# Uniform Inference in High-Dimensional Threshold Regression Models*†

Jiatong Li‡ Hongqiang Yan §

September 19, 2024

**Abstract**

We develop uniform inference for high-dimensional slope parameters in threshold regression models, allowing for either cross-sectional or time series data. We first establish oracle inequalities for prediction errors, and $\ell_1$ estimation errors for the Lasso estimator of the slope parameters and the threshold parameter, accommodating heteroskedastic non-subgaussian error terms and non-subgaussian covariates. Next, we derive the asymptotic distribution of tests involving an increasing number of slope parameters by debiasing (or desparsifying) the Lasso estimator in cases with no threshold effect and with a fixed threshold effect. We show that the asymptotic distributions in both cases are the same, allowing us to perform uniform inference without specifying whether the specification is a linear or threshold regression. Finally, we demonstrate the consistent performance of our estimator in both cases through simulation, and we apply the proposed estimator to analyze two empirical applications.

**Keywords:** High-dimensional data, Lasso, Tests, Threshold models, Uniform inference.

**JEL Codes:** C12, C13, C24.

* We received very useful comments from Professors Mehmet Caner, Harold Chiang, Barry Goodwin, Vincent Han, Atsushi Inoue, Ilze Kalnina, Tong Li, Zheng Li, Yuya Sasaki (Jiatong Li's advisor), Yulong Wang and Ping Yu (listed in alphabetical order), as well as participants in Bristol Econometrics Study Group.

† Authors are listed in alphabetical order.

‡Jiatong Li: jiatong.li@vanderbilt.edu. Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA

§Hongqiang Yan: hongqiang.yan@asu.edu. Morrison School of Agribusiness, Arizona State University, 7231 E Sonoran Arroyo Mall, Santan Hall Suite 235, Mesa, AZ 85212-6414,USA

# 1 Introduction

Consider the following threshold regression model

$$Y_i = X_i'\beta_0 + X_i'\delta_0 \mathbf{1}\{Q_i < \tau_0\} + U_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $X_i$ is a $p \times 1$ covariate vector, and $Q_i$ is the threshold variable determining regime switching; for example, well-developed countries may follow a different economic growth pattern from developing countries. $\tau_0$ is the unknown threshold parameter, and $U_i$ is the error term. In this paper, we focus on uniform inference for high-dimensional regression parameters $(\beta_0, \delta_0)$, allowing for $p > n$. The threshold autoregression (TAR) model, with the lag of the series as the threshold variable, was formally introduced by Tong and Lim (1980) to analyze cyclical time series data. It is a class of non-linear time series models and is parsimonious for nonparametric model estimation. [1] [2] Potter (1995) applies it to study the properties of US GNP and finds that the response of output to shocks is asymmetric throughout different stages of the business cycle.

Subsequently, threshold regression is utilized by Hansen (2000) to identify multiple regimes based on a particular predetermined variable, allowing for either time series or cross-sectional data. Since then, there has been growing interest in reanalyzing previous empirical applications using threshold models, particularly when multiple equilibria may exist. For example, Hansen (2000) and Lee et al. (2016) consider cross-country economic growth behaviors initially analyzed by Durlauf and Johnson (1995); Yu and Fan (2021) and Lee and Wang (2023) examine tipping-like behavior discussed in Card et al. (2008); and Grennes et al. (2010), Afonso and Jalles (2013), and Chudik et al. (2017) investigate the effect of government debt on economic growth originally studied by Reinhart and Rogoff (2010). In this paper, we reconfirm the existence of multiple steady states, as described by Durlauf and Johnson (1995), by showing that some values of $\delta_2$ are significantly different from 0. Subsequently, we apply the high-dimensional local projection threshold model to identify the threshold point that defines the state of the economy and reestimate the impulse response to a military

[1] For a survey paper, see Hansen (2011).

[2] Chan (1993) and Chan and Tsay (1998) study the limiting properties of the least square estimators in the threshold autoregression model.

spending news shock in government spending and GDP, as previously examined by Ramey and Zubairy (2018).

Additionally, variable selection is necessary to identify threshold effects. A linear model incorporating a broader set of regressors could outperform the threshold model with a specific set of covariates, as emphasized by Lee et al. (2016). Meanwhile, recent advancements in data collection and processing have enabled us to analyze high-dimensional datasets, where the number of variables may significantly exceed the sample size. High dimensionality can also arise from including interaction terms and polynomial terms as regressors for more flexible functional forms. To keep the model free from arbitrary variable selection, we study such high-dimensional settings. However, traditional estimation and inferential methods, such as OLS and MLE, are no longer valid even in high-dimensional linear regression models. Specifically, in the high-dimensional threshold model, due to sample splitting, the total number of parameters may be larger than the effective sample size in the regime with the fewest observations, particularly when multiple threshold points exist, thereby leading to poor estimation and out-of-sample prediction in finite samples. Many methods are available for high dimensional estimation and variable selection, for example, Lasso in Tibshirani (1996) and SCAD in Fan and Li (2001). We apply Lasso to estimate the high-dimensional threshold regression (allowing for $p > n$), as in Lee et al. (2016) and Callot et al. (2017).

In this paper, we first study cross-sectional data and use the concentration inequality from Chernozhukov et al. (2014) and Chernozhukov et al. (2015), as formulated in Lemma 2 of Chiang et al. (2023), and concentration inequality for the partial sum of random variables that we propose in Lemma 3 to derive oracle inequalities for both the prediction error and $\ell_1$ estimation error for slope parameters and the threshold parameter, which are qualitatively similar to those in Lee et al. (2016). We allow for heteroskedastic non-subgaussian error terms and non-subgaussian covariates.

Based on these oracle inequalities, we then obtain the desparsified Lasso estimator by using nodewise regression to construct the approximate empirical precision matrix. In our proof, we maintain the dependence assumption between covariates and the threshold variable, and we apply the inverse of a $2 \times 2$ block matrix to construct the precision matrix when the threshold effect may exist.[3] Next, based on the infer-

[3]The independence assumption would significantly simplify the proof, but it is uncommon in

ence theory of Caner and Kock (2018), we establish the asymptotic distribution of tests involving an increasing number of slope parameters, while depending on a fixed sequence of hypotheses in the cases with no threshold effect and with a fixed threshold effect. We show that the asymptotic distributions in both cases are identical. We also provide a uniformly consistent covariance matrix estimator in both cases. [4] We further construct asymptotically valid confidence intervals for the interest of the slope parameter, which are uniformly valid and contract at the optimal rate. Moreover, we generalize the uniform inference theory for the debiased Lasso estimator to the setting of the high-dimensional time series threshold regression model, with local projection threshold regression as a special case.

The existing literature on high-dimensional threshold regression has focused on deriving oracle inequalities for the prediction errors and estimation errors for the Lasso estimator of the slope parameters and the threshold parameter in the case of fixed design with gaussian errors (Lee et al. (2016)), and on model selection consistency in the case of random design with sub-gaussian covariates and errors (Callot et al. (2017)). Both studies focus on independent data. However, high-dimensional inference is another important topic in statistics and econometrics; for example, the estimation of impulse response functions is an essential part of econometric inference in time series models. Thus, in this paper, we perform uniform inference for high-dimensional threshold regression parameters, allowing for either cross-sectional or time series data, by applying the de-sparsified method of van de Geer et al. (2014) to complement the existing literature. This method desparsifies the estimator by constructing a reasonable approximate inverse of the singular empirical Gram matrix, thereby removing the bias from the estimation of the shrinkage method. Our asymptotic result is uniformly valid over the class of sparse models.

A growing body of literature applies the desparsified method of van de Geer et al. (2014) to perform inference in high-dimensional regression models. Gold et al. (2020) desparsified the Lasso estimator based on a two-stage least squares estimation, allowing both the numbers of instruments and of regressors to exceed the sample size. Semenova et al. (2023) desparsified the orthogonal Lasso estimator in their third

empirical applications.

[4] There is a slight difference between the limits of their asymptotic variances since, in the case of a fixed effect, there is a true value for the threshold parameter.

stage when heterogeneous treatment effects are present. Additionally, Adamek et al. (2023) and Adamek et al. (2024) constructed the desparsified Lasso estimator in high-dimensional time series models. The desparsified method has also been applied in high-dimensional panel data models to perform uniform inference, as shown in works by Kock (2016), Kock and Tang (2019), and Chiang et al. (2023). However, all of these studies test hypotheses for a bounded number of parameters. Caner and Kock (2018) later considered hypotheses involving an increasing number of parameters in linear regression models. We are the first to derive the asymptotic distribution of tests involving an increasing number of slope parameters in threshold regression models. Meanwhile, we demonstrate that the asymptotic distributions of the tests are identical in cases with no threshold effect and with a fixed threshold effect, implying that the researchers can perform inference without specifying the existence of a threshold effect.

**Organization:** The rest of the paper is organized as follows. Section 2 recalls the Lasso estimator of Lee et al. (2016) and establishes oracle inequalities under weaker conditions on covariates and error terms. We construct the debiased Lasso estimator and derive the asymptotic distribution of hypothesis tests in Section 3. Section 4 develops the uniform inference theory for high-dimensional time series threshold regression models. In Section 5, we investigate the finite sample properties of our debiased Lasso estimator, followed by two empirical applications. Section 6 concludes. All proofs are deferred to the Appendix.

## Notation

Denote the $\ell_q$ norm of a vector $a$ by $|a|_q$ and the empirical norm of $a \in \mathbb{R}^n$ by $||a||_n := (n^{-1}\sum_{i=1}^n a_i^2)^{1/2}$. For any $m \times n$ matrix $A$, the induced $l_1$-norm and $l_\infty$-norm of $A$ are defined as $\|A\|_{l_1} := \max_{1\le j\le n}\sum_{i=1}^m |A_{ij}|$ and $\|A\|_{l_\infty} := \max_{1\le i\le m}\sum_{j=1}^n |A_{ij}|$, respectively. Additionally, define $\|A\|_\infty := \max_{1\le i\le m, 1\le j\le n} |A_{ij}|$.

For $a \in \mathbb{R}^n$, denote the cardinality of $J(a)$ by $|J(a)|$, where $J(a) = \{j = 1, ..., n : a_j \neq 0\}$. Denote the number of non-zero elements of $a$ by $\mathcal{M}(a)$, characterizing the sparsity of $a$. Let $a_M$ denote the vector in $\mathbb{R}^n$ that has the same coordinates as $a$ on $M$ and zero coordinates on $M^c$. Let the superscript $^{(j)}$ denote the $j$th element of a vector or the $j$th column of a matrix.

Finally, define $f_{(\alpha,\tau)}(x,q) := x'\beta + x'\delta\mathbf{1}\{q<\tau\}$, $f_0(x,q) := x'\beta_0 + x'\delta_0\mathbf{1}\{q<\tau_0\}$, and $\widehat{f}(x,q) := x'\widehat{\beta} + x'\widehat{\delta}\mathbf{1}\{q<\widehat{\tau}\}$. NThe prediction norm is defined as $\left\|\widehat{f}-f_0\right\|_n := \sqrt{\frac{1}{n}\sum_{i=1}^n\left(\widehat{f}(X_i,Q_i)-f_0(X_i,Q_i)\right)^2}$.
In this paper, we use the terms "debiased Lasso estimator" and "desparsified Lasso estimator" interchangeably.

# 2 The Lasso Estimator and Oracle Inequalities

## 2.1 Lasso Estimation

The high-dimensional threshold regression model (1.1) can be rewritten as

$$Y_i = \begin{cases} X_i'\beta_0 + U_i, & \text{if } Q_i \geq \tau_0, \\ X_i'(\beta_0+\delta_0) + U_i, & \text{if } Q_i < \tau_0. \end{cases} \tag{2.1}$$

$Q_i$ is the threshold variable that splits the sample into two regimes and $\delta_0$ represents the threshold effect between two regimes. The model (1.1) thus captures a regime switch based on the observable variable $Q_i$. The parameter $\tau_0$ is the unknown threshold parameter, which lies within a compact parameter space $T = [t_0, t_1]$. There is no threshold effect when $\delta_0 = 0$, and the model reduces to a linear model.

Denoting a $(2p\times 1)$ vector by $\boldsymbol{X}_i(\tau) = (X_i', X_i'\mathbf{1}\{Q_i<\tau\})'$ and an $(n\times 2p)$ matrix by $\boldsymbol{X}(\tau)$, where the $i$-th row is $\boldsymbol{X}_i(\tau)'$. Let $X$ and $X(\tau)$ denote the first and last $p$ columns of $\boldsymbol{X}(\tau)$, respectively. Thus, we can rewrite (1.1) as

$$Y_i = \boldsymbol{X}_i(\tau_0)'\alpha_0 + U_i, \quad i = 1,\ldots,n. \tag{2.2}$$

where $\alpha_0 = (\beta_0', \delta_0')'$. In this paper, our interest lies in performing uniform inference for the high-dimensional slope parameter $\alpha_0$, allowing for $p > n$.

Let $\boldsymbol{Y} := (Y_1,\ldots,Y_n)'$. The residual sum of squares is

$$S_n(\alpha,\tau) = \frac{1}{n}\sum_{i=1}^n\left(Y_i - X_i'\beta - X_i'\delta\mathbf{1}\{Q_i<\tau\}\right)^2 = \|\boldsymbol{Y}-\boldsymbol{X}(\tau)\alpha\|_n^2. \tag{2.3}$$

The Lasso estimator for threshold regression can thus be defined as the one-step minimizer:

$$(\widehat{\alpha}, \widehat{\tau}) := \text{argmin}_{\alpha \in \mathcal{B} \subset \mathbb{R}^{2p}, \tau \in \mathbb{T} \subset \mathbb{R}} \left\{ S_n(\alpha, \tau) + \lambda \left| \boldsymbol{D}(\tau)\alpha \right|_1 \right\}, \tag{2.4}$$

where $\mathcal{B}$ is the parameter space for $\alpha_0$, and $\lambda$ is a tuning parameter. The $(2p \times 2p)$ diagonal weighting matrix is denoted as follows:

$$\boldsymbol{D}(\tau) := \text{diag}\left\{ \left\| \boldsymbol{X}^{(j)}(\tau) \right\|_n, \quad j = 1, ..., 2p \right\}. \tag{2.5}$$

Furthermore, we can rewrite the penalty term as

$$\begin{aligned} \lambda \left| \boldsymbol{D}(\tau)\alpha \right|_1 &= \lambda \sum_{j=1}^{2p} \left\| \boldsymbol{X}^{(j)}(\tau) \right\|_n \left| \alpha^{(j)} \right| \\ &= \lambda \sum_{j=1}^{p} \left[ \left\| X^{(j)} \right\|_n \left| \alpha^{(j)} \right| + \left\| X^{(j)}(\tau) \right\|_n \left| \alpha^{(p+j)} \right| \right]. \end{aligned}$$

Meanwhile, the one-step minimizer $(\widehat{\alpha}, \widehat{\tau})$ in (2.4) can be regarded as a two-step minimizer:
(i) For each $\tau \in \mathbb{T}$, $\widehat{\alpha}(\tau)$ is defined as

$$\widehat{\alpha}(\tau) := \text{argmin}_{\alpha \in \mathcal{B} \subset \mathbb{R}^{2p}} \left\{ S_n(\alpha, \tau) + \lambda \left| \boldsymbol{D}(\tau)\alpha \right|_1 \right\}; \tag{2.6}$$

(ii) Define $\widehat{\tau}$ as the estimator of $\tau_0$ such that:

$$\widehat{\tau} := \text{argmin}_{\tau \in \mathbb{T} \subset \mathbb{R}} \left\{ S_n(\widehat{\alpha}(\tau), \tau) + \lambda \left| \boldsymbol{D}(\tau)\widehat{\alpha}(\tau) \right|_1 \right\}. \tag{2.7}$$

Note that these estimators are weighted Lasso estimators that use a data-dependent $\ell_1$ penalty to balance covariates. Chiang et al. (2023) summarize various ways to impose weights depending on different situations in Remark B.1. Additionally, $\widehat{\tau}$ is an interval; we choose the maximum of the interval as the estimator $\widehat{\tau}$. For any $n$, we use grid search to find candidates for $\widehat{\tau}$ over $\{Q_1, \ldots, Q_n\}$. [5]

[5] If $n$ is very large, we can use only $N < n$ evaluations; see p.4 in Hansen (2000).

## 2.2 Oracle Inequalities

After recalling the Lasso estimator, we proceed to establish the oracle inequalities for the estimators in (2.4). First, we make the following assumptions, some of which are modified from Lee et al. (2016).

**Assumption 1.** *Let* $\{X_i, Q_i, U_i\}_{i=1}^n$ *denote a sequence of independently distributed random variables.*
*(i) For the parameter space* $\mathcal{B}$ *for* $\alpha_0$*, any* $\alpha := (\alpha_1, \ldots, \alpha_{2p}) \in \mathcal{B} \subset \mathbb{R}^{2p}$*, including* $\alpha_0$*, satisfies* $|\alpha|_\infty \leq C_1$*, for some constant* $C_1 > 0$*. Furthermore,* $\mathcal{M}(\alpha_0) \leq s_0$ *and* $\frac{s_0^2 |\delta_0|_1^2 \log p}{n} = o_p(1)$.
*(ii) The threshold variable* $Q_i$*,* $i = 1, ..., n$*, is continuously distributed with intensity function* $f_Q(\tau)$*. The parameter* $\tau_0$ *lies in* $\mathbb{T} = [t_0, t_1]$*, where* $0 < t_0 < t_1 < 1$.
*(iii) The covariates* $X_i$*,* $i = 1, ..., n$*, satisfy* $\max_{1\leq j\leq p} E\left[\left(X_i^{(j)}\right)^4\right] \leq C_2^4$ *and* $\min_{1\leq j\leq p}$ $E\left[\left(X_i^{(j)}(t_0)\right)^2\right] \geq C_3^2$*, for some constants* $C_2$ *and* $C_3$*. Additionally,* $E\left[X_i^{(j)} X_i^{(l)} | Q_i = \tau\right]$ *is continuous and bounded when* $\tau$ *is in a neighborhood of* $\tau_0$*, for all* $1 \leq j, l \leq p$.
*(iv) The error terms* $U_i$*,* $i = 1, ..., n$*, satisfy* $E(U_i | X_i, Q_i) = 0$ *and* $\max_{1\leq i\leq n} E(U_i^4) \leq C_4 < \infty$*, for a positive constant* $C_4$*. Additionally,* $E\left[X_i^{(j)} X_i^{(l)} U_i^2 | Q_i = \tau\right]$ *is continuous and bounded when* $\tau$ *is in a neighborhood of* $\tau_0$*, for all* $1 \leq j, l \leq p$.
*(v)* $\frac{\sqrt{EM_{UX}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$*, where* $M_{UX} = \max_{1\leq i\leq n} \max_{1\leq j\leq p} \left|U_i X_i^{(j)}\right|$.
*(vi)* $\frac{\sqrt{EM_{XX}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$*, where* $M_{XX} = \max_{1\leq i\leq n} \max_{1\leq j,l\leq p} \left|X_i^{(j)} X_i^{(l)}\right|$.

Assumption 1 imposes weaker conditions on covariates and error terms compared to the fixed covariates and gaussian errors in Lee et al. (2016) and the sub-gaussian covariates and errors in Callot et al. (2017), as we allow for heteroskedastic non-subgaussian error terms and non-subgaussian covariates.
The first part of Assumption 1 (i) restricts each slope parameter vector component. The second part implies that $s_0$ and $\|\delta_0\|_1$ can increase with n and that Assumption 6 in Lee et al. (2016) holds. [6] Assumption 1 (ii) ensures that there are no ties among the $Q_i$s. [7] Applying the Cauchy-Schwarz inequality under Assumptions 1 (iii) and (iv) yields $\max_{1\leq j,l\leq p} E\left[X_i^{(j)} X_i^{(l)}\right] \leq C_2^2$ uniformly in $i$; $\max_{1\leq j\leq p} \mathrm{Var}\left(U_i X_i^{(j)}\right)$,

[6] See Callot et al. (2017) for further dicussion.
[7] We maintain the dependence between $Q$ and $X$ in the proof, and we will show in Section 5.1 that the performance of our estimator does not depend on whether $Q$ is an element of $X$.

