies in either ball $X_{t-d} \in B(\mu_1, \delta) \cup B(\mu_2, \delta) = \{x;\ |x-\mu_1| \leq \delta \text{ or } |x-\mu_2| \leq \delta\}$. It is straightforward to show that outside the boundary set, we have $\mathbf{1}(X_{t-D} < \mu) = \mathbf{1}(X_{t-D} \geq \mu)$. Therefore, when t is not in the boundary set, $\boldsymbol{Z}_t(\boldsymbol{\gamma}_1) = \boldsymbol{Z}_t(\boldsymbol{\gamma}_2)$, where $\boldsymbol{\gamma}_1 = (d, \mu_1)$ and $\boldsymbol{\gamma}_2 = (d, \mu_2)$ with $|\mu_1 - \mu_2| \leq \delta$. Therefore,

$$\begin{aligned} \mathcal{B}_2 &= \sup_{\boldsymbol{\gamma}\in\Gamma} \left\| \frac{1}{N}\sum_{t=1}^{N} \left(\boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\boldsymbol{Z}_t(\boldsymbol{\gamma}_j)^\top\right) \right\| \\ &\leq \sup_{\boldsymbol{\gamma}\in\Gamma} \frac{1}{N}\sum_{t=1}^{N} \big\|\boldsymbol{Z}_t(\boldsymbol{\gamma}) - \boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\big\| \cdot \big\|\boldsymbol{Z}_t(\boldsymbol{\gamma}) + \boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\big\| \\ &\leq 0 \ + \ \frac{C\delta N \cdot \max_{t\in\mathcal{U}} \left(\sqrt{2}\|\boldsymbol{Z}_t\|\right)}{N} \cdot \sup_{\boldsymbol{\gamma}\in\Gamma} \frac{1}{N}\sum_{t=1}^{N} \big\|\boldsymbol{Z}_t(\boldsymbol{\gamma}) + \boldsymbol{Z}_t(\boldsymbol{\gamma}_j)\big\| \\ &\leq C\delta \max_t(\sqrt{2}\|\boldsymbol{Z}_t\|) \cdot \frac{1}{N}\sum_{t=1}^{N} \sqrt{(2\|\boldsymbol{Z}_t\|^2 + 2\|\boldsymbol{Z}_t\|)^2} \ \leq \ C'\delta, \end{aligned}$$

where the last inequalities are due to $\|\boldsymbol{Z}_t\|$ being bounded in probability as $\boldsymbol{Z}_t$ has finite moments (Assumption 2′-(i)). We also used the fact that X_{t-D} has a continuous

density $f_X(\cdot)$ bounded by some constant $\bar{f}$, and therefore, for some fixed $\mu_0 \in \mathbb{R}$,

$$P\left(\|X_{t-D} - \mu_0\| \leq \delta\right) \;=\; F_X(\mu_0 + \delta) - F_X(\mu - \delta) \int_{\mu_0 - \delta}^{\mu + \delta} f_X(x)dx \;\leq\; \bar{f} \cdot 2\delta,$$

and therefore $\text{Card}\left\{t;\ |X_{t-D} - \mu| \leq \delta\right\} \leq 2\delta \bar{f} N \;=\; O(\delta N)$.

For any $\epsilon > 0$ we can choose δ small that $C'\delta < \epsilon/2$, and also choose N large enough that the sums converge within $\epsilon/2$, yielding the desired uniform convergence in probability of $\mathcal{B}_2$ to 0 (and similarly for $\mathcal{B}_3$). Therefore, we have constructed (A.1).

Now by continuous mapping, we have

$$\sup_{\boldsymbol{\gamma} \in \Gamma} \left\| \left(\frac{1}{N} \sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right)^{-1} - \left(\mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \Big] \right)^{-1} \right\| \xrightarrow{p} \; 0 \tag{A.8}$$

as $N = nm \to \infty$. We remark that Assumption 2′-(ii) guarantees the existence of the inverse in (A.8) uniformly in $\boldsymbol{\gamma}$, as this requires $\boldsymbol{M}(\boldsymbol{\gamma}) = \mathbb{E}[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top]$ to be nonsingular for all $\boldsymbol{\gamma}$.