$\max_{1\le j,l\le p} \operatorname{Var}\left(X_i^{(j)} X_i^{(l)}\right)$, $\max_{1\le j,l\le p} \operatorname{Var}\left(X_i^{(j)} X_i^{(l)} \mathbf{1}\{Q_i < \tau_0\}\right)$, and $\max_{1\le j\le p}$ $\operatorname{Var}\left(X_i^{(j)}(t_0)\right)^2$ are bounded uniformly in $i$.

Define

$$\lambda = \frac{A}{\mu}\frac{\sqrt{\log p}}{\sqrt{n}} \tag{2.8}$$

as the tuning parameter in (2.4) for a constant $A \ge 0$ and a fixed constant $\mu \in (0,1)$.

**Lemma 1.** *Suppose that Assumption 1 holds. Let* $(\widehat{\alpha}, \widehat{\tau})$ *be the Lasso estimator defined by* (2.4)*. Then, with probability at least* $1 - C(logn)^{-1}$*, we have*

$$\left\|\widehat{f} - f_0\right\|_n \le \sqrt{(6+2\mu_2)C_1\sqrt{C_2^2 + \mu_1\lambda}}\sqrt{s_0\lambda}. \tag{2.9}$$

Lemma 1 provides a non-asymptotic upper bound on the prediction error, regardless of whether the specification is a linear or threshold regression, as in Theorem 1 of Lee et al. (2016). The prediction error is consistent as $n \to \infty$, $p \to \infty$, and $s_0\lambda \to 0$. This lemma plays an important role in deriving the oracle inequalities in Theorem 1 for linear models and Theorem 2 for threshold models.

Next, we address the standard assumptions in high-dimensional regression models. To this end, we define the population covariance matrix $\boldsymbol{\Sigma}(\tau) = E\left[1/n\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]$, $\boldsymbol{M} = E[1/n\sum_{i=1}^n X_iX_i']$, $\boldsymbol{M}(\tau) = E\left[1/n\sum_{i=1}^n X_i(\tau)X_i(\tau)'\right]$, and $\boldsymbol{N}(\tau) = \boldsymbol{M} - \boldsymbol{M}(\tau)$. $\boldsymbol{\Sigma}(\tau)$ can be represented as a $2\times 2$ matrix due to the properties of the indicator function, i.e.,

$$\boldsymbol{\Sigma}(\tau) = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{M}(\tau) \\ \boldsymbol{M}(\tau) & \boldsymbol{M}(\tau) \end{bmatrix}.$$

Meanwhile, we define the following population uniform adaptive restricted eigenvalue and impose certain assumptions,

$$\kappa(s_0, c_0, \mathbb{S}, \boldsymbol{\Sigma}) = \min_{\tau\in\mathbb{S}} \min_{J_0\subset\{1,\ldots,2p\},|J_0|\le s_0} \min_{\gamma\neq 0,\|\gamma_{J_0^c}\|_1\le c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{\left(\gamma' E\left[1/n\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma\right)^{1/2}}{|\gamma_{J_0}|_2}.$$

**Assumption 2.** *(i)* $\boldsymbol{M}(\tau)$ *and* $\boldsymbol{N}(\tau)$ *are non-singular.*
*(ii) (Uniform Adaptive Restricted Eigenvalue Condition) For some integer* $s_0$ *such that* $\mathcal{M}(\alpha_0) \le s_0 < p$*, for a positive number* $c_0$*, and some set* $\mathbb{S} \subset \mathbb{R}$*, the following*

*condition holds*

$$\kappa(s_0, c_0, \mathbb{S}, \boldsymbol{\Sigma}) > 0. \tag{2.10}$$

Assumption 2 (i) is a standard assumption for model estimation. Assumption 2 (ii) is a uniform adaptive restricted eigenvalue condition, which is a common and high-level condition in the literature of high-dimensional econometrics and statistics. This condition can be relaxed if $\boldsymbol{\Sigma}(\tau)$ has full rank. Moreover, $\boldsymbol{\Sigma}(\tau)$ is invertible by applying Theorem 2.1 (ii) in Lu and Shiou (2002) under Assumption 2 (i). We then can do the gaussian elimination to obtain

$$\boldsymbol{\Theta}(\tau) := \boldsymbol{\Sigma}(\tau)^{-1} = \begin{bmatrix} \boldsymbol{N}(\tau)^{-1} & -\boldsymbol{N}(\tau)^{-1} \\ -\boldsymbol{N}(\tau)^{-1} & \boldsymbol{M}(\tau)^{-1} + \boldsymbol{N}(\tau)^{-1} \end{bmatrix}. \tag{2.11}$$

Thus, Assumption 2 (ii) holds under the non-singularity conditions for $\boldsymbol{M}(\tau)$ and $\boldsymbol{N}(\tau)$. Lemma 7 shows that $1/n\boldsymbol{X}(\tau)\boldsymbol{X}(\tau)'$ uniformly converges to $\boldsymbol{\Sigma}(\tau)$; therefore, the empirical adaptive restricted condition holds as stated in Lemma 8.

Given that $\tau_0$ is unknown, we impose that the restricted eigenvalue condition holds uniformly over $\tau$. Here, we analyze two separate cases. When $\delta_0 = 0$, Assumption 2 is required to hold uniformly with $\mathbb{S} = \mathbb{T}$, the entire parameter space for $\tau_0$, since $\tau_0$ is not identified. When $\delta_0 \neq 0$, this condition needs to hold uniformly in a neighborhood of $\tau_0$, as $\tau_0$ can be identified. The Uniform Adaptive Restricted Eigenvalue (UARE) Condition is applied to tighten the bound in Lemma 1 for establishing the oracle inequalities for the prediction error as well as the $\ell_1$ estimation error. Although we consider two cases separately, similar to Lee et al. (2016), we can make predictions and estimate $\alpha_0$ without pretesting the existence of the threshold effect.

### 2.2.1 Case I. No Threshold Effect.

In the case where $\delta_0 = 0$, the true model simplifies to a linear model $Y_i = X_i'\beta_0 + U_i$. The model (2.2) is thus much more over-parameterized, but we can still estimate the slope parameter vector $\alpha_0$ precisely, as shown in Theorem 1.

**Theorem 1.** *Supposed that $\delta_0 = 0$ and that Assumptions 1-2 hold with $\kappa = \kappa\left(s_0, \frac{1+\mu}{1-\mu}, \mathbb{T}, \boldsymbol{\Sigma}\right)$. Let $(\widehat{\alpha}, \widehat{\tau})$ be the Lasso estimator from* (2.4) *with $\lambda$ satisfying* (2.8).

*Then, as $n \to \infty$, with probability at least $1 - C(logn)^{-1}$, we have*

$$\left\|\widehat{f} - f_0\right\|_n \leq \frac{2\sqrt{2}}{\kappa}\left(\sqrt{C_2^2 + \mu_1\lambda}\right)\sqrt{s_0}\lambda,$$
$$|\widehat{\alpha} - \alpha_0|_1 \leq \frac{4\sqrt{2}}{(1-\mu)\,\kappa^2}\frac{C_2^2 + \mu_1\lambda}{\sqrt{C_3^2 - \mu_1\lambda}}s_0\lambda.$$

*Furthermore, these bounds are valid uniformly over the $l_0$-ball*

$$\mathcal{A}_{\ell_0}^{(1)}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid |\beta_0|_\infty \leq C_1, \mathcal{M}(\beta_0) \leq s_0, \delta_0 = 0\right\}.$$

#### 2.2.2 Case II. Fixed Threshold Effect.

In the case where $\delta_0 \neq 0$, we assume that the true model has a well-identified and fixed threshold effect.

**Assumption 3** (Identifiability under Sparsity and Discontinuity of Regression)**.** *For a given $s_0 \geq \mathcal{M}(\alpha_0)$, and for any $\eta$ and $\tau$ such that $\eta < |\tau - \tau_0|$ and $\alpha \in \{\alpha : \mathcal{M}(\alpha) \leq s_0\}$, there exists a constant $C_4 > 0$ such that, with probability approaching one,*

$$\left\|f_{(\alpha,\tau)} - f_0\right\|_n^2 > C_4\eta.$$

Assumption 3 states identifiability of $\tau_0$. Its validity was studied in Appendix B.1 (pages A7–A8) of Lee et al. (2016) when Assumption 1 holds. [8] When $\tau_0$ is known, we only need the UARE condition to hold uniformly in a neighborhood of $\tau_0$. We derive an upper bound for $|\widehat{\tau} - \tau_0|$ in Lemma 10 and thus define

$$\eta^* = \frac{2(3+\mu_2)C_1}{C_4}\sqrt{C_2^2 + \mu_1\lambda}s_0\lambda, \quad \mathbb{S} = \{|\tau - \tau_0| \leq \eta^*\}.$$

**Assumption 4** (Smoothness of Design)**.** *For any $\eta > 0$, there exists a constant $C_5 < \infty$ such that with probability to one,*

$$\sup_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|<\eta}\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}X_i^{(l)}\right|\left|\mathbf{1}\{Q_i < \tau_0\} - \mathbf{1}\{Q_i < \tau\}\right| \leq C_5\eta, \tag{2.12}$$

[8] We omit the restriction $\eta \geq \min_i |Q_i - \tau_0|$ since $\eta > \min_i |Q_i - \tau_0|$ holds in the random design with a continuous threshold variable $Q$.

**Assumption 5** (Well-defined second moments)**.** *For any* $\eta$ *such that* $1/n \leq \eta \leq \sqrt{(6+2\mu_2)C_1\sqrt{C_2^2+\mu_1\lambda}\sqrt{s_0}\lambda}$, $h_n^2(\eta)$ *is bounded with probability approaching one, where*

$$h_n^2(\eta) = \frac{1}{2n\eta} \sum_{i=\min\{1,[n(\tau_0-\eta)]\}}^{\max\{[n(\tau_0+\eta]),n\}} \left(X_i'\delta_0\right)^2, \tag{2.13}$$

*and* $[\cdot]$ *denotes an integer part of any real number.*

Assumptions 4 and 5 are similar to those in Lee et al. (2016) and Callot et al. (2017). Lemma 11 shows that Assumption 4 holds automatically under Assumptions 1 and 5.

**Theorem 2.** *Suppose that* $\delta_0 \neq 0$ *and that Assumptions 1 and 2 hold with* $\kappa = \kappa(s_0, \frac{2+\mu}{1-\mu}, \mathbb{S}, \mathbf{\Sigma})$. *Furthermore, suppose that Assumptions 3, 4 and 5 hold. Let* $(\widehat{\alpha}, \widehat{\tau})$ *be the Lasso estimator from* (2.4) *with* $\lambda$ *satisfying* (2.8). *Then,* $n \to \infty$, *with probability at least* $1 - C(logn)^{-1}$, *we have*

$$\begin{aligned}
\left\|\widehat{f} - f_0\right\|_n &\leq 6\frac{\sqrt{C_2^2+\mu\lambda}}{\kappa}\sqrt{s_0}\lambda, \\
|\widehat{\alpha} - \alpha_0|_1 &\leq \frac{36(C_2^2+\mu\lambda)}{\kappa^2(1-\mu)\sqrt{C_3^2-\mu\lambda}} s_0\lambda, \\
|\widehat{\tau} - \tau_0| &\leq \left(\frac{3\,(1+\mu)\,\sqrt{(C_2^2+\mu\lambda)}}{(1-\mu)\sqrt{(C_3^2-\mu\lambda)}} + 1\right)\frac{12(C_2^2+\mu\lambda)}{\kappa^2 C_4} s_0\lambda^2.
\end{aligned}$$

*Furthermore, these bounds are valid uniformly over the* $l_0$*-ball*

$$\mathcal{A}_{\ell_0}^{(2)}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid |\alpha_0|_\infty \leq C_1, \mathcal{M}\left(\alpha_0\right) \leq s_0, \delta_0 \neq 0\right\}.$$

The bound on the prediction norm from Theorem 1 or 2 is much smaller than that in Lemma 1. Compared to the oracle inequality in the literature related to high-dimensional linear models (e.g. Bickel et al. (2009), van de Geer et al. (2014)), Theorem 1 provides results of the same magnitude, implying that our estimation method can accommodate the linear model even though our model is much more overparameterized. Additionally, the inequalities from Theorem 1 and Theorem 2 are of the same magnitude, up to a constant. These results thus hold uniformly over

$\mathcal{B}_{\ell_0}(s_0)$, where

$$\mathcal{B}_{\ell_0}(s_0) = \mathcal{A}^{(1)}_{\ell_0}(s_0) \cup \mathcal{A}^{(2)}_{\ell_0}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid |\alpha_0|_\infty \leq C, \mathcal{M}(\alpha_0) \leq s_0\right\}.$$

For the super-consistency result of $\widehat{\tau}$, Lee et al. (2016) explains that the least squares objective function is locally linear rather than locally quadratic around $\tau_0$.

The main contribution of this section is our extension of the inequality results from Lee et al. (2016) to a high-dimensional threshold regression with non-subgaussian random regressors and heteroskedastic non-subgaussian error terms. We then apply these oracle inequalities to develop the uniform inference theory.

# 3 The Debiased Lasso Estimator and Uniform Inference

## 3.1 The Debiased Lasso Estimator

To perform uniform inference for the slope parameters, we first construct the desparsified Lasso estimator proposed by van de Geer et al. (2014) in our high-dimensional threshold regression model as follows: [9]

$$\widehat{a}(\widehat{\tau}) = \widehat{\alpha}(\widehat{\tau}) + \widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'(Y - \boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}))/n, \tag{3.1}$$

where $\widehat{\alpha}(\widehat{\tau})$ is obtained from (2.4), and $\widehat{\mathbf{\Theta}}(\widehat{\tau})$ is an approximate inverse of the empirical Gram matrix $\widehat{\mathbf{\Sigma}}(\widehat{\tau}) = \boldsymbol{X}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'/n$ because $\widehat{\mathbf{\Sigma}}(\widehat{\tau})$ is singular in our high-dimensional model.

We will discuss the debiased Lasso estimator in two separate cases. Additionally, we will apply nodewise regression from Meinshausen and Bühlmann (2006) to construct the approximate inverse matrix.

[9] This estimator is obtained by inverting the Karush-Kuhn-Tucker (KKT) conditions. For more details, see van de Geer et al. (2014).

### 3.1.1 Case I. No Threshold Effect

In the case where $\delta_0 = 0$, the true model simplifies to a linear model $Y = X\beta_0 + U$. Substituting this into (3.1) yields

$$
\begin{aligned}
\widehat{a}(\widehat{\tau}) &= \widehat{\alpha}(\widehat{\tau}) + \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'(X\beta_0 + U - \boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}))/n \\
&= \widehat{\alpha}(\widehat{\tau}) + \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'(\boldsymbol{X}(\widehat{\tau})\alpha_0 + U - \boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}))/n \\
&= \alpha_0 - \alpha_0 + \widehat{\alpha}(\widehat{\tau}) - \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}(\widehat{\tau})(\widehat{\alpha}(\widehat{\tau}) - \alpha_0) + \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n \\
&= \alpha_0 + \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n - \Delta(\widehat{\tau})/n^{1/2},
\end{aligned} \tag{3.2}
$$

where $\Delta(\tau) = \sqrt{n}\left(\widehat{\boldsymbol{\Theta}}(\tau)\widehat{\boldsymbol{\Sigma}}(\tau) - I_{2p}\right)(\widehat{\alpha}(\widehat{\tau}) - \alpha_0)$. The second equality holds due to $\delta_0 = 0$.

To derive the asymptotic distribution of tests involving an increasing number of parameters, we define a $(2p \times 1)$ vector $g$ with $|g|_2 = 1$ and let $H = \{j = 1, \ldots, 2p \mid g_j \neq 0\}$ with cardinality $|H| = h < p$. $H$ contains the indices of the coefficients involved in the hypothesis to be tested. We allow for $h \to \infty$ but require $h/n \to 0$, as $n \to \infty$. By the Cauchy–Schwarz inequality, we have $|g|_1 \leq \sqrt{h}$. In particular, $g = e_j$ represents the case where we test only a single coefficient, where $e_j$ is the $2p \times 1$ unit vector with the $j$-th element being one.

Our focus is on

$$
\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0) = g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\Delta(\widehat{\tau}), \tag{3.3}
$$

and we will derive its asymptotic distribution by applying a central limit theorem to $g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}$ and by showing that $g'\Delta(\widehat{\tau})$ is asymptotically negligible.

### 3.1.2 Case II. Fixed Threshold Effect

This subsection explores the case where the threshold effect is well-identified and fixed. Following a procedure similar to Section 3.1.1, we substitute $Y = \boldsymbol{X}(\tau_0)\alpha_0 + U$ into (3.1), yielding

$$
\begin{aligned}
\widehat{\alpha}(\widehat{\tau}) = \alpha_0 &+ \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'\left(\boldsymbol{X}(\tau_0)\alpha_0 - \boldsymbol{X}(\widehat{\tau})\alpha_0\right)/n \\
&- \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\lambda\boldsymbol{D}(\widehat{\tau})\widehat{\rho} + \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n - \Delta(\widehat{\tau})/n^{1/2}.
\end{aligned} \tag{3.4}
$$

In this case, our focus is on

$$\sqrt{n}g'(\widehat{a}(\widehat{\tau})-\alpha_0) = g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) \\ + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}, \tag{3.5}$$

and we will derive its asymptotic distribution by applying a central limit theorem to $g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}$ and by showing that $g'\widehat{\mathbf{\Theta}}\,(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\,\alpha_0/n^{1/2}$ and $g'\Delta(\widehat{\tau})$ are asymptotically negligible.

### 3.1.3 Constructing the Approximate Inverse $\widehat{\mathbf{\Theta}}(\tau)$

In this subsection, we formalize the process of constructing the approximate inverse matrix $\widehat{\mathbf{\Theta}}(\tau)$ of the singular empirical Gram matrix. The approach closely follows that of van de Geer et al. (2014), with the additional requirement of verifying that the specified conditions are met.

We seek a well-behaved $\widehat{\mathbf{\Theta}}(\tau)$ and examine the asymptotic properties of $\widehat{\mathbf{\Theta}}(\tau)$ uniformly across $\tau \in \mathbb{T}$. Recalling (2.11), we have

$$\mathbf{\Theta}(\tau) = \mathbf{\Sigma}(\tau)^{-1} = \begin{bmatrix} \boldsymbol{N}(\tau)^{-1} & -\boldsymbol{N}(\tau)^{-1} \\ -\boldsymbol{N}(\tau)^{-1} & \boldsymbol{M}(\tau)^{-1} + \boldsymbol{N}(\tau)^{-1} \end{bmatrix}.$$

Define $\boldsymbol{A}(\tau) = \boldsymbol{M}(\tau)^{-1}$ and $\boldsymbol{B}(\tau) = \boldsymbol{N}(\tau)^{-1}$. We construct the approximate inverse $\widehat{\boldsymbol{A}}(\tau)$ of $\widehat{\boldsymbol{M}}(\tau)$ and $\widehat{\boldsymbol{B}}(\tau)$ of $\widehat{\boldsymbol{N}}(\tau)$, where $\widehat{\boldsymbol{M}}(\tau) = \frac{1}{n}\sum_{i=1}^n X_i X_i{}'\mathbf{1}\{Q_i < \tau\}$ and $\widehat{\boldsymbol{N}}(\tau) = \frac{1}{n}\sum_{i=1}^n X_i X_i{}'\mathbf{1}\{Q_i \geq \tau\}$, to build the approximate inverse matrix $\widehat{\mathbf{\Theta}}(\tau)$. Denote the $(p\times 1)$ vector by $\widetilde{X}_i^{(j)}(\tau) = X_i^{(j)}\mathbf{1}\{Q_i \geq \tau\}$ and the $(n\times p)$ matrix by $\widetilde{X}(\tau)$. Let $X^{(-j)}(\tau)$ and $\widetilde{X}^{(-j)}(\tau)$ denote the submatrices of $X(\tau)$ and $\widetilde{X}(\tau)$, respectively, without the $j$-th column. We study the following nodewise regression models with covariates orthogonal to the error terms in $L_2$ for all $j = 1, \ldots, p$, we consider

$$X^{(j)}(\tau) = X^{(-j)}(\tau)'\gamma_{0,j}(\tau) + \upsilon^{(j)},$$

$$\widetilde{X}^{(j)}(\tau) = \widetilde{X}^{(-j)}(\tau)'\widetilde{\gamma}_{0,j}(\tau) + \widetilde{\upsilon}^{(j)}.$$ [10]

$\upsilon^{(j)}$ and $\widetilde{\upsilon}^{(j)}$ may be functions of $\tau$ even though they are independence of $Q$.

[10] See Appendix B of Caner and Kock (2018) for the details about the covariance matrix's representation of the regression coefficients.

We then impose the following assumption to control the tail distribution of $\left|\upsilon_i^{(j)} X_i^{(l)}\right|$ and $\left|\widetilde{\upsilon}_i^{(j)} X_i^{(l)}\right|$, allowing us to apply the oracle inequalities in Section 2.2 to the nodewise regressions.

**Assumption 6.** *(i)* $\max_{1\le j\le p} |\gamma_{0,j}|_\infty \le C$ *and* $\max_{1\le j\le p} |\widetilde{\gamma}_{0,j}|_\infty \le C'$, *for some positive constants* $C$ *and* $C'$.
*(ii) For* $i = 1, ..., n$, *and* $j = 1, ..., p$, $E\left[\upsilon_i^{(j)} \middle| X_i, Q_i\right] = 0$ *and* $E\left[\widetilde{\upsilon}_i^{(j)} \middle| X_i, Q_i\right] = 0$; $E\left[\left(\upsilon_i^{(j)}\right)^2\right]$ *and* $E\left[\left(\widetilde{\upsilon}_i^{(j)}\right)^2\right]$ *are uniformly bounded in* $j = 1, \dots, p$.
*(iii)* $\frac{\sqrt{EM_{\upsilon X}^2}\sqrt{\log p}}{\sqrt{n}} < \infty$ *and* $\frac{\sqrt{EM_{\widetilde{\upsilon} X}^2}\sqrt{\log p}}{\sqrt{n}} < \infty$, *where* $M_{\upsilon X} = \max_{1\le i\le n}\max_{1\le l\le p}\left|\upsilon_i^{(j)} X_i^{(l)}\right|$ *and* $M_{\widetilde{\upsilon} X} = \max_{1\le i\le n}\max_{1\le l\le p}\left|\widetilde{\upsilon}_i^{(j)} X_i^{(l)}\right|$.

Now, we begin constructing $\widehat{\boldsymbol{A}}(\tau)$. Given any $\tau \in \mathbb{T}$, for each $j = 1, ..., p$, the Lasso estimator for the nodewise regression is given by

$$\widehat{\gamma}_j(\tau) = \operatorname{argmin}_{\gamma\in\mathbb{R}^{p-1}} \|X^{(j)}(\tau) - X^{(-j)}(\tau)\gamma_j\|_n^2 + \lambda_{node,j}\left|\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}(\tau)\gamma_j\right|_1, \tag{3.6}$$

where $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}(\tau) := \operatorname{diag}\left\{\left\|X^{(l)}(\tau)\right\|_n, l = 1, ..., p, l \ne j\right\}$, with components of $\widehat{\gamma}_j(\tau) = \{\widehat{\gamma}_j^{(k)}(\tau);\ k = 1, ..., p,\ k \ne j\}$. We choose $\lambda_{node,j} = \lambda_{node}$, required for the validity of Lemma 16. Define

$$\widehat{\boldsymbol{C}}(\tau) = \begin{pmatrix} 1 & -\widehat{\gamma}_1^{(2)}(\tau) & \cdots & -\widehat{\gamma}_1^{(p)}(\tau) \\ -\widehat{\gamma}_2^{(1)}(\tau) & 1 & \cdots & -\widehat{\gamma}_2^{(p)}(\tau) \\ \dots & \dots & \ddots & \dots \\ -\widehat{\gamma}_p^{(1)}(\tau) & -\widehat{\gamma}_p^{(2)}(\tau) & \cdots & 1 \end{pmatrix}. \tag{3.7}$$

and $\widehat{\boldsymbol{Z}}(\tau)^2 = diag\left(\widehat{z}_1(\tau)^2, \dots, \widehat{z}_p(\tau)^2\right)$, where

$$\widehat{z}_j(\tau)^2 = \left|\left|X^{(j)}(\tau) - X^{(-j)}(\tau)\widehat{\gamma}_j(\tau)\right|\right|_n^2 + \lambda_{node}\left|\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}(\tau)\widehat{\gamma}_j(\tau)\right|_1. \tag{3.8}$$

We thus construct

$$\widehat{\boldsymbol{A}}(\tau) = \widehat{\boldsymbol{Z}}(\tau)^{-2}\widehat{\boldsymbol{C}}(\tau). \tag{3.9}$$

Next, we show that $\widehat{\boldsymbol{A}}(\tau)$ is an approximate inverse matrix of $\widehat{\boldsymbol{M}}(\tau)$. Let $\widehat{A}_j(\tau)$ denote the $j$-th row of $\widehat{\boldsymbol{A}}(\tau)$. Thus, $\widehat{A}_j(\tau) = \widehat{C_j}(\tau)/\widehat{z}_j(\tau)^2$. Denoting by $\widetilde{e}_j$ the $j$-th unit vector, the KKT conditions imply that

$$\left|\widehat{A}_j(\tau)'\widehat{\boldsymbol{M}}(\tau) - \widetilde{e}_j'\right|_\infty \leq \left|\widehat{\boldsymbol{\Gamma}}_j(\tau)\right| \frac{\lambda_{node}}{\widetilde{z}_j(\tau)^2}. \tag{3.10}$$

Similarly, we use the same process to construct $\widehat{\boldsymbol{B}}(\tau)$. Given any $\tau \in \mathbb{T}$, for each $j = 1, ..., p$, define

$$\begin{aligned}
\widehat{\widetilde{\gamma}}_j(\tau) &= \operatorname{argmin}_{\gamma \in \mathbb{R}^{p-1}} \left|\left|\widetilde{X}^{(j)}(\tau) - \widetilde{X}^{(-j)}(\tau)'\widetilde{\gamma}_j\right|\right|_n^2 + \lambda_{\text{node}} \left|\widehat{\widetilde{\boldsymbol{\Gamma}}}_{\boldsymbol{j}}(\tau)\widetilde{\gamma}_j\right|_1, \\
\widehat{\widetilde{\boldsymbol{\Gamma}}}_{\boldsymbol{j}}(\tau) &= \operatorname{diag}\left\{\left\|\widetilde{X}^{(l)}(\tau)\right\|_n, \ l = 1, \ldots, p, \ l \neq j\right\},
\end{aligned}$$

with components of $\widehat{\widetilde{\gamma}}_j(\tau) = \{\widehat{\widetilde{\gamma}}_j^{(k)}(\tau) : k = 1, ..., p, k \neq j\}$. Meanwhile, define

$$\widehat{\widetilde{\boldsymbol{C}}}(\tau) = \begin{pmatrix} 1 & -\widehat{\widetilde{\gamma}}_1^{(2)}(\tau) & \cdots & -\widehat{\widetilde{\gamma}}_1^{(p)}(\tau) \\ -\widehat{\widetilde{\gamma}}_2^{(1)}(\tau) & 1 & \cdots & -\widehat{\widetilde{\gamma}}_2^{(p)}(\tau) \\ \cdots & \cdots & \ddots & \cdots \\ -\widehat{\widetilde{\gamma}}_p^{(1)}(\tau) & -\widehat{\widetilde{\gamma}}_p^{(2)}(\tau) & \cdots & 1 \end{pmatrix} \tag{3.11}$$

and $\widehat{\widetilde{\boldsymbol{Z}}}(\tau)^2 = \operatorname{diag}\left(\widehat{\widetilde{z}}_1(\tau)^2, \ldots, \widehat{\widetilde{z}}_p(\tau)^2\right)$ with

$$\widehat{\widetilde{z}}_j(\tau)^2 = \|\widetilde{X}^{(j)}(\tau) - \widetilde{X}^{(-j)}(\tau)'\widehat{\widetilde{\gamma}}_j(\tau)\|_n^2 + \lambda_{\text{node}} \left|\widehat{\widetilde{\boldsymbol{\Gamma}}}_j(\tau)\widehat{\widetilde{\gamma}}_j(\tau)\right|_1.$$

We then construct

$$\widehat{\boldsymbol{B}}(\tau) = \widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\widetilde{\boldsymbol{C}}}(\tau). \tag{3.12}$$

Therefore, we obtain

$$\widehat{\boldsymbol{\Theta}}(\tau) = \begin{bmatrix} \widehat{\boldsymbol{B}}(\tau) & -\widehat{\boldsymbol{B}}(\tau) \\ -\widehat{\boldsymbol{B}}(\tau) & \widehat{\boldsymbol{A}}(\tau) + \widehat{\boldsymbol{B}}(\tau) \end{bmatrix}, \tag{3.13}$$

and provide the asymptotic properties of $\widehat{\boldsymbol{\Theta}}(\tau)$ in Lemma 16.

### 3.2 Uniform Inference for the Debiased Lasso Estimator

This section derives the asymptotic distribution of tests in cases with no threshold effect and a fixed threshold effect, showing that their distributions are the same. Furthermore, we construct uniform confidence intervals for the parameters of interest that contract at the optimal rate. To this end, we first impose some assumptions to establish the validity of the asymptotically gaussian inference.

**Assumption 7.** *(i)* $\max_{1\le j\le p} E\left[\left(X_i^{(j)}\right)^{12}\right]$ *and* $E\left[U_i^8\right]$ *are bounded uniformly in* $i$.
$\frac{\sqrt{EM_{X^6}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$, $\frac{\sqrt{EM_{X^2U^2}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$ *and* $\frac{\sqrt{EM_{X^4U^2}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$,
*where* $M_{X^6} = \max_{1\le i\le n}\max_{1\le k,l,j\le p}\left|\left(X_i^{(k)}X_i^{(l)}X_i^{(j)}\right)^2\right|$, $M_{X^2U^2} = \max_{1\le i\le n}\max_{1\le j,l\le p}\left|X_i^{(j)}X_i^{(l)}U_i^2\right|$, *and* $M_{X^4U^2} = \max_{1\le i\le n}\max_{1\le j,l\le p}\left|\left(X_i^{(j)}X_i^{(l)}U_i\right)^2\right|$.
*(ii)*

$$(h)^{\frac{3}{2}}s_0^2\bar{s}^2\frac{\log p}{\sqrt{n}} = o_p(1); \quad \frac{(h\bar{s})^3}{n} = o_p(1).$$

*(iii)* $\kappa(s_0, c_0, \mathbb{T}, \mathbf{\Sigma}_{xu})$ *and* $\kappa(s_0, c_0, \mathbb{T}, \mathbf{\Sigma})$ *are bounded away from zero.* $\phi_{\max}(\mathbf{\Sigma}_{xu}(\tau))$ *and* $\phi_{\max}(\mathbf{\Sigma}(\tau))$ *are bounded from above, for* $\tau \in \mathbb{T}$.

Assumption 7 provides sufficient conditions for applying the central limit theorem. Assumption 7 (i) controls the tail behavior of the covariates and the error terms. By Assumption 7 (i), $\max_{1\le j,l\le p}\mathrm{Var}\left(X_i^{(j)}X_i^{(l)}U_i^2\right)$, $\max_{1\le k,l,j\le p}\mathrm{Var}\left(X_i^{(k)}X_i^{(l)}X_i^{(j)}\right)^2$, and $\max_{1\le j,l\le p}\mathrm{Var}\left(X_i^{(j)}X_i^{(l)}U_i\right)^2$ are bounded from above uniformly in $i$. Assumption 7 (ii) restricts the number of covariates ($p$), the number of parameters included in conducting joint inference ($h$), the sparsity of the population covariance matrix ($\bar{s}$), where $\bar{s} = \sup_{\tau\in\mathbb{T}}\max_{j\in H} s_j(\tau)$, $s_j(\tau) = |S_j(\tau)|$, and $S_j(\tau) = \{i = 1, ..., 2p : \Theta_{j,i}(\tau) \neq 0\}$, and the sparsity of the slope parameters ($s_0$). Notably, the second part of Assumption 7 (ii) is to verify the Lyapunov condition. Assumption 7 (iii) restricts the eigenvalues of $\mathbf{\Sigma}_{xu}(\tau)$ and $\mathbf{\Sigma}(\tau)$, where $\mathbf{\Sigma}_{xu}(\tau) = E\left[1/n\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i'(\tau)U_i^2\right]$.

**Theorem 3.** *Suppose that Assumptions 1 to 7 hold. Then, as* $n \to \infty$, *we have*

$$\frac{\sqrt{n}g'\left(\widehat{a}\left(\widehat{\tau}\right) - \alpha_0\right)}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \xrightarrow{d} N(0,1), \tag{3.14}$$

*uniformly in* $\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)$.

*Furthermore,*

$$\sup_{\alpha_0 \in \mathcal{A}_{\ell_0}^{(1)}(s_0)} \left| g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g \right| = o_p(1) \tag{3.15}$$

$$\sup_{\alpha_0 \in \mathcal{A}_{\ell_0}^{(2)}(s_0)} \left| g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\tau_0)\mathbf{\Theta}(\tau_0)'g \right| = o_p(1), \tag{3.16}$$

*where* $\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})' \left(\widehat{U}_i(\widehat{\tau})\right)^2$.

Thus, we establish the asymptotic distribution of tests involving an increasing number of slope parameters in the cases with no threshold effect and a fixed threshold effect, showing that their asymptotic distributions are identical. Additionally, we provide a uniformly consistent covariance matrix estimator for both cases. However, there is a slight difference between the limits of their asymptotic variances, since there is a true value for the threshold parameter in the latter case. Because we lack prior knowledge of the existence of a threshold effect, we simultaneously impose the assumptions of Theorems 1 and 2 to establish Theorem 3. Here, the number of parameters involved in hypotheses is allowed to grow to infinity at a rate restricted by Assumption 7(ii).

In the case of a fixed number of the parameters being tested, by (3.14), we have

$$\left| \left(\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'\right)_{H,H}^{-\frac{1}{2}} \sqrt{n}(\widehat{a}(\widehat{\tau})_H - \alpha_{0,H}) \right|_2^2 \xrightarrow{d} \chi^2(h), \tag{3.17}$$

for a fixed cardinality $h$. Thus, a $\chi^2$ test can be applied to test a hypothesis involving $h$ parameters simultaneously.

Next, we establish confidence intervals for the parameters of interest. Let $\Phi(t)$ denote the cumulative distribution function (CDF) of the standard normal distribution, and let $z_{1-\alpha/2}$ be the $1-\alpha/2$ percentile of the standard normal distribution. Define $\widehat{\sigma}_j(\widehat{\tau}) = \sqrt{e_j'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'e_j}$ for all $j \in \{1, ..., 2p\}$. Let $\mathrm{diam}([a, b])$ denote the length of the interval $[a, b] \subset \mathbb{R}$.

**Theorem 4.** *Suppose that Assumptions 1, 2, 3, 4, 6 and 7 hold. Then, as* $n \to \infty$, *we have*

$$\sup_{t\in\mathbb{R}}\sup_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)}\left|\mathbb{P}\left\{\frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau})-\alpha_0)}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}}\leq t\right\}-\Phi(t)\right|\rightarrow 0. \tag{3.18}$$

*Furthermore, for all* $j\in\{1,\ldots,2p\}$,

$$\lim_{n\rightarrow\infty}\inf_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)}\mathbb{P}\left\{\alpha_0^{(j)}\in\left[\widehat{a}^{(j)}(\widehat{\tau})-z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_j(\widehat{\tau})}{\sqrt{n}},\widehat{a}^{(j)}(\widehat{\tau})+z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_j(\widehat{\tau})}{\sqrt{n}}\right]\right\}=1-\alpha, \tag{3.19}$$

*and*

$$\sup_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)} diam\left(\left[\widehat{a}^{(j)}(\widehat{\tau})-z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_j(\widehat{\tau})}{\sqrt{n}},\widehat{a}^{(j)}(\widehat{\tau})+z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_j(\widehat{\tau})}{\sqrt{n}}\right]\right)=O_p\left(\frac{1}{\sqrt{n}}\right). \tag{3.20}$$

Therefore, we show that the convergence of a linear combination of the parameters of $\widehat{a}(\widehat{\tau})$ to the standard normal distribution is uniformly valid over the $\ell_0$-ball with a radius of at most $s_0$. Researchers can perform uniform inference for high-dimensional slope parameters without specifying whether the specification is a linear or threshold regression. In addition, these confidence intervals are asymptotically honest and contract at the optimal rate.

# 4 Time Series Model

We establish the uniform inference theory for the debiased Lasso estimator in the high-dimensional time series threshold regression model, extending the model of Adamek et al. (2023) by allowing for the existence of a threshold effect, with local projection threshold regression as a special case. We now assume that $(Y_i, X_i, Q_i)$ is a sequence of dependent data while still focusing on the model (2.2) as follows:

$$Y_i=\boldsymbol{X}_i(\tau_0)'\alpha_0+U_i,\quad i=1,\ldots,n. \tag{4.1}$$

## 4.1 Oracle Inequalities and Inference

We continue studying the Lasso estimator from equation (2.4). We follow the assumptions of Adamek et al. (2023), but list them here for completeness and briefly discuss them. In the notation, $C$ represents an arbitrary positive finite constant, and its value may vary from line to line.

**Assumption 8.** *Let* $\{X_i, Q_i, U_i\}_{i=1}^n$ *denote a sequence of dependent random variables. Define* $\boldsymbol{W}_i = (X_i', U_i)'$ *and suppose that there exist some constants* $\bar{m} > m > 2$*, and* $d \geq \max\{1, (\bar{m}/m - 1)/(\bar{m} - 2)\}$ *such that*
*(i)* $\mathrm{E}[\boldsymbol{W}_i] = 0$*,* $\mathrm{E}[X_i U_i] = 0$*,* $\mathrm{E}[Q_i U_i] = 0$*, and* $\max_{1 \leq j \leq p+1,\ 1 \leq i \leq N} E\left|W_i^{(j)}\right|^{2\bar{m}} \leq C$*, for a positive constant* $C$*.*
*(ii) Let* $\boldsymbol{s}_{N,i}$ *denote a* $k(N)$*-dimensional triangular array that is* $\alpha$*-mixing of size* $-d/(1/m - 1/\bar{m})$ *with* $\sigma$*-field* $\mathcal{F}_i^{\boldsymbol{s}} := \sigma\{\boldsymbol{s}_{n,i}, \boldsymbol{s}_{n,i-1}, \ldots\}$ *such that* $\boldsymbol{W}_i$ *is* $\mathcal{F}_i^{\boldsymbol{s}}$*-measurable. The process* $\left\{W_i^{(j)}\right\}$ *is* $L_{2m}$*-near-epoch-dependent (NED) of size* $-d$ *on* $\boldsymbol{s}_{N,i}$ *with positive bounded NED constants* [11]*, uniformly over* $j = 1, \ldots, n+1$*.*
*(iii) The threshold variable* $Q_i$*,* $i = 1, ..., n$*, is continuously distributed. The parameter* $\tau_0$ *lies in* $\mathbb{T} = [t_0, t_1]$*, where* $0 < t_0 < t_1 < 1$*.*
*(iv)* $E\left[X_i^{(j)} X_i^{(l)} | Q_i = \tau\right]$ *and* $E\left[X_i^{(j)} X_i^{(l)} U_i^2 | Q_i = \tau\right]$ *are continuous and bounded when* $\tau$ *is in a neighborhood of* $\tau_0$*, for all* $1 \leq j, l \leq p$*.*

This assumption is the time series analogue of Assumption 1. We assume $\boldsymbol{W}_i$ to be NED; thus $\boldsymbol{W}_i$ can be approximated by a mixing process. [12]

We assume that $\alpha_0$ is weakly sparse, which is a more general condition than the exact sparsity assumption for cross-sectional data. Additionally, establishing oracle inequalities under this condition for independent data is not difficult.