On noting that $\mathbb{E}(\boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t) = 0$, by following the same argument that led to (A.1) via the blocking method (see, for example, the proof for Theorem 3-(ii) below), we immediately have

$$\sup_{\boldsymbol{\gamma} \in \Gamma} \left\| \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t - 0 \right\| \xrightarrow{p} \; 0$$

from which it follows by the definition and Minkowski's inequality that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^*$ converges to 0 in probability, uniformly over $\boldsymbol{\gamma} \in \Gamma$. This completes the proof of Theorem 3-(i).

We now move on to the asymptotic normality, where we have

$$\sqrt{N} \left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^* \right] = \left(\frac{1}{N} \sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t \right).$$

As we employed above in the Bernstein's blocking procedure, we choose sequences r_n and d_n, both of which tend to ∞ as $n \to \infty$, and satisfy $r_n + d_n \leq N$, and $r_n + d_n = o(n)$. The blocks and gaps B_j and G_j, as well as the number of blocks J_n are defined as before, so that

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t \;=\; \frac{1}{\sqrt{N}} \sum_{j=1}^{J_n} \sum_{t \in B_j} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t + \frac{1}{\sqrt{N}} \sum_{t \in R} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t$$

$$= \frac{1}{\sqrt{N}} \sum_{j=1}^{J_n} \xi_j + \frac{1}{\sqrt{N}} \sum_{t \in R} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t. \tag{A.9}$$

For notational simplicity we prove the rest in the univariate setting; multivariate extension is straightforward via the Cramér-Wold device. As for the "remainder" in (A.9), we can apply Theorem 1 of Doukhan and Louhichi (1999) and yield

$$\frac{1}{N} \mathbb{E} \left(\sum_{t \in R} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t \right)^2 \to 0,$$

which suggests that the second term is $o_p(1)$.

For the first term in (A.9), by applying Hölder's inequality on the moment condition in Assumption 2′, it follows that (I): $\mathbb{E}|\boldsymbol{Z}_t(\boldsymbol{\gamma})U_t|^\ell < \infty$ where $\ell = 2v > 2$. Also, the mixing rate is $\alpha(j) = O(j^{-A})$ for $A > \frac{2\nu}{2\nu-1} > 0$, and therefore (II): $\sum_{j \geq 1} \alpha(j)^{1-1/2v} < \infty$. With (I) and (II), we can apply Theorem 1.5 of Bosq (1998), which yields

$$r_n^{-1} \mathbb{E} \xi_1 \xi_1^\top \to S(\boldsymbol{\gamma}) = \mathbb{E}\big(\boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t\big)^2$$

Consequently, both the *Feller condition*:

$$\frac{1}{N} \sum_{j=1}^{J_n} \mathbb{E} \xi_j^2 = \frac{J_n r_n}{N} \cdot \frac{1}{r_n} \mathbb{E} \xi_1^2 \to S(\cdot),$$

and the *Lindeberg condition*: for any $\epsilon > 0$,

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^{J_n} \mathbb{E} \left(\xi_j^2 1_{|\xi_j| \geq \epsilon S \sqrt{N}} \right) &\leq \frac{1}{N} \sum_{j=1}^{J_n} \left\{ \mathbb{E} \xi_1^4 \right\}^{1/2} P\left(|\xi_1| \geq \epsilon S \sqrt{N} \right) \\ &\leq C J_n r_n \cdot \frac{\mathbb{E} \xi_1^2}{N^2 \epsilon^2 S^2} = O\left(\frac{r_n}{N} \right) \to 0 \end{aligned}$$

are satisfied, and the required conditions for the central limit theorem (e.g., Peligrad, 1986), are met.

Therefore, as $n \to \infty$ we have

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t \xrightarrow{\mathcal{D}} N\left(0, \mathbb{E}\big(\boldsymbol{Z}_t(\boldsymbol{\gamma}) U_t U_t \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\big)\right).$$

Extending this to the distributional convergence for a finite collection of $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_K$ is trivial, since the central limit theorem we employ holds in multivariate sense. Con-

sequently, by continuous mapping and (A.8), the limiting distribution as given in (25) follows. The proof of Theorem 3 is now complete. □

A.3 Proof of Theorem 1 (Regular Midastar)

Proof. In the regular Midastar, recall that the conditional least squares estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ is given by

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left\{ \sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right\}^{-1} \left\{ \sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma}) y_t \right\},$$

where $\boldsymbol{z}_t = (1, y_t, \ldots, y_{t+1-p})^\top$ and

$$\boldsymbol{Z}_t(\boldsymbol{\gamma}) = \begin{bmatrix} \boldsymbol{z}_{t-1} \mathbf{1}(x^*_{t-d/m} < \mu) \\ \boldsymbol{z}_{t-1} \mathbf{1}(x^*_{t-d/m} \geq \mu) \end{bmatrix}, \qquad t \in \mathbb{L} = \{1, \ldots, n\}.$$