**Assumption 9.** *For some* $0 \leq r < 1$ *and sparsity level* $s_r$*, define the* $2p$*-dimensional sparse compact parameter space*

$$\mathcal{B}_{2p}(r, s_r) := \left\{\alpha \in \mathbb{R}^{2p} : |\alpha|_r^r \leq s_r,\ |\alpha|_\infty \leq C,\ \exists C < \infty\right\},$$

*and assume that* $\alpha_0 \in \mathcal{B}_{2p}(r, s_r)$*.*

In addition, we define $\mathcal{A}_{2p}^{(1)}(r, s_r) = \{\alpha \in \mathbb{R}^{2p} : |\alpha|_r^r \leq s_r,\ |\alpha|_\infty \leq C, \delta_0 = 0\}$ and $\mathcal{A}_{2p}^{(2)}(r, s_r)$ $= \{\alpha \in \mathbb{R}^{2p} : |\alpha|_r^r \leq s_r,\ |\alpha|_\infty \leq C, \delta_0 \neq 0\}$. Thus, $\mathcal{B}_{2p}(r, s_r) = \mathcal{A}_{2p}^{(1)}(r, s_r) \cup \mathcal{A}_{2p}^{(2)}(r, s_r)$.

We impose the following standard compatibility conditions for high-dimensional regression models, as in Adamek et al. (2023), which are stronger than the uniform

[11] A sequence $\boldsymbol{W}_N$ is of size $-d$ if $\boldsymbol{W}_N = O(N^{-d-\varepsilon})$ for some $\varepsilon > 0$.
[12] NED is a more general form of dependence; refer to Adamek et al. (2023).

adaptive restricted eigenvalue conditions in Assumption 2 (ii), while they are still considered regular conditions for establishing the consistency of the Lasso estimator.

**Assumption 10.** *Recall* $\mathbf{\Sigma}(\tau) := 1/n \sum_{i=1}^{n} \mathrm{E}\left[\boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]$. *For a general index set* $S$ *with cardinality* $|S|$, *define the compatibility constant*

$$\phi^2_{\mathbf{\Sigma}(\tau)}(S) := \min_{\tau \in \mathbb{S}} \min_{\{\gamma \neq 0 : |\gamma_{S^c}|_1 \leq 3|\gamma_S|_1\}} \left\{ \frac{|S|\gamma'\mathbf{\Sigma}(\tau)\gamma}{|\gamma_S|_1^2} \right\}.$$

*Assume that* $\phi^2_{\mathbf{\Sigma}(\tau)}(S_\lambda) \geq 1/C$, *which implies that*

$$|\gamma_{S_\lambda}|_1^2 \leq \frac{|S_\lambda|\gamma'\mathbf{\Sigma}(\tau)\gamma}{\phi^2_{\mathbf{\Sigma}(\tau)}(S_\lambda)} \leq C|S_\lambda|\gamma'\mathbf{\Sigma}(\tau)\gamma,$$

*for all* $\gamma$ *satisfying* $|\gamma_{S_\lambda^c}|_1 \leq 3|\gamma_{S_\lambda}|_1 \neq 0$.

The compatibility conditions for $\boldsymbol{M}(\tau)$ and $\boldsymbol{M}$ can be defined accordingly.

**Theorem 5.** *Suppose that Assumptions 3, 4, 8, 9, 10 and the conditions of Lemma 28 hold. Then, as* $n \to \infty$, *with probability at least* $1 - C \ln\ln n^{-1}$, *we have*

$$\left|\left|\widehat{f} - f_0\right|\right|_n^2 \leq C\lambda^{2-r}s_r,$$
$$|\widehat{\alpha} - \alpha_0|_1 \leq C\lambda^{1-r}s_r,$$

*when the fixed threshold effect exists,*

$$|\widehat{\tau} - \tau_0|_1 \leq C\lambda^{2-r}s_r.$$

Theorem 5 provides oracle inequalities for dependent data, in comparison to Theorems 1 and 2. We do not separate the results into two cases, as the oracle inequalities are qualitatively the same whether or not a fixed threshold effect exists. Additionally, we provide a non-asymptotic bound for $\widehat{\tau}$ when the threshold effect does exist.

We will establish the uniform asymptotic normality of the debiased Lasso estimator based on a finite number of parameters, a common consideration in time series inference. We can include an increasing number of tested parameters at the cost of assuming an $\alpha$-mixing process instead of the NED framework. [13] We utilize the

[13] See Section 4.3 in Adamek et al. (2023) for further discussion.

same nodewise regression approach as in Section 3.1.3 to construct the approximate inverses of the empirical Gram matrices $\widehat{\boldsymbol{M}}(\tau)$ and $\widehat{\boldsymbol{N}}(\tau)$, denoted by $\widehat{\boldsymbol{A}}(\tau)$ and $\widehat{\boldsymbol{B}}(\tau)$, for obtaining the debiased Lasso estimator. Thus, for all $j = 1, \ldots, p$,

$$X^{(j)}(\tau) = X^{(-j)}(\tau)'\gamma_{0,j}(\tau) + \upsilon^{(j)},$$

$$\widetilde{X}^{(j)}(\tau) = \widetilde{X}^{(-j)}(\tau)'\widetilde{\gamma}_{0,j}(\tau) + \widetilde{\upsilon}^{(j)}.$$

We provide the following assumptions for applying Theorem 5 to the nodewise Lasso regressions.

**Assumption 11.** *(i) For some $0 \le r < 1$ and sparsity levels $s_r^{(j)}$, $\widetilde{s}_r^{(j)}$, let $\gamma_{0,j} \in \mathcal{B}_{p-1}\left(r, s_r^{(j)}\right)$ and $\widetilde{\gamma}_{0,j} \in \mathcal{B}_{p-1}\left(r, \widetilde{s}_r^{(j)}\right)$, $\forall j = 1, ..., p$.*
*(ii) For $i = 1, ..., n$, and $j = 1, ..., p$, $E\left[\upsilon_i^{(j)}\right] = 0$; $E\left[\widetilde{\upsilon}_i^{(j)}\right] = 0$, $E\left[\upsilon_i^{(j)}\middle| X_i, Q_i\right] = 0$ and $E\left[\widetilde{\upsilon}_i^{(j)}\middle| X_i, Q_i\right] = 0$; $\max_{1\le j\le p,\ 1\le i\le n} E\left[\left|\upsilon_i^{(j)}\right|^{2\bar{m}}\right] \le C$, $\max_{1\le j\le p,\ 1\le i\le n} E\left[\left|\widetilde{\upsilon}_i^{(j)}\right|^{2\bar{m}}\right] \le C'$.*
*(iii) $\boldsymbol{M}(\tau)$ and $\boldsymbol{N}(\tau)$ are non-singular.*

Assumption 11 is analogous to Assumptions 4 and 5 in Adamek et al. (2023). Assumption 11 (iii) implies that $\boldsymbol{\Sigma}(\tau)$ is invertible, as discussed below Assumption 2. Therefore, we have

$$\widehat{\boldsymbol{\Theta}}(\tau) = \begin{bmatrix} \widehat{\boldsymbol{B}}(\tau) & -\widehat{\boldsymbol{B}}(\tau) \\ -\widehat{\boldsymbol{B}}(\tau) & \widehat{\boldsymbol{A}}(\tau) + \widehat{\boldsymbol{B}}(\tau) \end{bmatrix} = \begin{bmatrix} \widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\widetilde{\boldsymbol{C}}}(\tau) & -\widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\widetilde{\boldsymbol{C}}}(\tau) \\ -\widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\widetilde{\boldsymbol{C}}}(\tau) & \widehat{\boldsymbol{Z}}(\tau)^{-2}\widehat{\boldsymbol{C}}(\tau) + \widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\widetilde{\boldsymbol{C}}}(\tau) \end{bmatrix},$$

as in (2.11).

Recall the debiased Lasso estimators from Section 3: in the case of no threshold effect,

$$\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0) = g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\Delta(\widehat{\tau}),$$

and in the case of fixed threshold effect,

$$\begin{aligned}\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0) &= g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) \\ &\quad + g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}.\end{aligned}$$

Define $\boldsymbol{Z}(\tau_0)^2 = diag\left(z_1(\tau_0)^2, \ldots, z_p(\tau_0)^2\right)$ and $\widetilde{\boldsymbol{Z}}(\tau_0)^2 = diag\left(\widetilde{z}_1(\tau_0)^2, \ldots, \widetilde{z}_p(\tau_0)^2\right)$, where $z_j(\tau_0)^2 := 1/n \sum_{i=1}^n E\left[\left(\upsilon_i^{(j)}(\tau_0)\right)^2\right]$ and $\widetilde{z}_j(\tau_0)^2 := 1/n \sum_{i=1}^n E\left[\left(\widetilde{\upsilon}_i^{(j)}(\tau_0)\right)^2\right]$ from population nodewise regression. Furthermore, we define the long-run covariance matrices $\boldsymbol{\Omega}_{p,n} = E\left[1/n\left(\sum_{i=1}^n \boldsymbol{w}_i\right)\left(\sum_{i=1}^n \boldsymbol{w}_i'\right)\right]$, $\widetilde{\boldsymbol{\Omega}}_{p,n} = E\left[1/n\left(\sum_{i=1}^n \widetilde{\boldsymbol{w}}_i\right)\left(\sum_{i=1}^n \widetilde{\boldsymbol{w}}_i'\right)\right]$, and $\overline{\boldsymbol{\Omega}}_{p,n} = E\left[1/n\left(\sum_{i=1}^n \boldsymbol{w}_i\right)\left(\sum_{i=1}^n \widetilde{\boldsymbol{w}}_i'\right)\right]$, where $\boldsymbol{w}_i = \left(v_i^{(1)}u_i, \ldots, v_i^{(p)}u_i\right)'$ and $\widetilde{\boldsymbol{w}}_i = \left(\widetilde{v}_i^{(1)}u_i, \ldots, \widetilde{v}_i^{(p)}u_i\right)'$. $\boldsymbol{\Omega}_{p,n}$, $\widetilde{\boldsymbol{\Omega}}_{p,n}$, and $\overline{\boldsymbol{\Omega}}_{p,n}$ can be rewritten as $\boldsymbol{\Omega}_{p,n} = \boldsymbol{\Xi}(0) + \sum_{l=1}^{n-1}\left(\boldsymbol{\Xi}(l) + \boldsymbol{\Xi}'(l)\right)$, $\widetilde{\boldsymbol{\Omega}}_{p,n} = \widetilde{\boldsymbol{\Xi}}(0) + \sum_{l=1}^{n-1}\left(\widetilde{\boldsymbol{\Xi}}(l) + \widetilde{\boldsymbol{\Xi}}'(l)\right)$ and $\overline{\boldsymbol{\Omega}}_{p,n} = \overline{\boldsymbol{\Xi}}(0) + \sum_{l=1}^{n-1}\left(\overline{\boldsymbol{\Xi}}(l) + \overline{\boldsymbol{\Xi}}'(l)\right)$ where $\boldsymbol{\Xi}(l) = \frac{1}{n}\sum_{i=l+1}^n E\left[\boldsymbol{w}_i\boldsymbol{w}_{i-l}'\right]$, $\widetilde{\boldsymbol{\Xi}}(l) = \frac{1}{n}\sum_{i=l+1}^n E\left[\widetilde{\boldsymbol{w}}_i\widetilde{\boldsymbol{w}}_{i-l}'\right]$, and $\overline{\boldsymbol{\Xi}}(l) = \frac{1}{n}\sum_{i=l+1}^n E\left[\boldsymbol{w}_i\widetilde{\boldsymbol{w}}_{i-l}'\right]$, as in Adamek et al. (2023).

Let $\underline{\lambda} = \min_{j=1,\ldots,2p}\lambda_j$ and $\bar{\lambda} = \max_{j=1,\ldots,2p}\lambda_j$, satisfy (8.2) and let $\bar{s}_r = \max\left\{\max_{j=1,\ldots,p} s_r^{(j)}, \max_{j=1,\ldots,p} \widetilde{s}_r^{(j)}\right\}$. Define $\lambda_{\min} = \min\{\lambda, \underline{\lambda}\}$, $\lambda_{\max} = \max\{\lambda, \bar{\lambda}\}$, $s_{r,\max} = \max\{s_r, \bar{s}_r\}$. Similar to Section 3, we define $g$ as a $(2p \times 1)$ vector with $|g|_2 = 1$ and let $H = \{j = 1, \ldots, 2p \mid g_j \neq 0\}$ with cardinality $|H| = h < C$. $H$ contains the indices of the coefficients involved in the hypothesis to be tested. The following theorem establishes the asymptotic normality of the debiased Lasso estimator in both cases.

**Theorem 6.** *Suppose that Assumptions 3, 4 and 8 to 11 hold, that* $s_{r,max}^{3/2} logp/\sqrt{n} \to 0$, *and that the smallest eigenvalues of* $\boldsymbol{\Omega}_{p,n}$, $\widetilde{\boldsymbol{\Omega}}_{p,n}$ *and* $\overline{\boldsymbol{\Omega}}_{p,n}$ *are bounded away from 0. Furthermore, assume that* $\lambda_{\max}^2 \leq (\ln\ln n)\lambda_{\min}^r\left[\sqrt{n}s_{r,\max}\right]^{-1}$, *and*

$$0 < r < 1: \quad \lambda_{\min} \geq (\ln\ln n)\left[s_{r,\max}\left(\frac{p^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}}{\sqrt{n}}\right)^{\frac{1}{\left(\frac{1}{d}+\frac{m}{m-1}\right)}}\right]^{\frac{1}{r}}$$

$$r = 0: \quad s_{0,\max} \leq (\ln\ln n)^{-1}\left[\frac{\sqrt{n}}{p^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}}\right]^{\frac{1}{\left(\frac{1}{d}+\frac{m}{m-1}\right)}}, \quad \lambda_{\min} \geq (\ln\ln n)\frac{p^{1/m}}{\sqrt{n}}.$$

*Then, as* $n \to \infty$, *we have*

$$\frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\boldsymbol{\Psi}\left(\widehat{\tau}\right)g}} \xrightarrow{d} N(0,1),$$

*uniformly in* $\alpha_0 \in \mathcal{B}_{2p}(r, s_r)$, *where*

$$\boldsymbol{\Psi}\left(\tau\right) := \begin{bmatrix} \widetilde{\boldsymbol{Z}}(\tau)^{-2}\widetilde{\boldsymbol{\Omega}}_{p,n}\widetilde{\boldsymbol{Z}}(\tau)^{-2} & \widetilde{\boldsymbol{Z}}(\tau)^{-2}\overline{\boldsymbol{\Omega}}_{p,n}\boldsymbol{Z}(\tau)^{-2} - \widetilde{\boldsymbol{Z}}(\tau)^{-2}\widetilde{\boldsymbol{\Omega}}_{p,n}\widetilde{\boldsymbol{Z}}(\tau)^{-2} \\ \widetilde{\boldsymbol{Z}}(\tau)^{-2}\overline{\boldsymbol{\Omega}}_{p,n}\boldsymbol{Z}(\tau)^{-2} - \widetilde{\boldsymbol{Z}}(\tau)^{-2}\widetilde{\boldsymbol{\Omega}}_{p,n}\widetilde{\boldsymbol{Z}}(\tau)^{-2} & \boldsymbol{Z}(\tau)^{-2}\boldsymbol{\Omega}_{p,n}\boldsymbol{Z}(\tau)^{-2} + \widetilde{\boldsymbol{Z}}(\tau)^{-2}\widetilde{\boldsymbol{\Omega}}_{p,n}\widetilde{\boldsymbol{Z}}(\tau)^{-2} - 2\widetilde{\boldsymbol{Z}}(\tau)^{-2}\overline{\boldsymbol{\Omega}}_{p,n}\boldsymbol{Z}(\tau)^{-2} \end{bmatrix}.$$

To estimate the asymptotic variance, we consider the long-run variance kernel estimator, as in Adamek et al. (2023), $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Xi}}(0) + \sum_{l=1}^{\widehat{k}_n - 1} K\left(\frac{l}{\widehat{k}_n}\right)\left(\widehat{\boldsymbol{\Xi}}(l) + \widehat{\boldsymbol{\Xi}}'(l)\right)$, $\widehat{\widetilde{\boldsymbol{\Omega}}} = \widehat{\widetilde{\boldsymbol{\Xi}}}(0) + \sum_{l=1}^{\widetilde{k}_n - 1} K\left(\frac{l}{\widetilde{k}_n}\right)\left(\widehat{\widetilde{\boldsymbol{\Xi}}}(l) + \widehat{\widetilde{\boldsymbol{\Xi}}}'(l)\right)$, and $\widehat{\overline{\boldsymbol{\Omega}}} = \widehat{\overline{\boldsymbol{\Xi}}}(0) + \sum_{l=1}^{\overline{k}_n - 1} K\left(\frac{l}{\overline{k}_n}\right)\left(\widehat{\overline{\boldsymbol{\Xi}}}(l) + \widehat{\overline{\boldsymbol{\Xi}}}'(l)\right)$, where $\widehat{\boldsymbol{\Xi}}(l) = \frac{1}{n-l}\sum_{i=l+1}^{n} \widehat{\boldsymbol{w}}_i \widehat{\boldsymbol{w}}_{i-l}'$ with $\widehat{w}_i^{(j)} = \widehat{v}_i^{(j)}\widehat{u}_i$, $\widehat{\widetilde{\boldsymbol{\Xi}}}(l) = \frac{1}{n-l}\sum_{i=l+1}^{n} \widehat{\widetilde{\boldsymbol{w}}}_i \widehat{\widetilde{\boldsymbol{w}}}_{i-l}'$ with $\widehat{\widetilde{w}}_i^{(j)} = \widehat{\widetilde{v}}_i^{(j)}\widehat{u}_i$, and $\widehat{\overline{\boldsymbol{\Xi}}}(l) = \frac{1}{n-l}\sum_{i=l+1}^{n} \widehat{\boldsymbol{w}}_i \widehat{\widetilde{\boldsymbol{w}}}_{i-l}'$, the kernel $K(\cdot)$ can be taken as the Bartlett kernel $K(l/k_n) = \left(1 - \frac{l}{k_n}\right)$ (Newey and West (1987)) and the bandwidths $k_n$, $\widetilde{k}_n$ and $\overline{k}_n$ should increase with the sample size at an appropriate rate. Define $k_n = \max\left\{\widehat{k}_n, \widetilde{k}_n, \overline{k}_n\right\}$.