Under $H_0^* : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_0^*$, we have $y_t = \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \boldsymbol{\beta}_0^* + u_t, \forall \boldsymbol{\gamma} \in \Gamma$, from which we see that

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) \;=\; \boldsymbol{\beta}_0^* + \left(\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma}) u_t \right).$$

As before in Theorem 3, we write

$$\Delta_t(\boldsymbol{\gamma}) \;:=\; \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \Big].$$

Since $\{(y_t, x_t^*, u_t)\}$ is jointly stationary and α-mixing, by standard results, *inter alia* Ibragimov and Linnik (1971), its finite-dimensional measurable transformation $\Delta_t(\boldsymbol{\gamma})$ is ergodic stationary.

Therefore, by Birkhoff's Ergodic Theorem we have

$$\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \Big] \xrightarrow{p} 0 \tag{A.10}$$

for each $\boldsymbol{\gamma} = (d, \mu)^\top \in \Gamma$. The uniform convergence can be constructed via the same VC-type covering method as in the proof of Theorem 3 above, and is not repeated for brevity. Consequently, we have

$$\sup_{\boldsymbol{\gamma} \in \Gamma} \left\| \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top - \mathbb{E}\Big[\boldsymbol{Z}_t(\boldsymbol{\gamma}) \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top \Big] \right\| \xrightarrow{p} 0,$$

as $n \to \infty$, and as before, it follows that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^*$ converges to 0 in probability, uniformly over $\boldsymbol{\gamma} \in \Gamma$. This completes the proof of the first part of Theorem 1.
As for the asymptotic normality, we note that

$$\sqrt{n}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0^*\right] = \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma})u_t\right).$$

Since stationarity and α-mixing properties are preserved under any *finite-dimensional* measurable transformations, it is straightforward to check Assumption 1 meets the required conditions for Theorem 1.7 of Bosq (1998). Consequently, we have as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{Z}_t(\boldsymbol{\gamma})u_t \xrightarrow{\mathcal{D}} N\left(0, \mathbb{E}\big(\boldsymbol{Z}_t(\boldsymbol{\gamma})u_t u_t \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\big)\right). \tag{A.11}$$

The desired asymptotic normality follows as before in Theorem 3. Corollary 1 now immediately follows – Recall that the local alternative hypothesis (24) is given by $H_1^*(\boldsymbol{\lambda}_n) : \boldsymbol{\beta}_n = \bar{\boldsymbol{\beta}}_1 + n^{-\frac{1}{2}}\bar{\boldsymbol{\lambda}}_n$. Replacing $\boldsymbol{\beta}$ with $\boldsymbol{\beta}_n$ we can write

$$\begin{aligned}
\sqrt{n}\left\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_n\right\} &= \sqrt{n}\left[\frac{1}{n}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\sum_{t\in\mathbb{L}} \boldsymbol{Z}_t(\boldsymbol{\gamma})\left\{\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\boldsymbol{\beta}_n + u_t\right\} - \boldsymbol{\beta}_n\right] \\
&= \sqrt{n}\left\{\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\boldsymbol{M}_n(\boldsymbol{\gamma})\boldsymbol{\beta}_n + \frac{1}{n}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\sum_{t\in\mathbb{L}} \boldsymbol{s}_t(\boldsymbol{\gamma}) - \boldsymbol{\beta}_n\right\} \\
&= \frac{1}{\sqrt{n}}\boldsymbol{M}_n(\boldsymbol{\gamma})^{-1}\sum_{t\in\mathbb{L}} \boldsymbol{s}_t(\boldsymbol{\gamma}) \xrightarrow{\mathcal{D}} \mathcal{G}(\boldsymbol{\gamma})
\end{aligned}$$

in view of (A.10) and (A.11). Since $\lim_{n\to\infty}\|\boldsymbol{\lambda}_n\| < \infty$, we have $\sqrt{n}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \bar{\boldsymbol{\beta}}_1\} \xrightarrow{\mathcal{D}} \bar{\boldsymbol{\lambda}} + \mathcal{G}(\boldsymbol{\gamma})$, as required. The same arguments apply for the proof of consistency. □