**Theorem 7.** *Take* $\widehat{\boldsymbol{\Omega}}$, $\widehat{\widetilde{\boldsymbol{\Omega}}}$ *and* $\widehat{\overline{\boldsymbol{\Omega}}}$ *with* $k_n \to \infty$ *as* $n \to \infty$, *such that* $k_n h^2 (\sqrt{n} h^2)^{-\frac{1}{1/d + m/(m-2)}} \to 0$. *Suppose that*

$$\lambda_{\max}^{2-r} \leq (\ln\ln n)^{-1} \min\left\{\left[\sqrt{k_n}\sqrt{n} s_{r,\max}\right]^{-1}, \left[k_n h^{1/m} n^{1/m} s_{r,\max}\right]^{-1}, \left[k_n^2 h^{3/m} n^{(3-m)/m} s_{r,\max}\right]^{-1}, \left[k_n^{2/3} h^{1/(3m)} n^{(m+1)/3m} s_{r,\max}\right]^{-1}\right\},$$

$$\lambda_{\max}^{2} \leq (\ln\ln n)^{-1} \lambda_{\min}^{r} \left[\sqrt{n} h^{2/m} s_{r,\max}\right]^{-1}, \ \textit{and}$$

$$0 < r < 1: \quad \lambda_{\min} \geq (\ln\ln n)\left[s_{r,\max}\left(\frac{(hp)^{\left(\frac{2}{d} + \frac{2}{m-1}\right)}}{\sqrt{n}}\right)^{\frac{1}{\left(\frac{1}{d} + \frac{m}{m-1}\right)}}\right]^{\frac{1}{r}},$$

$$r = 0: \quad s_{0,\max} \leq (\ln\ln n)^{-1}\left[\frac{\sqrt{n}}{(hp)^{\left(\frac{2}{d} + \frac{2}{m-1}\right)}}\right]^{\frac{1}{\left(\frac{1}{d} + \frac{m}{m-1}\right)}}, \quad \lambda_{\min} \geq (\ln\ln n)\frac{(hp)^{1/m}}{\sqrt{n}},$$

*and that Assumptions 3, 4 and 8 to 11 hold, then, we have*

$$\sup_{\alpha^0 \in \mathcal{A}_{2p}^{(1)}(r, s_r)} \left| g'\widehat{\boldsymbol{\Psi}}\left(\widehat{\tau}\right) g - g'\boldsymbol{\Psi}\left(\widehat{\tau}\right) g\right|_1 = o_p(1), \tag{4.2}$$

$$\sup_{\alpha^0 \in \mathcal{A}_{2p}^{(2)}(r,s_r)} \left| g'\widehat{\boldsymbol{\Psi}}\left(\widehat{\tau}\right)g - g'\boldsymbol{\Psi}\left(\tau_0\right)g \right|_1 = o_p(1), \tag{4.3}$$

*where* $\widehat{\boldsymbol{\Psi}}\left(\widehat{\tau}\right) = \begin{bmatrix} \widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2}\widehat{\widetilde{\boldsymbol{\Omega}}}\widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2} & \widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\overline{\boldsymbol{\Omega}}}\widehat{\boldsymbol{Z}}(\tau)^{-2} - \widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2}\widehat{\widetilde{\boldsymbol{\Omega}}}\widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2} \\ \widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\overline{\boldsymbol{\Omega}}}\widehat{\boldsymbol{Z}}(\tau)^{-2} - \widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2}\widehat{\widetilde{\boldsymbol{\Omega}}}\widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2} & \widehat{\boldsymbol{Z}}(\widehat{\tau})^{-2}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{Z}}(\widehat{\tau})^{-2} + \widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2}\widehat{\widetilde{\boldsymbol{\Omega}}}\widehat{\widetilde{\boldsymbol{Z}}}(\widehat{\tau})^{-2} - 2\widehat{\widetilde{\boldsymbol{Z}}}(\tau)^{-2}\widehat{\overline{\boldsymbol{\Omega}}}\widehat{\boldsymbol{Z}}(\tau)^{-2} \end{bmatrix}$.

We provide a uniformly consistent covariance matrix estimator for both cases in Theorem 7. Similar to (3.15) and (3.16), there is a slight difference between the limits of the two asymptotic variances, since there is a true value for the threshold parameter in the case of a fixed threshold effect.

**Theorem 8.** *Suppose that Assumptions 3, 4 and 8 to 11 hold, that* $s_{r,max}^{3/2} logp/\sqrt{n} \to 0$*, that the smallest eigenvalues of* $\boldsymbol{\Omega}_{p,n}$*,* $\overline{\boldsymbol{\Omega}}_{p,n}$ *and that* $\widetilde{\boldsymbol{\Omega}}_{p,n}$ *are bounded away from 0, and* $k_n n^{-\frac{1}{2/d+2m/(m-2)}} \to 0$ *for some* $k_n \to \infty$*. Further, assume that* $\lambda \sim \lambda_{\max} \sim \lambda_{\min}$*, and that*

$$0<r<1: \quad (\ln\ln n)^{-1} s_{r,\max}^{1/r}\left[\frac{p^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}}{\sqrt{n}}\right]^{\frac{1}{r\left(\frac{1}{d}+\frac{m}{m-1}\right)}} \leq \lambda \leq \ \ln\ln n\left[k_n^2\sqrt{n}s_{r,\max}\right]^{-1/(2-r)},$$

$$r=0: \quad (\ln\ln n)^{-1}\frac{p^{1/m}}{\sqrt{n}} \leq \lambda \leq \ln\ln n\left[k_n^2\sqrt{n}s_{0,\max}\right]^{-1/2}.$$

*Assume that* $k_n^r s_{r,\max} p^{(2-r)\left(\frac{d+m-1}{dm+m-1}\right)} n^{\frac{1}{4}\left(r-\frac{d(m-1)(2-r)}{dm+m-1}\right)} \to 0$*, and that* $k_n^2 s_{0,\max}\frac{p^{2/m}}{\sqrt{n}} \to 0$ *if* $r=0$*. Then, we have*

$$\sup_{t\in\mathbb{R}} \sup_{\alpha_0\in\mathcal{B}_{2p}(r,s_r)} \left| P\left(\frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau})-\alpha_0)}{\sqrt{g'\widehat{\boldsymbol{\Psi}}\left(\widehat{\tau}\right)g}} \leq t\right) - \Phi(t)\right| \to 0.$$

Similar to the result in Theorem 4, we show that the convergence of a linear combination of the parameters of the debiased estimator $\widehat{a}(\widehat{\tau})$ to the standard normal distribution is uniformly valid over the $\ell_r$-ball. This allows researchers to perform uniform inference without specifying whether the specification is a linear or threshold regression.

### 4.2 Local Projection Inference

In this section, we develop the uniform inference theory for the debiased impulse response parameters in the high-dimensional local projection (HDLP) threshold model. We focus on the following local projection threshold regression:

$$Y_{i+h} = \begin{cases} \beta_{h,0} + \phi_h x_i + \rho_h Y_i + \boldsymbol{\eta}_h' \boldsymbol{w}_{s,i} + \sum_{k=1}^{K} \boldsymbol{\Delta}_{h,k}' \boldsymbol{z}_{t-i} + U_{h,i}, \quad \text{if } Q_i \geq \tau_0, \\ (\beta_{h,0} + \delta_{h,0}) + (\phi_h + \delta_{h,x,0}) x_i + (\rho_h + \delta_{h,y,0}) Y_i + (\boldsymbol{\eta}_h' + \delta_{h,\boldsymbol{\eta},0}') \boldsymbol{w}_{s,i} + \sum_{k=1}^{K} (\boldsymbol{\Delta}_{h,k}' + \delta_{h,k,\boldsymbol{\Delta},0}') \boldsymbol{z}_{t-i} + U_{h,i}, \\ \qquad \text{if } Q_i < \tau_0. \end{cases} \tag{4.4}$$

where $h = 0, 1, \ldots, h_{\max}$, [14] $\beta_h = (\beta_{h,0}, \phi_h, \rho_h, \boldsymbol{\eta}_h', \boldsymbol{\Delta}_{h,k}')'$ represents the projection parameters when the threshold variable is above the threshold point $\tau_0$, while $\beta_{h,0} + \delta_{h,0}, \phi_h + \delta_{h,x,0}, \rho_h + \delta_{h,y,0}, \boldsymbol{\eta}_h + \delta_{h,\boldsymbol{\eta},0}, \boldsymbol{\Delta}_{h,k} + \delta_{h,k,\boldsymbol{\Delta},0}$ are the projection parameters when the threshold variable is below the threshold point. $U_{h,i}$ is the projection error and $\boldsymbol{z}_i = \left(\boldsymbol{w}_{s,i}', Y_i, x_i, \boldsymbol{w}_{f,i}'\right)'$ includes the response $Y_i$, the shock variable $x_i$, and the vectors of control variables consisting of "slow" variables $\boldsymbol{w}_{s,i} \in \mathbb{R}^{n_s}$, and the "fast" variables $\boldsymbol{w}_{f,i} \in \mathbb{R}^{n_f}$ for identification purposes.[15] We are interested in $\phi_h$ and $\phi_h + \delta_{h,x,0}$, either of which represents the response at horizon $h$ of $y_i$ after an impulse in $x_i$. When $\delta_0 = (\delta_{h,x,0}, \delta_{h,y,0}, \delta_{h,\boldsymbol{\eta},0}', \delta_{h,k,\boldsymbol{\Delta},0}')' = 0$, it reduces to the local projection regression, similar to equation (1) in Adamek et al. (2024). We focus on a small number of parameters, allowing us to rewrite equation (4.4) as

$$Y_i = \underset{1\times 2H}{\boldsymbol{X}_{\mathcal{H},i}'(\tau_0)} \underset{2H\times 1}{\alpha_{\mathcal{H},0}} + \underset{1\times(2p-2H)}{\boldsymbol{X}_{-\mathcal{H},i}'(\tau_0)} \underset{(2p-2H)\times 1}{\alpha_{-\mathcal{H},0}} + U_i, \quad i = 1, \ldots, n, \tag{4.5}$$

and furthermore,

$$\boldsymbol{Y} = \boldsymbol{X}_{\mathcal{H}}(\tau_0)\alpha_{\mathcal{H},0} + \boldsymbol{X}_{-\mathcal{H}}(\tau_0)\alpha_{-\mathcal{H},0} + \boldsymbol{U}.$$

We now have two groups of parameters. The parameters of interest are $\alpha_{\mathcal{H},0} = (\beta_{H,0}, \delta_{H,0})$, which belong to the first group while the second group includes the parameters for control variables, where $\mathcal{H}$ is the index set representing the $2H$ variables of in-

[14] We assume that the unknown threshold parameter ($\tau_0$) is the same when estimating the impulse response function across different horizons and that the number of horizons is finite.

[15] See Adamek et al. (2024) for more detailed explanations.

terest. Without loss of generality, we order the variables in $\boldsymbol{X}(\tau) = (\boldsymbol{X}_{\mathcal{H}}(\tau), \boldsymbol{X}_{-\mathcal{H}}(\tau))$. We then apply the penalization method, as in Adamek et al. (2024), penalizing only the parameters for control variables $\alpha_{-\mathcal{H},0}$. Thus, shrinkage bias does not affect the unpenalized parameters of interest. We suppress the dependence on $h$ since each horizon of the LPs is estimated separately. The Lasso estimator is given as follows:

$$(\widehat{\alpha}(\widehat{\tau}), \widehat{\tau}) = \left( \left( \widehat{\alpha}'_{\mathcal{H}}(\widehat{\tau}), \widehat{\alpha}'_{-\mathcal{H}}(\widehat{\tau}) \right)', \widehat{\tau} \right) = \underset{\alpha \in \mathbb{R}^{2p}, \tau \in \mathbb{T}}{\arg\min} \, ||\boldsymbol{Y} - \boldsymbol{X}(\tau)\alpha||_n^2 + 2\lambda |\boldsymbol{D}(\tau)\alpha|_1, \quad (4.6)$$

where $\boldsymbol{D}(\tau)$ is an $2p \times 2p$ diagonal matrix with $\boldsymbol{D}_{i,i}(\tau) = 0$ for $i \in \mathcal{H}$ and $\boldsymbol{D}_{i,i}(\tau)$ can be defined as in (2.5) otherwise. [16] The oracle inequalities are qualitatively the same as those in Theorem 5.

Next, we consider the debiased Lasso estimator

$$\widehat{a}_{\mathcal{H}}(\widehat{\tau}) \; = \widehat{\alpha}_{\mathcal{H}}(\widehat{\tau}) + \widehat{\boldsymbol{\Theta}}(\widehat{\tau}) \boldsymbol{X}(\widehat{\tau})'(\boldsymbol{Y} - \boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}))/n, \quad (4.7)$$

where $\widehat{\boldsymbol{\Theta}}(\widehat{\tau})$ is an $2H \times 2p$ submatrix of an approximate inverse of $\widehat{\boldsymbol{\Sigma}}(\widehat{\tau}) := \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})/n$. We still use nodewise regression and follow the same process as in Section 4.1 to obtain $\widehat{\boldsymbol{\Theta}}(\widehat{\tau})$. We construct

$$\widehat{\boldsymbol{C}}_H(\tau) := \begin{pmatrix} 1 & -\widehat{\gamma}_1^{(2)}(\tau) & \dots & -\widehat{\gamma}_1^{(H)}(\tau) & \dots & -\widehat{\gamma}_1^{(p)}(\tau) \\ -\widehat{\gamma}_2^{(1)}(\tau) & 1 & \dots & -\widehat{\gamma}_2^{(H)}(\tau) & \dots & -\widehat{\gamma}_2^{(p)}(\tau) \\ \vdots & \vdots & \ddots & \vdots & & \\ -\widehat{\gamma}_H^{(1)}(\tau) & -\widehat{\gamma}_H^{(2)}(\tau) & \dots & 1 & \dots & -\widehat{\gamma}_H^{(p)}(\tau) \end{pmatrix},$$

and $\widehat{\boldsymbol{Z}}_H(\tau)^2 := \mathrm{diag}\,(\widehat{z}_1(\tau)^2, \dots, \widehat{z}_H(\tau)^2)$, where $\widehat{z}_j(\tau)^2 := ||X^{(j)}(\tau) - X^{(-j)}(\tau)\widehat{\gamma}_j(\tau)||_n^2 + 2\lambda_j |\widehat{\gamma}_j(\tau)|_1$, we thus obtain $\widehat{A}_H(\tau) = \widehat{\boldsymbol{Z}}_H(\tau)^{-2}\widehat{\boldsymbol{C}}_H(\tau)$. Similarly, we have $\widehat{B}_H(\tau) = \widehat{\widehat{\boldsymbol{Z}}}_H(\tau)^{-2}\widehat{\widehat{\boldsymbol{C}}}_H(\tau)$. Define $\boldsymbol{\Omega}_{H,p,n}$, $\overline{\boldsymbol{\Omega}}_{H,p,n}$, and $\widetilde{\boldsymbol{\Omega}}_{H,p,n}$ as the top-left $H \times H$ submatrices of $\boldsymbol{\Omega}_{p,n}$, $\overline{\boldsymbol{\Omega}}_{p,n}$ and $\widetilde{\boldsymbol{\Omega}}_{p,n}$, respectively.

[16] As in Section 2.1 of Adamek et al. (2024), the Lasso estimator can be derived by

$$(\widehat{\alpha}_{-\mathcal{H}}(\widehat{\tau}), \widehat{\tau}) = \underset{\alpha \in \mathbb{R}^{(2p-2H)}, \tau \in \mathbb{T}}{\arg\min} \left|\left| \boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{H}}(\tau))}\boldsymbol{Y} - \boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{H}}(\tau))}\boldsymbol{X}_{-\mathcal{H}}(\tau)\alpha \right|\right|_n^2 + 2\lambda|\alpha|_1,$$
$$\widehat{\alpha}_{\mathcal{H}}(\widehat{\tau}) = \widehat{\boldsymbol{\Sigma}}_{\mathcal{H}}^{-1}\boldsymbol{X}_{\mathcal{H}}(\widehat{\tau})' \left( \boldsymbol{Y} - \boldsymbol{X}_{-\mathcal{H}}(\widehat{\tau})\widehat{\alpha}_{-\mathcal{H}} \right)/n,$$

where $\boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{H}}(\tau))} := I - \boldsymbol{X}_{\mathcal{H}}(\tau)(\boldsymbol{X}_{\mathcal{H}}(\tau)'\boldsymbol{X}_{\mathcal{H}}(\tau))^{-1}\boldsymbol{X}_{\mathcal{H}}(\tau)'$, and $\widehat{\boldsymbol{\Sigma}}_{\mathcal{H}} := \boldsymbol{X}_{\mathcal{H}}(\tau)'\boldsymbol{X}_{\mathcal{H}}(\tau)/n$.

**Theorem 9.** *Suppose that Assumptions 3, 4 and 8 to 11 hold, that* $H \leq C$, *that* $s_{r,max}^{3/2} logp/\sqrt{n} \to 0$, *that the smallest eigenvalues of* $\mathbf{\Omega}_{H,p,n}$, $\overline{\mathbf{\Omega}}_{H,p,n}$, *and that* $\widetilde{\mathbf{\Omega}}_{H,p,n}$ *are bounded away from 0, and* $k_n n^{-\frac{1}{2/d+2m/(m-2)}} \to 0$ *for some* $k_n \to \infty$. *Further, assume that* $\lambda \sim \lambda_{\max} \sim \lambda_{\min}$, *and that*

$$0 < r < 1: \quad (\ln\ln n)^{-1} s_{r,\max}^{1/r} \left[\frac{p^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}}{\sqrt{n}}\right]^{\frac{1}{r\left(\frac{1}{d}+\frac{m}{m-1}\right)}} \leq \lambda \leq \ \ln\ln n \left[k_n^2\sqrt{n}s_{r,\max}\right]^{-1/(2-r)},$$

$$r = 0: \quad (\ln\ln n)^{-1}\frac{p^{1/m}}{\sqrt{n}} \leq \lambda \leq \ln\ln n\left[k_n^2\sqrt{n}s_{0,\max}\right]^{-1/2}.$$

*Assume that* $k_n^r s_{r,\max} p^{(2-r)\left(\frac{d+m-1}{dm+m-1}\right)} n^{\frac{1}{4}\left(r-\frac{d(m-1)(2-r)}{dm+m-1}\right)} \to 0$, *and that* $k_n^2 s_{0,\max}\frac{p^{2/m}}{\sqrt{n}} \to 0$ *if* $r = 0$. *Then, for* $g \in \mathbb{R}^{\mathcal{H}}$, *we have*

$$\sup_{t\in\mathbb{R}} \sup_{\alpha_0 \in \boldsymbol{B}_{2p}(r,s_r)} \left| P\left(\frac{\sqrt{n}g'(\widehat{a}_{\mathcal{H}}(\widehat{\tau}) - \alpha_{\mathcal{H},0})}{\sqrt{g'\widehat{\mathbf{\Psi}}_{\mathcal{H}}(\widehat{\tau})g}} \leq t\right) - \Phi(t)\right| = o_p(1),$$

*where*

$$\widehat{\mathbf{\Psi}}_{\mathcal{H}}(\widehat{\tau}) = \begin{bmatrix} \widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2}\widehat{\widetilde{\mathbf{\Omega}}}_{\mathcal{H}}\widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2} & \widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\tau)^{-2}\widehat{\overline{\mathbf{\Omega}}}_{\mathcal{H}}\widehat{\boldsymbol{Z}}_{\mathcal{H}}(\tau)^{-2} - \widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2}\widehat{\widetilde{\mathbf{\Omega}}}_{\mathcal{H}}\widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2} \\ \widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\tau)^{-2}\widehat{\overline{\mathbf{\Omega}}}_{\mathcal{H}}\widehat{\boldsymbol{Z}}_{\mathcal{H}}(\tau)^{-2} - \widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2}\widehat{\widetilde{\mathbf{\Omega}}}_{\mathcal{H}}\widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2} & \widehat{\boldsymbol{Z}}_{\mathcal{H}}(\widehat{\tau})^{-2}\widehat{\mathbf{\Omega}}_{\mathcal{H}}\widehat{\boldsymbol{Z}}_{\mathcal{H}}(\widehat{\tau})^{-2} + \widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2}\widehat{\widetilde{\mathbf{\Omega}}}_{\mathcal{H}}\widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\widehat{\tau})^{-2} - 2\widehat{\widetilde{\boldsymbol{Z}}}_{\mathcal{H}}(\tau)^{-2}\widehat{\overline{\mathbf{\Omega}}}_{\mathcal{H}}\widehat{\boldsymbol{Z}}_{\mathcal{H}}(\tau)^{-2} \end{bmatrix}.$$

As in Adamek et al. (2023) and Adamek et al. (2024), we also use an autocorrelation robust Newey-West long-run covariance estimator, as in Section 4.1.

The uniformly consistent covariance results for the cases with no threshold effect and a fixed threshold effect are similar to those in Theorem 7. The proof of Theorem 9 follows from the proofs of Theorem 8, and Theorem 1 in Adamek et al. (2024). Therefore, we omit the detailed proof.

# 5 Monte Carlo Simulation and Applications

We study the finite sample properties of the proposed desparsified Lasso estimator for high-dimensional threshold regression through Monte Carlo experiments. We compare our debiased Lasso estimators for threshold regression with those from linear regression in cases with no threshold effect and a fixed threshold effect. Additionally, We apply our method to two empirical applications, one related to the multiple steady states of economic growth by Durlauf and Johnson (1995), and the other concerning the effect of a military spending news shock on government spending and GDP by Ramey and Zubairy (2018).

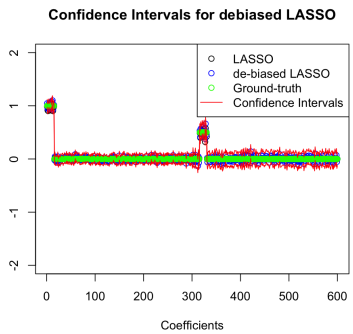
## 5.1 Monte Carlo Simulation

The implementation of the debiased Lasso method for the linear model is publicly available at `https://web.stanford.edu/~montanar/sslasso/code.html`. To choose the tuning parameter $\lambda$, we utilize the regularization parameter $\lambda = 4\widehat{\sigma}\sqrt{(2logp)/n}$ from equation (31) in Javanmard and Montanari (2014) and ten-fold cross-validation. However, according to our simulation results, cross-validation does not significantly enhance the performance, while the processing time is considerably longer. Therefore, we use regularization parameters for both $\lambda$ and $\lambda_{node}$.

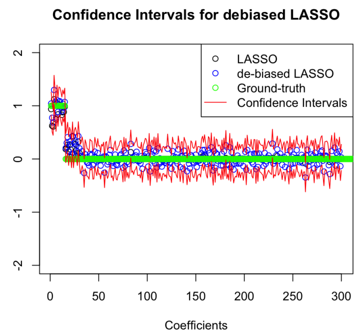
Before discussing the results, we briefly describe the data-generating process. We consider the threshold regression model (1.1), where the rows of the design matrix are i.i.d realizations from $N(0, \mathbf{\Sigma})$, with $\Sigma_{j,k} = 0.5^{|j-k|}$, a Toeplitz structure, and error terms are $U_i \sim N(0, 0.25)$. When the threshold variable $Q_i$ is independent of $X_i$, we take $Q_i \sim$ uniform$(0, 1)$. [17] We also consider the case where the threshold variable correlates with the covariates. We take $\tau_0 = 0.5$ unless otherwise specified. We use the grid search method to find $\tau$ from 0.15 to 0.85 by steps of 0.05. Without loss of generality, we assume that $\beta_0$ is a $p \times 1$ vector with the first $s_0$ elements being $b$, the remaining $p - s_0$ elements being zeros and that $\delta_0$ is a $p \times 1$ vector with the first $s_0$ elements being 0, the next $s_0$ elements being $b1$, and the remaining elements being zeros.

In Figures 1 and 2, we plot the constructed 95% confidence intervals for the realizations $(n, 2p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (400, 600, 15, 1, 0.5, 0.5)$ and $(n, 2p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (400, 600, 15, 1, 0, 0.5)$ for both threshold and linear regression models. Our debiased estimator for threshold regression performs much better than the debiased estimator for linear regression when there is a fixed threshold effect. Even when the threshold effect does not exist, our estimator for the threshold regression still performs comparably to that for the linear regression.

[17] We also consider the case where the rows of the design matrix are i.i.d. realizations from a binomial distribution with a success probability of 0.15. The results are similar, so we do not report them.

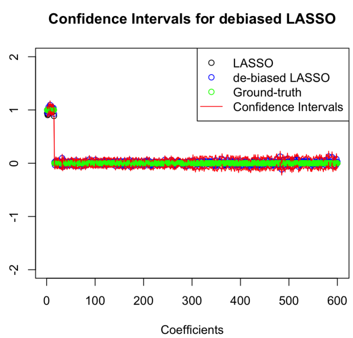



(a) Our estimator

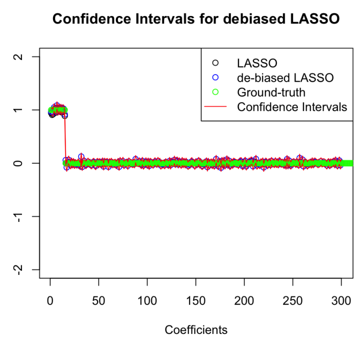



(b) J&M estimator

Figure 1: 95% confidence intervals for one realization $(n, 2p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (200, 600, 15, 1, 0.5, 0.5)$ (with a fixed threshold effect).




(a) Our estimator




(b) J&M estimator

Figure 2: 95% confidence intervals for one realization $(n, 2p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (200, 600, 15, 1, 0, 0.5)$ (without threshold effect).

Additionally, we consider 20 independent realizations for each parameter $\alpha_{0,i}$ for each model specification. We focus only on the parameters $\beta_{0,i}$ as we study two separate cases: one with a fixed threshold effect and the other without a threshold effect ($\delta_{0,i} = 0$). We compute the average length of the corresponding confidence interval $\mathrm{Avglength}\,(\mathrm{J_i}(\beta))$, and the average

$$\ell \equiv p^{-1} \sum_{i \in [p]} \mathrm{Avglength}(\mathrm{J_i}(\beta)). \tag{5.1}$$

| Configuration \ Measure | $\lvert\widehat{\tau}-\tau_0\rvert$ | $\ell$ | $\ell_S$ | $\ell_{S^c}$ | $\widehat{Cov}$ | $\widehat{Cov}_S$ | $\widehat{Cov}_{S^c}$ |
|---|---|---|---|---|---|---|---|
| $(400, 600, 15, 1, 0.5, 0)$ | 0.007 | 0.1717 | 0.1719 | 0.1717 | 0.9958 | 0.9767 | 0.9968 |
| $(400, 600, 15, 1, 0, 0)$ | - | 0.1576 | 0.1571 | 0.1576 | 0.9907 | 0.94 | 0.9933 |
| $(400, 600, 45, 1, 0.5, 0)$ | 0.017 | 0.2171 | 0.2171 | 0.2170 | 0.9935 | 0.9667 | 0.9982 |
| $(400, 600, 45, 1, 0, 0)$ | - | 0.1813 | 0.1811 | 0.1814 | 0.9913 | 0.9544 | 0.9978 |
| $(400, 600, 15, 0.5, 0.25, 0)$ | 0.019 | 0.1696 | 0.1689 | 0.1697 | 0.9963 | 0.9867 | 0.9968 |
| $(400, 600, 15, 0.5, 0, 0)$ | - | 0.1593 | 0.1588 | 0.1594 | 0.9907 | 0.96 | 0.9923 |
| $(400, 600, 15, 1, 0.5, 0.5)$ | 0.018 | 0.1722 | 0.1720 | 0.1722 | 0.9953 | 0.9767 | 0.9963 |
| $(400, 600, 15, 1, 0, 0.5)$ | - | 0.1621 | 0.1625 | 0.1621 | 0.9908 | 0.95 | 0.9930 |
| $(400, 600, 15, 1, 0.5, 0)\ \tau_0 = 0.4$ | 0.012 | 0.1668 | 0.1657 | 0.1669 | 0.9933 | 0.9867 | 0.9937 |
| $(400, 600, 15, 1, 0, 0)$ | - | 0.1490 | 0.1484 | 0.1490 | 0.9852 | 0.94 | 0.9875 |

Table 1: Simulation results for absolute threshold parameter estimation error, average length of confidence intervals, and average coverage.

We also compute the average length of intervals for the active and inactive parameters,

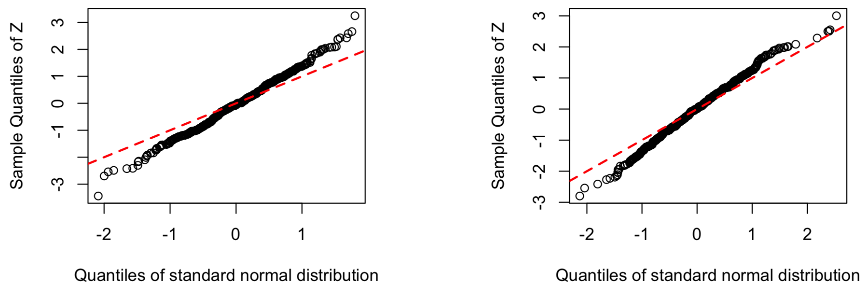
$$\ell_S \equiv s_0^{-1} \sum_{i\in S} \text{Avglength}(\mathrm{J_i}(\beta))\,, \quad \ell_{S^c} \equiv (\mathrm{p}-\mathrm{s_0})^{-1} \sum_{\mathrm{i}\in \mathrm{S^c}} \text{Avglength}(\mathrm{J_i}(\beta)), \tag{5.2}$$

and the average coverage for individual parameters,

$$\begin{aligned} &\widehat{Cov} \equiv p^{-1} \sum_{i\in[p]} \widehat{\mathbb{P}}[\beta_{0,i} \in J_i(\beta)], \quad \widehat{Cov}_S \equiv s_0^{-1} \sum_{i\in S} \widehat{\mathbb{P}}[\beta_{0,i} \in J_i(\beta)], \\ &\widehat{Cov}_{S^c} \equiv (p-s_0)^{-1} \sum_{i\in S^c} \widehat{\mathbb{P}}[0 \in J_i(\beta)], \end{aligned} \tag{5.3}$$

where $\widehat{\mathbb{P}}$ denotes the empirical probability computed based on the 20 realizations for each configuration. The results are reported in Table 1. The debiased estimator performs better for the active parameters. We show the robustness of the debiased estimator concerning the presence of a fixed threshold effect, the correlation between the threshold variable and the covariates, and different magnitudes of effects and varying levels of sparsity.

Figures 3a and 3b show the sample quantiles of $Z$, where $Z = (z_i)_{i=1}^{2p}$ with $z_i = \sqrt{n}\left(\widehat{a}^{(i)}(\widehat{\tau}) - \alpha_0^{(i)}\right) / \sqrt{\left[\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'\right]_{i,i}}$, versus the quantiles of the standard normal distribution for realizations of the configuration $(n, p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (400, 600, 10, 1, 0.5, 0)$ and $(n, p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (400, 600, 10, 1, 0, 0)$. The scattered points are close to the line with a slope of one and an intercept of zero, which is consistent with the result of Theorem 4 regarding the standard normality of $z_i$.




(a) Q-Q plot for one realization $(n, 2p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (400, 600, 15, 1, 0.5, 0)$ (with a fixed threshold effect).

(b) Q-Q plot for one realization $(n, 2p, s_0, b, b1, \rho_{Q,X^{(2)}}) = (400, 600, 15, 1, 0, 0)$ (without threshold effect).

Additionally, we consider a test for the family of hypotheses $\left\{ H_0^{(j)} : \alpha_0^{(j)} = 0 \right\}$ for $j = 1, \ldots, 2p$. We report the familywise error rate (FWER) based on the Bonferroni procedure and the empirical power,

$$\text{Power} = s_0^{-1} \sum_{i \in S} P \left( H_{0,i} \text{ is rejected} \right).$$

We compare our results from threshold models with those from linear regression in Table 2. The FWER based on the threshold model is close to the preassigned significance level of 0.05 and is robust to the magnitude of the threshold effect and the level of sparsity. Our debiased Lasso estimator for the threshold model has more power than that for the linear model, even when the threshold effect is small.

## 5.2 Economic Growth Rate

Durlauf and Johnson (1995) provide a theoretical background for the existence of multiple steady states in economic growth models. They also consider a broad set of control variables to check the robustness, but Lee et al. (2016) argue that this approach still restricts variable selection. Therefore, they apply the Lasso method to simultaneously select covariates and choose between linear and threshold models when

| | Threshold Model | | Linear Model | |
|---|---|---|---|---|
| Configuration | FWER | Power | FWER | Power |
| $(400, 600, 15, 0.5, 0.25, 0)$ | 0.049 | 0.973 | 0.053 | 0.535 |
| $(400, 600, 30, 0.5, 0.25, 0)$ | 0.094 | 0.941 | 0.097 | 0.483 |
| $(400, 600, 15, 0.5, 0.1, 0)$ | 0.032 | 0.59 | 0.051 | 0.503 |
| $(1000, 1200, 15, 0.5, 0.25, 0)$ | 0.050 | 1 | 0.061 | 0.615 |
| $(1000, 1200, 30, 0.5, 0.25, 0)$ | 0.101 | 1 | 0.107 | 0.533 |
| $(1000, 1200, 15, 0.5, 0.1, 0)$ | 0.069 | 0.747 | 0.053 | 0.521 |

Table 2: Simulation results for FWER and Power from threshold models and linear models

dealing with high-dimensional data. To further identify the relevant covariates, we continue applying a threshold regression model to study countries' economic growth and analyze the significance of covariates by the debiased Lasso estimator. Our setup follows Equation (3.1) in Lee et al. (2016),

$$gr_i = \beta_0 + \beta_1 lgdp60_i + X_i'\beta_2 + \mathbf{1}\{Q_i < \tau\}\left(\delta_0 + \delta_1 lgdp60_i + X_i'\delta_2\right) + U_i, \tag{5.4}$$

where $gr_i$ is the annualized GDP growth rate for each country $i$ during the period 1960-1985, $lgdp60_i$ represents the log GDP in 1960, and $X_i$ is a vector of additional covariates, including education, demographic characteristics, market openness, politics, and interaction terms. [18] We use either the initial GDP or the adult literacy rate in 1960 as the threshold variable $Q_i$ following Durlauf and Johnson (1995) and Lee et al. (2016). The grid interval ranges from the 10th to the 90th percentiles of the threshold variable. We utilize covariates from Lee et al. (2016) and the dataset originating from Barro and Lee (1994) and Durlauf and Johnson (1995). With initial GDP as the threshold variable, we have 79 countries with 46 covariates. With literacy rate as the threshold variable, we have 67 countries with 47 covariates. The traditional OLS and MLE are no longer valid because of $2p > n$.

Tables 3 and 4 summarize the results with initial gdp and adult literacy rate as the threshold variables. We reconfirm the existence of the threshold point, as some values of $\delta_2$ are significantly different from 0. These results thus verify that multiple steady states exist in growth models. The significance of covariates varies depending

[18] Table 1 in Lee et al. (2016) provides a detailed introduction to the covariates.

on the regime. Firstly, when initial gdp is the threshold variable, except for some common variables that affect both developing and developed countries, a positive trade shock and a higher percentage of secondary school completion will accelerate a developing country's economic growth, whereas participating in an external war will decrease a developing country's economic growth. Secondly, when the adult literacy rate is the threshold variable, in addition to some common variables that affect both well-educated and less-educated countries, a higher percentage of secondary school completion will accelerate a less-educated country's economic growth. Additionally, the variables selected and the significance of covariates in the threshold model differ from those in linear regressions. Finally, our selected relevant covariates differ from those in Lee et al. (2016), but our results are more convincing as we report the standard error for each significant covariate.

## 5.3 Government Spending and GDP

Ramey and Zubairy (2018) provide theoretical and empirical background on the effect of a military news shock on government spending and GDP and use state-dependence LP to estimate impulse responses. Later, Adamek et al. (2024) reestimate the impulse responses through a state-dependence HDLP specification while including more lags for a robustness check. However, the threshold point defining the state of the economy by Ramey and Zubairy (2018) and Adamek et al. (2024) is based on a predetermined standard. For example, defining the state as slack when the unemployment rate exceeds 6.5 percent (the US Federal Reserve's standard). A state is defined as a zero lower bound (ZLB) when the 3-month Treasury bill rates are below 0.5 percent. We thus apply a high-dimensional local projection threshold model to find a theoretically more accurate threshold point to define the state and estimate the impulse responses accordingly. The model is as follows,

$$Y_{i+h} = \beta_{h,0}+\beta_{h,1}x_i+\sum_{k=1}^{K}\boldsymbol{z}'_{i-k}\beta_{h,2,k}+\mathbf{1}\{Q_{i-1} < \tau\}\left(\delta_{h,0}+\delta_{h,1}x_i+\sum_{k=1}^{K}\boldsymbol{z}'_{i-k}\delta_{h,2,k}\right)+U_{h,i}, \tag{5.5}$$

where $Y_i$ includes real per capita GDP and government spending, $x_i$ is the military spending news shock, $\boldsymbol{z}_i$ includes lags of the news, GDP, government spending, and tax. We use a quarterly dataset from 1889Q1 to 2015Q4 which is available at `https:`

//econweb.ucsd.edu/~vramey/research.html#govt. [19] The time series length is 161, spanning a long U.S. history. It includes many prolonged periods of slack and extended periods of near-zero bound, allowing us to estimate the impulse responses over reasonable horizons with non-changing states. We use one lag of the 3-month Treasury bill rate or one lag of the unemployment rate as the threshold variable. [20]

When one lag of the 3-month Treasury bill rate is the threshold variable, we take $K = 20$, and search for the optimal point in the interval $[0, 4]$ (0 and 4.23 are the lowest value and the 70th percentile, respectively, of the threshold variable) by steps of 0.05. We try horizons from $h = 1$ to $h = 5$ for a robustness check and derive that the potential threshold points are 0.45, 0.5, and 0.55. There are only two periods with a 3-month Treasury bill rate in the range of $[0.45, 0.55]$, so any of the three points can be the optimal threshold point. We then define the state as ZLB when the rate is below 0.45 (the estimate when $h = 1$) and the normal state when it is above 0.45. Figure 4 shows the impulse responses to a military spending news shock in government spending and GDP. Compared to Figure 11 in Ramey and Zubairy (2018), for government spending, the peak of the responses occurs at the same horizon, and the peak magnitude is slightly larger in the ZLB state. In the normal state, the peak occurs at the same horizon, and the magnitude is slightly smaller. However, for GDP, the peak occurs 2 quarters earlier, and the peak magnitude is larger in the ZLB while the response is insignificant in the normal state.