A.4 Proof of Theorem 2 (Bootstrap consistency)

Proof. 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 $\boldsymbol{R} = (\boldsymbol{I}_{p+1}, -\boldsymbol{I}_{p+1})$ is the selection matrix for H_0^*; $\boldsymbol{Q}(\boldsymbol{\gamma}) = \boldsymbol{M}(\boldsymbol{\gamma})^{-1}\boldsymbol{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0)$; $\boldsymbol{M}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \mathbb{E}\{\boldsymbol{Z}_t(\boldsymbol{\gamma}_1)\boldsymbol{Z}_t(\boldsymbol{\gamma}_2)^\top\}$; $\boldsymbol{\gamma}_0$ is the fixed true value of $\boldsymbol{\gamma}$ when $\boldsymbol{\lambda}_n \neq \mathbf{0}$.

Incorporate all possible values of $\boldsymbol{\gamma}$ in $\mathcal{W}^\lambda(\boldsymbol{\gamma})$ with one of the three transformations:

$$\sup\mathcal{W}^\lambda = \sup_{\boldsymbol{\gamma}\in\Gamma} \mathcal{W}^\lambda(\boldsymbol{\gamma}), \qquad \text{ave}\mathcal{W}^\lambda = \int_\Gamma \mathcal{W}^\lambda(\boldsymbol{\gamma}) d\mu^*(\boldsymbol{\gamma}), \tag{A.12}$$

$$\exp\mathcal{W}^\lambda = \ln\left[\int_\Gamma \exp\left\{\frac{\mathcal{W}^\lambda(\boldsymbol{\gamma})}{2}\right\} d\mu^*(\boldsymbol{\gamma})\right],$$

where some subset of Γ has positive measure with respect to μ^* (e.g., Davies, 1977, 1987, Andrews and Ploberger, 1994). The quantities in (A.12) are the asymptotic counterparts to those in (20).

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 $\{\xi_t\}_{t\in\mathbb{L}}$ be iid standard normal random variables. Define:

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

$$\widehat{\boldsymbol{v}}_n(\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}}\sum_{t\in\mathbb{L}} \widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma})\xi_t.$$

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

To prove Theorem 2.(i), impose H_0^*. Then, the asymptotic distribution of $\widehat{p}_n$ is Uniform on $[0,1]$. Application of the Glivenko-Cantelli Theorem implies that $\widehat{p}_n^B(H_0^*) - \widehat{p}_n \xrightarrow{p} 0$ 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 $\widehat{p}_n \Rightarrow_p p^\lambda$. 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 2 (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). □

A.5 Proof of Theorem 4 (Bootstrap consistency)

The proof for reverse Midastar broadly follows that of Theorem 2. As before, we focus on the Wald tests in this proof, as the LM tests can be done similarly. We 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 $\boldsymbol{R} = (\boldsymbol{I}_{p+1}, -\boldsymbol{I}_{p+1})$ is the selection matrix for H_0^*; $\boldsymbol{Q}(\boldsymbol{\gamma}) = \boldsymbol{M}(\boldsymbol{\gamma})^{-1}\boldsymbol{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0)$; $\boldsymbol{M}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \mathbb{E}\{\boldsymbol{Z}_t(\boldsymbol{\gamma}_1)\boldsymbol{Z}_t(\boldsymbol{\gamma}_2)^\top\}$; $\boldsymbol{\gamma}_0$ is the fixed true value of $\boldsymbol{\gamma}$ when $\boldsymbol{\lambda}_N \neq \boldsymbol{0}$; $\bar{\boldsymbol{\beta}}_1 := (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_1^\top)^\top$; and $\bar{\boldsymbol{\lambda}}_N := (\boldsymbol{0}_{p+1}^\top, \boldsymbol{\lambda}_N^\top)^\top$, so that under $H_1^*(\boldsymbol{\lambda}_N)$ we can write $\boldsymbol{\beta}_N = \bar{\boldsymbol{\beta}}_1 + N^{-1/2}\bar{\boldsymbol{\lambda}}_N$. We incorporate all possible values of $\boldsymbol{\gamma}$ in $\mathcal{W}^\lambda(\boldsymbol{\gamma})$ with one of the three transformations: $\sup\mathcal{W}^\lambda, \text{ave}\mathcal{W}^\lambda$, and $\exp\mathcal{W}^\lambda$ in (A.12), which are the asymptotic counterparts to those in (20) with n replaced by N.