When one lag of the unemployment rate is the threshold variable, we take $K = 20$, and search for the optimal point in the interval $[3, 10]$ (3.36 and 10.57 are the 10th and 90th percentiles, respectively, of the threshold variable) by steps of 0.1. We also try horizons from $h = 1$ to $h = 5$ for a robustness check and obtain that the potential threshold points are 5.6, 5.8, 6.0, and 6.2. The impulse response patterns are similar when using these potential points to split the sample. Therefore, we define the state as the low unemployment state when the unemployment rate is below 5.8 (the estimate when $h = 1$) and the high unemployment state (slack state) when it is above 5.8. Figure 5 shows the impulse responses to a military spending news shock in government spending and GDP. Compared to Figure 5 in Ramey and Zubairy

[19] Section II.B in Ramey and Zubairy (2018) provides a more detailed description of the data and variables.

[20] Section IV. A. and Section V.A in Ramey and Zubairy (2018) present the narrative reasons for the choice of the variable to define the state of the economy.

(2018) and Figure 4 in Adamek et al. (2024), for government spending, the peak of the responses occurs earlier but lasts longer, and the peak magnitude is slightly smaller in the slack state. In the low unemployment state, the patterns are very similar. For GDP, compared to Figure 5 in Ramey and Zubairy (2018), the peak of the responses occurs 5 periods earlier, and the peak magnitude is slightly smaller in the slack state. In the low unemployment state, the peak occurs at a similar horizon, and the magnitude is slightly larger. While comparing to Figure 4 in Adamek et al. (2024), the peak of the responses occurs 2 periods earlier, and the peak magnitude is slightly smaller in the slack state. In the low unemployment state, the pattern is similar and the peak magnitude is slightly larger. In addition, the impulse responses are insignificant at horizon 0 in the linear, high, and low unemployment states.

# 6 Conclusion

In this paper, we propose a debiased Lasso estimator for high-dimensional slope parameters in threshold regression models, allowing for either cross-sectional or time series data. We derive the asymptotic distribution of tests involving an increasing number of slope parameters and construct uniformly valid confidence bands. We show that the asymptotic distributions are the same in the cases with no threshold effect and a fixed threshold effect. Notably, our study allows for less restrictive assumptions than existing research in high-dimensional threshold models, accommodating heteroskedastic non-subgaussian error terms and non-subgaussian covariates. Future research directions could include generalizing uniform inference theory to dynamic panel data models with threshold effects and extending the current framework to models with multiple threshold points or multiple threshold variables.

Table 3: Estimates and standard errors (in parentheses) with $Q = gdp60$

| | Linear Model | Threshold Model $\widehat{\tau} = 2978$ | |
|---|---|---|---|
| | | $\widehat{\beta}$ | $\widehat{\delta}$ |
| *lgdp60* | $-0.0172^{***}$ | $-0.0139^{***}$ | - |
| | (0.0004) | (0.0004) | |
| $ls_k$ | $0.0078^{***}$ | $0.0070^{***}$ | - |
| | (0.0018) | (0.0018) | |
| *pyrf60* | - | - | $-0.0002^{*}$ |
| | | | (0.0024) |
| *syrf60* | - | - | $(-7.9203 \times 10^{-5})^{*}$ |
| | | | (0.0015) |
| *nof60* | $0.0019^{***}$ | - | $(8.1229 \times 10^{-5})^{***}$ |
| | (0.0008) | | (0.0009) |
| *seccm60* | $0.0047^{***}$ | - | $0.0062^{***}$ |
| | (0.0019) | | (0.0024) |
| *seccf60* | - | $(-9.8805 \times 10^{-5})^{*}$ | - |
| | | (0.0016) | |
| *llife* | $0.0349^{***}$ | $0.0200^{***}$ | - |
| | (0.0008) | (0.0008) | |
| *lfert* | $-0.0136^{***}$ | $-0.0127^{***}$ | - |
| | (0.0021) | (0.0021) | |
| *edu/gdp* | $0.0615^{***}$ | $0.1203^{***}$ | - |
| | (0.0766) | (0.0772) | |
| *gcon/gdp* | $-0.0767^{***}$ | $-0.0840^{***}$ | - |
| | (0.0274) | (0.0272) | |
| *revol* | $-0.0007^{**}$ | $-0.0017^{***}$ | - |
| | (0.0101) | (0.0107) | |
| *wardum* | $-0.0036^{***}$ | $-0.0003^{*}$ | $-0.0033^{***}$ |
| | (0.0041) | (0.0039) | (0.0044) |
| *wartime* | $-0.0140^{***}$ | $-0.0190^{***}$ | - |
| | (0.0057) | (0.0132) | |
| *lbmp* | $-0.0230^{***}$ | $-0.0234^{***}$ | $-0.0007^{**}$ |
| | (0.0091) | (0.0085) | (0.0088) |
| *tot* | $0.0798^{***}$ | - | $0.0971^{***}$ |
| | (0.0526) | | (0.0931) |
| *lgdp60* × *pyrf60* | $(1.6110 \times 10^{-5})^{**}$ | - | $(-3.4547 \times 10^{-5})^{**}$ |
| | (0.0003) | | (0.0004) |
| *lgdp60* × *syrf60* | - | - | $(3.0148 \times 10^{-5})^{***}$ |
| | | | (0.0002) |
| *lgdp60* × *prim60* | $-0.0004^{***}$ | $-0.0004^{***}$ | - |
| | (0.0001) | (0.0001) | |
| *lgdp60* × *seccf60* | - | $(-1.4203 \times 10^{-5})^{*}$ | - |
| | - | $(9.3643 \times 10^{-5})$ | - |
| *lgdp60* × *seccm60* | - | - | $(2.5449 \times 10^{-5})^{**}$ |
| | | | (0.0003) |

*Note:* *** p<0.01, ** p<0.05, * p<0.10.

Table 4: Estimates and standard errors (in parentheses) with $Q = lr$

| | Linear Model | Threshold Model $\widehat{\tau} = 83$ | |
|---|---|---|---|
| | | $\widehat{\beta}$ | $\widehat{\delta}$ |
| *lgdp60* | $-0.0138^{***}$ (0.0004) | $-0.0090^{***}$ (0.0004) | - |
| $ls_k$ | $0.0086^{***}$ (0.0018) | $0.0085^{***}$ (0.0018) | - |
| *pyrf60* | - | - | $-0.0002^{*}$ (0.0023) |
| *seccm60* | $0.0045^{***}$ (0.0019) | - | $0.0061^{***}$ (0.0025) |
| *llife* | $0.0267^{***}$ (0.0008) | $0.0158^{***}$ (0.0008) | - |
| *lfert* | $-0.0138^{***}$ (0.0021) | $-0.0132^{***}$ (0.0021) | - |
| *gcon/gdp* | $-0.0717^{***}$ (0.0284) | $-0.0732^{***}$ (0.0275) | - |
| *revol* | $-0.0013^{***}$ (0.0105) | $-0.0009^{**}$ (0.0111) | - |
| *wardum* | $-0.0011^{***}$ (0.0042) | $-0.0010^{***}$ (0.0040) | $-0.0003^{*}$ (0.0043) |
| *wartime* | $-0.0195^{***}$ (0.0074) | $-0.0245^{***}$ (0.0173) | $-0.0023^{**}$ (0.0205) |
| *lbmp* | $-0.0176^{***}$ (0.0092) | $-0.0148^{***}$ (0.0085) | $-0.0008^{**}$ (0.0090) |
| *tot* | $0.1210^{***}$ (0.0436) | $-0.0637^{***}$ (0.0789) | $0.0112^{**}$ (0.0813) |
| $lgdp60 \times pyrf60$ | $(-8.6678 \times 10^{-5})^{***}$ (0.0003) | $(-2.2235 \times 10^{-5})^{**}$ (0.0003) | $(4.3976 \times 10^{-5})^{***}$ (0.0004) |
| $lgdp60 \times hyrm60$ | $0.0001^{***}$ (0.0002) | $(5.9845 \times 10^{-5})^{***}$ (0.0001) | - |
| $lgdp60 \times nof60$ | $0.0003^{***}$ (0.0001) | $(1.0465 \times 10^{-5})^{***}$ $(4.1041 \times 10^{-5})$ | - |
| $lgdp60 \times prim60$ | $-0.0002^{***}$ (0.0001) | $-0.0003^{***}$ (0.0001) | - |
| $lgdp60 \times prif60$ | $(-8.0069 \times 10^{-5})^{***}$ (0.0001) | $-0.0001^{***}$ (0.0001) | $(-4.1852 \times 10^{-5})^{***}$ (0.0002) |
| $lgdp60 \times seccm60$ | - | - | $(3.4412 \times 10^{-5})^{**}$ (0.0004) |
| $lgdp60 \times seccf60$ | $(-1.0598 \times 10^{-5})^{**}$ (0.0002) | - | - |

*Note:* *** p<0.01, ** p<0.05, * p<0.10.

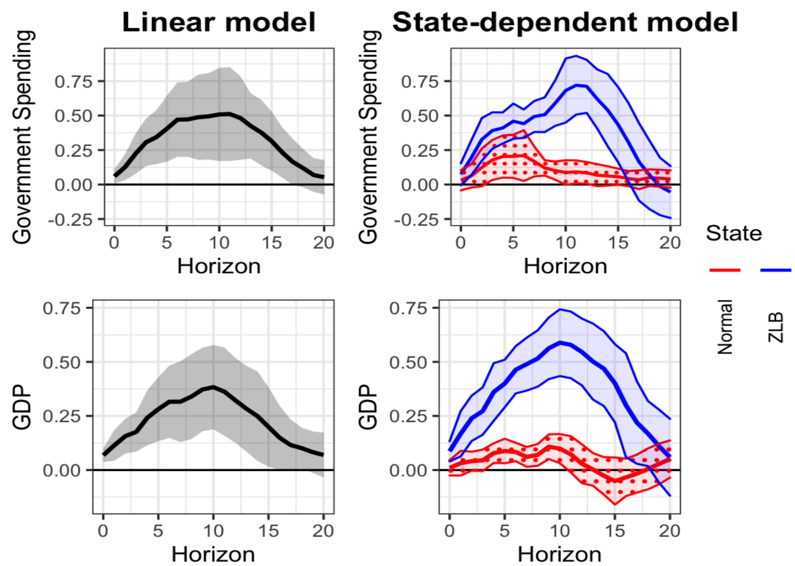



Figure 4: Impulse responses to a military spending news shock of the size of 1% of GDP in government spending and GDP.

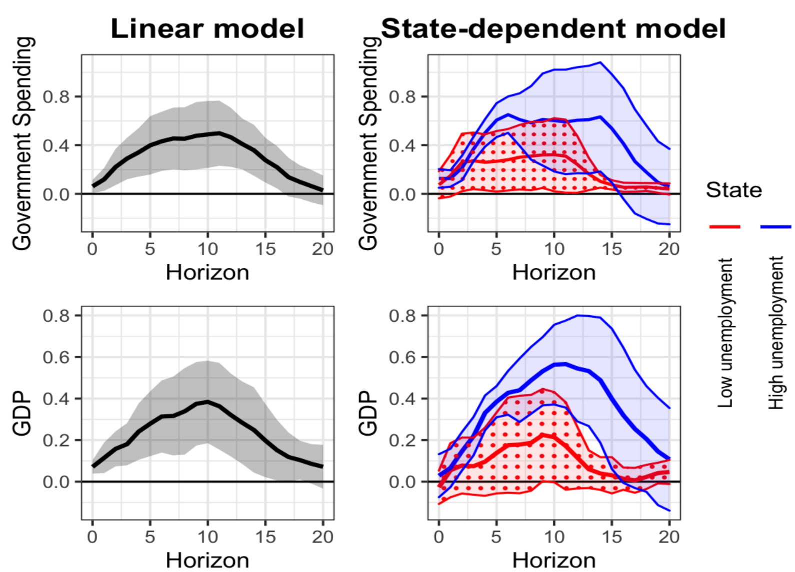



Figure 5: Impulse responses to a military spending news shock of the size of 1% of GDP in government spending and GDP.

# 7 Appendix A

We first recall the concentration inequality from Chernozhukov et al. (2014) and Chernozhukov et al. (2015), as formulated in Lemma 2 of Chiang et al. (2023), to derive oracle inequalities. For notation, $C$ is an arbitrary positive finite constant, and its value may vary from line to line.

**Lemma 2** (A Concentration Inequality)**.** *Let* $\{X_i\}_{i=1}^n$ *be* $p$*-dimensional independent random vectors,* $B = \sqrt{E\left[\max_{1\leq i\leq n}\max_{1\leq j\leq p}\left|X_i^{(j)}\right|^2\right]}$ *and* $\sigma^2 = \max_{1\leq j\leq p} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}\right)^2\right]$. *For* $C > 0$, *with probability at least* $1 - C(logn)^{-1}$,

$$\max_{1\leq j\leq p}\left|\frac{1}{n}\sum_{i=1}^n\left(X_i^{(j)} - E\left[X_i^{(j)}\right]\right)\right| \lesssim \sqrt{\frac{\sigma^2 log(p\vee n)}{n}} + \frac{Blog(p\vee n)}{n}.$$

Without loss of generality, we will assume $p > n$ throughout the appendix.

**Lemma 3** (A Concentration Inequality for Partial Sum of Random Variables)**.** *Let* $\{X_i\}_{i=1}^n$ *be* $p$*-dimensional independent random vectors,* $B = \sqrt{E\left[\max_{1\leq i\leq n}\max_{1\leq j\leq p}\left|X_i^{(j)}\right|^2\right]}$ *and* $\sigma^2 = \max_{1\leq j\leq p} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}\right)^2\right]$. *For* $C > 0$, *with probability at least* $1 - C(logn)^{-1}$, *we have*

$$(i)\ \max_{1\leq j\leq p}\max_{1\leq k\leq n}\left|\frac{1}{n}\sum_{i=1}^k\left(X_i^{(j)} - E\left[X_i^{(j)}\right]\right)\right| \lesssim \sqrt{\frac{\sigma^2 log(pn)}{n}} + \frac{Blog(pn)}{n},$$

$$(ii)\ \max_{1\leq j\leq p}\max_{1\leq q\leq k\leq n}\left|\frac{1}{n}\sum_{i=q}^k\left(X_i^{(j)} - E\left[X_i^{(j)}\right]\right)\right| \lesssim \sqrt{\frac{\sigma^2 log(pn^2)}{n}} + \frac{Blog(pn^2)}{n}.$$

*Proof of Lemma 3.* Denote a deterministic upper triangular matrix with all elements equal to one by $\Xi_{n,n}$ :

$$\Xi_{n,n} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{7.1}$$

Let $\xi_i^{(k)}$ be the $i$-th row, $k$-th column element of $\Xi_{n,n}$, and $\widetilde{\xi}_i^{(q)}$ be the $i$-th row and $q$-th column of the transpose of $\Xi_{n,n}$, denoted by $\Xi_{n,n}^T$.

To prove (i), write

$$\max_{1\le j\le p}\max_{1\le k\le n}\left|\sum_{i=1}^{k}\left(X_i^{(j)}-E\left[X_i^{(j)}\right]\right)\right| = \max_{1\le j\le p}\max_{1\le k\le n}\left|\sum_{i=1}^{n}\left(X_i^{(j)}-E\left[X_i^{(j)}\right]\right)\xi_i^{(k)}\right|$$

and $X_i^{(j)}\xi_i^{(k)}$ is a independent random variable. Due to the speciality of matrix $\Xi_{n,n}$, we obtain $\max_{1\le j\le p}\max_{1\le k\le n} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}\xi_i^{(k)}\right)^2\right] = \max_{1\le j\le p} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}\right)^2\right] = \sigma^2$ and $\sqrt{E\left[\max_{1\le i\le n}\max_{1\le k\le n}\max_{1\le j\le p}\left|X_i^{(j)}\xi_i^{(k)}\right|^2\right]} = \sqrt{E\left[\max_{1\le i\le n}\max_{1\le j\le p}\left|X_i^{(j)}\right|^2\right]}$ $= B$. Then applying Lemma 2 yields

$$\max_{1\le j\le p}\max_{1\le k\le n}\left|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}-E\left[X_i^{(j)}\right]\right)\xi_i^{(k)}\right| \lesssim \sqrt{\frac{\sigma^2 log(pn)}{n}}+\frac{Blog(pn)}{n},$$

(i) thus holds with probability at least $1-C(logn)^{-1}$.

Next for (ii), write

$$\max_{1\le j\le p}\max_{1\le q\le k\le n}\left|\sum_{i=q}^{k}\left(X_i^{(j)}-E\left[X_i^{(j)}\right]\right)\right| = \max_{1\le j\le p}\max_{1\le q\le k\le n}\left|\sum_{i=1}^{n}\left(X_i^{(j)}-E\left[X_i^{(j)}\right]\right)\xi_i^{(k)}\widetilde{\xi}_i^{(q)}\right|$$

and $X_i^{(j)}\xi_i^{(k)}\widetilde{\xi}_i^{(q)}$ is a independent random variable. Similarly, due to the speciality of matrix $\Xi_{n,n}$ and $\Xi_{n,n}^T$, we obtain $\max_{1\le j\le p}\max_{1\le q\le k\le n} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}\xi_i^{(k)}\widetilde{\xi}_i^{(q)}\right)^2\right] =$ $\max_{1\le j\le p} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}\right)^2\right] = \sigma^2$ and $\sqrt{E\left[\max_{1\le i\le n}\max_{1\le q\le k\le n}\max_{1\le j\le p}\left|X_i^{(j)}\xi_i^{(k)}\widetilde{\xi}_i^{(q)}\right|^2\right]}$

$= \sqrt{E\left[\max_{1\le i\le n}\max_{1\le j\le p}\left|X_i^{(j)}\right|^2\right]} = B$. Then applying Lemma 2 yields

$$\max_{1\le j\le p}\max_{1\le q\le k\le n}\left|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)} - E\left[X_i^{(j)}\right]\right)\xi_i^{(k)}\widetilde{\xi}_i^{(q)}\right| \lesssim \sqrt{\frac{\sigma^2 log(n^2p)}{n}} + \frac{Blog(n^2p)}{n},$$

(ii) thus holds with probability at least $1 - C(logn)^{-1}$. □

## 7.1 Proofs for Section 2.2

To establish the prediction consistency of the Lasso estimator, we define some regularized events and provide some inequalities.

**Lemma 4** (Regularized events $\mathbb{A}_1$ and $\mathbb{A}_2$)**.** *Suppose that Assumption 1 holds and set* $\lambda$ *by* (2.8)*. Let* $\mu_1 = \mu/A$ *and define the events*

$$\mathbb{A}_1 = \left\{\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}\right)^2 \le C_2^2 + \mu_1\lambda\right\}, \quad \mathbb{A}_2 = \left\{\min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}(t_0)\right)^2 \ge C_3^2 - \mu_1\lambda\right\},$$

*In particular* $\mathbb{A}_2 \subseteq \mathbb{A}_2' = \left\{\min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}\right)^2 \ge C_3^2 - \mu_1\lambda\right\}$*, then* $P(\mathbb{A}_1) \ge 1 - C(logn)^{-1}$, $P(\mathbb{A}_2) \ge 1 - C(logn)^{-1}, P(\mathbb{A}_2') \ge 1 - C(logn)^{-1}$*, and* $P(\mathbb{A}_1 \cap \mathbb{A}_2) \ge 1 - C(logn)^{-1}$.

*Proof of Lemma 4.* Under Assumption 1, let $\sigma^2 = \max_{1\le j\le p} 1/n\sum_{i=1}^{n} E\left[\left(X_i^{(j)}\right)^4\right] \le C_2^4$, which is bounded, and $B = \sqrt{M_{XX}^2}$; by Lemma 2,

$$\max_{1\le j\le p}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\left(X_i^{(j)}\right)^2 - E\left[\left(X_i^{(j)}\right)^2\right]\right)\right| \lesssim \sqrt{\frac{log(p)}{n}}$$

holds with probability at least $1 - C(logn)^{-1}$. Thus, with probability at least $1 - C(logn)^{-1}$,

$$\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}\right)^2 \lesssim \max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}E\left[\left(X_i^{(j)}\right)^2\right] + \sqrt{\frac{log(p)}{n}} \lesssim C_2^2 + \sqrt{\frac{log(p)}{n}} = C_2^2 + \mu_1\lambda,$$

which implies that $\mathbb{A}_1$ holds.