As in the Proof of Theorem 2, 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, and note that $g(\cdot)$ is a continuous functional of the Gaussian process $\mathcal{G}(\boldsymbol{\gamma})$. Application of Theorem 3 together with the continuous mapping theorem 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

$$\mathcal{W}^0(\boldsymbol{\gamma}) = \mathcal{G}(\boldsymbol{\gamma})^\top\boldsymbol{R}^\top\{\boldsymbol{R}\boldsymbol{V}(\boldsymbol{\gamma})\boldsymbol{R}^\top\}^{-1}\boldsymbol{R}\mathcal{G}(\boldsymbol{\gamma}),$$

and 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 $p_N = 1 - F^0(g_N)$, where $g_N = g(\mathcal{W}_N)$. Then, $p_N \Rightarrow p^\lambda = 1 - F^0(g^\lambda)$ by monotonicity and continuity of F^0. Therefore, 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. We define:

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

$$\widehat{\boldsymbol{v}}_N(\boldsymbol{\gamma}) = \frac{1}{\sqrt{N}}\sum_{t=1}^N \widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma})\xi_t, \qquad \widehat{\boldsymbol{s}}_t(\boldsymbol{\gamma}) = \boldsymbol{Z}_t(\boldsymbol{\gamma})\,\widehat{u}_t(\boldsymbol{\gamma}),$$

where $\boldsymbol{M}_N(\boldsymbol{\gamma}) = N^{-1}\sum_{t=1}^N \boldsymbol{Z}_t(\boldsymbol{\gamma})\boldsymbol{Z}_t(\boldsymbol{\gamma})^\top$, $\widehat{\boldsymbol{V}}_N(\boldsymbol{\gamma})$ is the reverse-sample analogue of $\widehat{\boldsymbol{V}}_n(\boldsymbol{\gamma})$, and $\widehat{u}_t(\boldsymbol{\gamma}) = Y_t - \boldsymbol{Z}_t(\boldsymbol{\gamma})^\top\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ is the reverse-sample residual constructed from the conditional LS estimator. Note that $\widehat{\boldsymbol{v}}_N(\boldsymbol{\gamma})$ is a mean zero Gaussian process

with a well-defined covariance kernel conditional on the sample. Let $\widehat{F}_N$ denote the distribution function of $\widehat{g}_N = g(\widehat{\mathcal{W}}_N)$ conditional on the sample. Let $\widehat{p}_N = 1 - \widehat{F}_N(g_N)$. Then, under the strengthened absolute-regularity condition, the bootstrap argument of Hansen (1996, Theorem 2) implies that $\widehat{p}_N \Rightarrow_p p^\lambda$.

To prove Theorem 4(i), impose H_0^*. Then, the asymptotic distribution of $\widehat{p}_N$ is Uniform on $[0, 1]$. With the Glivenko-Cantelli Theorem we have $\widehat{p}_N^B(H_0^*) - \widehat{p}_N \xrightarrow{p} 0$ as $N \to \infty$ and $B \to \infty$, and this establishes Theorem 4.(i).

To prove Theorem 4(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 $\widehat{p}_N \Rightarrow_p p^\lambda$. Therefore, it is sufficient to show that $g^\lambda = g(\mathcal{W}^\lambda)$ diverges almost surely as $\|\boldsymbol{\lambda}_N\| \to \infty$. By Assumption $2'$ (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,

$$\left\{\boldsymbol{RQ}(\boldsymbol{\gamma})\bar{\boldsymbol{\lambda}}_N\right\}^\top \{\boldsymbol{RV}(\boldsymbol{\gamma})\boldsymbol{R}^\top\}^{-1}\boldsymbol{RQ}(\boldsymbol{\gamma})\bar{\boldsymbol{\lambda}}_N \to \infty$$

almost surely as $\|\boldsymbol{\lambda}_N\| \to \infty$ for every $\boldsymbol{\gamma} \in \Gamma$. Hence, $g^\lambda \to \infty$ almost surely as $\|\boldsymbol{\lambda}_N\| \to \infty$, which implies $p^\lambda \to 0$ and therefore $\widehat{p}_N \xrightarrow{p} 0$. Finally, $\widehat{p}_N^B(H_0^*) - \widehat{p}_N \xrightarrow{p} 0$ as $N \to \infty$ and $B \to \infty$ implies $\widehat{p}_N^B(H_0^*) \xrightarrow{p} 0$. This establishes Theorem 4(ii) and completes the proof. □