Next, consider $\mathbb{A}_2$. Similarly, under Assumption 1, let $\sigma^2 = \max_{1\le j\le p} 1/n \sum_{i=1}^n E\left[\left(X_i^{(j)}(t_0)\right)^4\right]$ $\le C_2^4$, which is bounded, and $B = \sqrt{M_{Xt_0}^2} \le \sqrt{M_{XX}^2}$; by Lemma 2,

$$\max_{1\le j\le p}\left|\frac{1}{n}\sum_{i=1}^n\left(\left(X_i^{(j)}(t_0)\right)^2 - E\left[\left(X_i^{(j)}(t_0)\right)^2\right]\right)\right| \lesssim \sqrt{\frac{log(p)}{n}}$$

holds with probability at least $1 - C(logn)^{-1}$. Thus, with probability at least $1 - C(logn)^{-1}$,

$$\min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^n\left(X_i^{(j)}(t_0)\right)^2 \gtrsim \max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^n E\left[\left(X_i^{(j)}(t_0)\right)^2\right] - \sqrt{\frac{log(p)}{n}} \gtrsim C_3^2 - \sqrt{\frac{log(p)}{n}} = C_3^2 - C^{-1}\mu\lambda,$$

which implies that $\mathbb{A}_2$ holds. By the same steps, we can obtain that $\mathbb{A}_2'$ holds with probability at least $1 - C(logn)^{-1}$.

Since $P(\mathbb{A}_1 \cap \mathbb{A}_2) \ge 1 - P(\mathbb{A}_1^c) - P(\mathbb{A}_2^c)$, we prove the lemma. □

**Lemma 5** (Regularized events $\mathbb{A}_3$ and $\mathbb{A}_4$)**.** *Suppose that Assumption 1 hold and set $\lambda$ by (2.8). Let $\mu_2 = 2\mu_1/C_3$ and define*

$$\mathbb{A}_3 := \left\{\max_{1\le j\le p}\frac{1}{\|X^{(j)}\|_n}\left|\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)}\right| \le \frac{\mu_2\lambda}{2}\right\},$$

$$\mathbb{A}_4 := \left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\|X^{(j)}(\tau)\|_n}\left|\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)}\mathbf{1}\{Q_i < \tau\}\right| \le \frac{\mu_2\lambda}{2}\right\},$$

*Then,* $P(\mathbb{A}_3) \ge 1 - C(logn)^{-1}, P(\mathbb{A}_4) \ge 1 - C(logn)^{-1}$, *and* $P(\mathbb{A}_3 \cap \mathbb{A}_4) \ge 1 - C(logn)^{-1}$.

*Proof of Lemma 5.* Under Assumption 1, let $\sigma^2 = \max_{1\le j\le p} 1/n \sum_{i=1}^n E\left[\left(U_i X_i^{(j)}\right)^2\right]$, which is bounded, and $B = \sqrt{M_{UX}^2}$, by Lemma 2,

$$\max_{1\le j\le p}\left|\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)}\right| \lesssim \sqrt{\frac{log(p)}{n}},$$

holds with probability at least $1 - C(logn)^{-1}$, and conditional on $\mathbb{A}_2'$,

$$\max_{1\le j\le p}\left|\frac{1}{\|X^{(j)}\|_n}\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)}\right| \lesssim \frac{1}{\min_{1\le j\le p}\|X^{(j)}\|_n}\sqrt{\frac{log(p)}{n}} \le \frac{\mu_1\lambda}{C_3}.$$

We thus obtain that $\mathbb{A}_3$ holds with probability at least $1 - C(logn)^{-1}$.

Next, consider the event $\mathbb{A}_4$. Since we have $n$ observations, sort $\{X_i, U_i, Q_i\}_{i=1}^n$ by $(Q_1, \ldots, Q_n)$ in ascending order. Given the sorted $Q_i$, the supremum over $\tau$ is achieved at one of the points $Q_{(i)}$. Thus, for $j = 1, \ldots, p$,

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\|X^{(j)}(\tau)\|_n}\left|\frac{1}{n}\sum_{i=1}^n U_iX_i^{(j)}(\tau)\right|\le\frac{\mu_2\lambda}{2}\right\}\\
\ge&\mathbb{P}\left\{\frac{1}{\min_{1\le j\le p}\|X^{(j)}(t_0)\|_n}\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\left|\frac{1}{n}\sum_{i=1}^n U_iX_i^{(j)}(\tau)\right|\le\frac{\mu_2\lambda}{2}\right\}\\
=&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}\left|\frac{1}{n}\sum_{i=1}^k U_iX_i^{(j)}\right|\le\frac{\mu_2\lambda}{2}\min_{1\le j\le p}\left\|X^{(j)}(t_0)\right\|_n\right\}\\
\ge&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}\left|\frac{1}{n}\sum_{i=1}^k U_iX_i^{(j)}\right|\le\sqrt{\frac{log(p)}{n}}\right\}.
\end{aligned}
$$

Under Assumption 1, let $\sigma^2 = \max_{1\le j\le p} 1/n\sum_{i=1}^n E\left[\left(U_iX_i^{(j)}\right)^2\right]$, which is bounded, and $B \le \sqrt{M_{UX}^2}$, by Lemma 3 (i), with probability at least $1 - C(logn)^{-1}$,

$$
\max_{1\le j\le p}\max_{1\le k\le n}\left|\frac{1}{n}\sum_{i=1}^k U_iX_i^{(j)}\right|\lesssim\sqrt{\frac{log(p)}{n}},
$$

which implies $\mathbb{A}_4$ holds. □

We will follow appendix C in Lee et al. (2016) to establish our oracle inequalities. Define $J_0 = J(\alpha_0)$, $\widehat{\boldsymbol{D}} = \widehat{\boldsymbol{D}}(\widehat{\tau})$, $\boldsymbol{D} = \boldsymbol{D}(\tau_0)$ and $R_n = R_n(\alpha_0, \tau_0)$, where

$$
R_n(\alpha,\tau) := 2n^{-1}\sum_{i=1}^n U_iX_i'\delta\left\{1(Q_i<\widehat{\tau})-1(Q_i<\tau)\right\}.
$$

**Lemma 6.** *Conditional on the events* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ *and* $\mathbb{A}_4$, *for* $0 < \mu < 1$, *we have*

$$
\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right|_1\le 2\lambda\left|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right|_1+\lambda\left|\left|\widehat{\boldsymbol{D}}\alpha_0\right|_1-\left|\boldsymbol{D}\alpha_0\right|_1\right|+R_n, \tag{7.2}
$$

$$\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right|_1\leq 2\lambda\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1+\left\|f_{(\alpha_0,\widehat{\tau})}-f_0\right\|_n^2. \quad (7.3)$$

Lemma 6 directly follows from Lemma 5 in Lee et al. (2016), so the proof is omitted.

*Proof of Lemma 1.* Conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, the three terms on the right-hand side of (7.2) can be bounded as follows by using Hölder's inequality:

$$|R_n|\leq 2\mu_2\lambda\sum_{j=1}^{p}\left\|X^{(j)}\right\|_n\left|\delta_0^{(j)}\right|\leq 2\mu_2\left|\delta_0\right|_1\lambda\sqrt{C_2^2+\mu_1\lambda.}, \quad (7.4)$$

$$\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1\leq\left\|\widehat{\boldsymbol{D}}\right\|_\infty|(\widehat{\alpha}-\alpha_0)_{J_0}|_1\leq|(\widehat{\alpha}-\alpha_0)_{J_0}|_1\sqrt{C_2^2+\mu_1\lambda}, \quad (7.5)$$

$$\left|\left|\widehat{\boldsymbol{D}}\alpha_0\right|_1-|\boldsymbol{D}\alpha_0|_1\right|\leq\left|(\widehat{\boldsymbol{D}}-\boldsymbol{D})\alpha_0\right|_1\leq\left\|\widehat{\boldsymbol{D}}-\boldsymbol{D}\right\|_\infty|\alpha_0|_1\leq 2\left|\alpha_0\right|_1\sqrt{C_2^2+\mu_1\lambda} \quad (7.6)$$

Combining (7.4), (7.5) and (7.6) with (7.2) yields

$$\begin{aligned}\left\|\widehat{f}-f_0\right\|_n^2\leq&\Big(2\left\|(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+2\left\|\alpha_0\right\|_1+2\mu_2\left\|\delta_0\right\|_1\Big)\lambda\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}\\ \leq&(6+2\mu_2)C_1\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}s_0\lambda.\end{aligned}$$

□

## 7.2 Proofs for Section 2.2.1

Our first result is a preliminary lemma that can be used to prove adaptive restricted eigenvalue condition.

**Lemma 7.** *Suppose that Assumption 1 hold, with probability at least* $1-C(logn)^{-1}$, *we have*

$$\left\|\frac{1}{n}\sum_{i=1}^{n}X_iX_i'-\frac{1}{n}\sum_{i=1}^{n}E\left[X_iX_i'\right]\right\|_\infty=O_P\left(\sqrt{\frac{log(p)}{n}}\right),$$

$$\sup_{\tau\in\mathbb{T}}\left\|\frac{1}{n}\sum_{i=1}^{n}X_i(\tau)X_i(\tau)'-\frac{1}{n}\sum_{i=1}^{n}E\left[X_i(\tau)X_i(\tau)'\right]\right\|_\infty=O_P\left(\sqrt{\frac{log(p)}{n}}\right).$$

*Proof of Lemma 7.* Under Assumption 1, let $\sigma^2 = \max_{1\leq j,l\leq p} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}X_i^{(l)}\right)^2\right]$, which is bounded, and $B = \sqrt{M_{XX}^2}$, by Lemma 2, with probability at least $1 - C(logn)^{-1}$,

$$\max_{1\leq j,l\leq p}\left|\frac{1}{n}\sum_{i=1}^n\left(X_i^{(j)}X_i^{(l)} - E\left[(X_i^{(j)}X_i^{(l)}\right]\right)\right| \lesssim \sqrt{\frac{log(p)}{n}}, \tag{7.7}$$

implying that

$$\max_{1\leq j,l\leq p}\left|\frac{1}{n}\sum_{i=1}^n\left(X_i^{(j)}X_i^{(l)}\right) - \frac{1}{n}\sum_{i=1}^n E\left[X_i^{(j)}X_i^{(l)}\right]\right| = O_P\left(\sqrt{\frac{log(p)}{n}}\right).$$

Next, sort $(X_i, U_i, Q_i)_{i=1}^n$ by $(Q_1, ..., Q_n)$ in ascending order, we have

$$\begin{aligned}&\mathbb{P}\left\{\max_{1\leq j,l\leq p}\sup_{\tau\in\mathbb{T}}\left|\frac{1}{n}\sum_{i=1}^n\left(X_i^{(j)}X_i^{(l)}1\left(Q_i<\tau\right) - 1\left(Q_i<\tau\right)E\left[X_i^{(j)}X_i^{(l)}\right]\right)\right|\leq t\right\}\\ \geq&\mathbb{P}\left\{\max_{1\leq j,l\leq p}\max_{1\leq k\leq n}\left|\frac{1}{n}\sum_{i=1}^k\left(X_i^{(j)}X_i^{(l)} - E\left[X_i^{(j)}X_i^{(l)}\right]\right)\right|\leq t\right\}.\end{aligned} \tag{7.8}$$

Under Assumption 1, let $\sigma^2 = \max_{1\leq j,l\leq p} 1/n\sum_{i=1}^n E\left[\left(X_i^{(j)}X_i^{(l)}\right)^2\right]$, which is bounded, and $B \leq \sqrt{M_{XX}^2}$, by Lemma 3 (i), with probability at least $1 - C(logn)^{-1}$,

$$\max_{1\leq j,l\leq p}\max_{1\leq k\leq n}\left|\frac{1}{n}\sum_{i=1}^k\left(X_i^{(j)}X_i^{(l)} - E\left[X_i^{(j)}X_i^{(l)}\right]\right)\right| \lesssim \sqrt{\frac{log(p)}{n}},$$

implying that

$$\sup_{\tau\in\mathbb{T}}\max_{1\leq j,l\leq p}\left|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}1\left(Q_i<\tau\right) - \frac{1}{n}\sum_{i=1}^n E\left[X_i^{(j)}X_i^{(l)}1\left(Q_i<\tau\right)\right]\right| = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

□

Now we consider the empirical UARE condition. Define

$$\widehat{\kappa}(s_0, c_0, \mathbb{T}, \widehat{\mathbf{\Sigma}}) = \min_{\tau\in\mathbb{T}} \min_{J_0\subset\{1,\ldots,2p\},|J_0|\le s_0} \min_{\gamma\neq 0,|\gamma_{J_0^c}|_1\le c_0|\gamma_{J_0}|_1} \frac{(\gamma' 1/n \boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma)^{1/2}}{\|\gamma_{J_0}\|_2},$$

and recall Assumption 2 (2.10)

$$\kappa(s_0, c_0, \mathbb{T}, \mathbf{\Sigma}) = \min_{\tau\in\mathbb{T}} \min_{J_0\subset\{1,\ldots,2p\},|J_0|\le s_0} \min_{\gamma\neq 0,|\gamma_{J_0^c}|_1\le c_0|\gamma_{J_0}|_1} \frac{(\gamma' E\left[1/n\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma)^{1/2}}{\|\gamma_{J_0}\|_2} > 0.$$

**Lemma 8.** *Suppose that Assumptions 1-2 hold, let*

$$\mathbb{A}_5 := \left\{ \frac{\kappa(c_0, \mathbb{T}, \mathbf{\Sigma})^2}{2} < \widehat{\kappa}(c_0, \mathbb{T}, \widehat{\mathbf{\Sigma}})^2 \right\},$$

$\mathbb{A}_5$ *holds with probability at least* $1 - C(logn)^{-1}$.

*Proof of Lemma 8.* Write

$$\begin{aligned}
\left|\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma\right| &= \left|\gamma'\left(\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right] + E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\right)\gamma\right| \\
&\ge \left|\gamma' E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma\right| - \left|\gamma'\left(\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\right)\gamma\right|.
\end{aligned}$$

By Holders' inequality,

$$\left|\gamma'\left(\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\right)\gamma\right| \le |\gamma|_1^2 \left\|\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right\|_\infty. \tag{7.9}$$

According to the defined restriction set, we have

$$|\gamma|_1 \le |\gamma_{J_0}|_1 + \left|\gamma_{J_0^c}\right|_1 \le (1+c_0)\,|\gamma_{J_0}|_1 \le (1+c_0)\sqrt{s_0}\,|\gamma_{J_0}|_2\,,$$

thus $\frac{|\gamma|_1}{|\gamma_{J_0}|_2} \le (1+c_0)\sqrt{s_0}$. Dividing (7.9) by $|\gamma_{J_0}|_2^2$ yields

$$
\begin{aligned}
&\left| \frac{|\gamma' \frac{1}{n} \boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma|}{|\gamma_{J_0}|_2^2} - \frac{|\gamma' E\left[1/n \sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma|}{|\gamma_{J_0}|_2^2} \right| \\
&\leq \frac{|\gamma|_1^2}{|\gamma_{J_0}|_2^2} \left\| \frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right] \right\|_\infty \\
&\leq (1+c_0)^2 s_0 \left\| \frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right] \right\|_\infty .
\end{aligned}
$$

Meanwhile,

$$
\begin{aligned}
&\frac{|\gamma' E\left[1/n \sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma|}{|\gamma_{J_0}|_2^2} - \frac{|\gamma' 1/n \boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma|}{|\gamma_{J_0}|_2^2} \\
&\leq \left| \frac{|\gamma' 1/n \boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma|}{|\gamma_{J_0}|_2^2} - \frac{|\gamma' E\left[1/n \sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma|}{|\gamma_{J_0}\|_2^2} \right| .
\end{aligned}
$$

We thus obtain

$$
\begin{aligned}
\frac{|\gamma' 1/n \boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma|}{|\gamma_{J_0}|_2^2} \geq & \frac{|\gamma' E\left[1/n \sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma|}{|\gamma_{J_0}|_2^2} \\
& - (1+c_0)^2 s_0 \left\| \frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right] \right\|_\infty .
\end{aligned}
$$

Minimizing the right hand side over $\tau \in \mathbb{T}$ yields

$$
\begin{aligned}
\frac{|\gamma' 1/n \boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma|}{|\gamma_{J_0}|_2^2} \geq & \min_{\tau\in\mathbb{T}} \frac{|\gamma' E\left[1/n \sum_{i=1}^n \boldsymbol{X}_i(\tau)\boldsymbol{X}_i(\tau)'\right]\gamma|}{|\gamma_{J_0}|_2^2} \\
& - (1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}} \left\| \frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right] \right\|_\infty .
\end{aligned}
$$

Then minimizing the right hand side over $\left\{\gamma \neq 0, |\gamma_{J_0^c}|_1 \leq c_0 |\gamma_{J_0}|_1\right\}$ yileds

$$
\frac{|\gamma' 1/n \boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma|}{|\gamma_{J_0}|_2^2} \geq \kappa(c_0, \mathbb{T}, \Sigma)^2 - (1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}} \left\| \frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau) - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right] \right\|_\infty .
$$

The above inequality holds for all $\tau \in \mathbb{T}$ and $\left\{\gamma \neq 0, |\gamma_{J_0^c}|_1 \leq c_0 |\gamma_{J_0}|_1\right\}$, so we minimize the left hand side over $\tau \in \mathbb{T}$ and $\left\{\gamma \neq 0, |\gamma_{J_0^c}|_1 \leq c_0|\gamma_{J_0}|_1\right\}$ and derive

$$\widehat{\kappa}(c_0, \mathbb{T}, \widehat{\boldsymbol{\Sigma}})^2 \geq \kappa(c_0, \mathbb{T}, \boldsymbol{\Sigma})^2 - (1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty . \tag{7.10}$$

By Lemma 7, with probability at least $1 - C(logn)^{-1}$,

$$(1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - \frac{1}{n} \sum_{i=1}^{n} E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty \lesssim \left( \sqrt{\frac{\log p}{n}} \right), \tag{7.11}$$

thus, with probability at least $1 - C(logn)^{-1}$, $(1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} ||1/n \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[1/n \sum_{i=1}^n \boldsymbol{X}_i(\tau) \boldsymbol{X}_i(\tau)'\right] ||_\infty \leq \frac{\kappa(c_0, \mathbb{T}, \Sigma)^2}{2}$, we prove the lemma. □

**Lemma 9.** *Suppose that $\delta_0 = 0$ and that Assumptions 1 amd 2 hold with $\kappa = \kappa\left(\frac{1+\mu}{1-\mu}, \mathbb{T}, \boldsymbol{\Sigma}\right)$ for $\mu \in (0,1)$. Let $(\widehat{\alpha}, \widehat{\tau})$ be the Lasso estimator defined by (2.4) with $\lambda = \frac{C}{\mu} \frac{\sqrt{\log p}}{\sqrt{n}}$. Then, conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, we have*

$$\left\| \widehat{f} - f_0 \right\|_n \leq \frac{2\sqrt{2}}{\kappa} \left( \sqrt{C_2^2 + \mu_1 \lambda} \right) \sqrt{s_0} \lambda,$$
$$|\widehat{\alpha} - \alpha_0|_1 \leq \frac{4\sqrt{2}}{(1-\mu)\kappa^2} \frac{C_2^2 + \mu_1 \lambda}{\sqrt{C_3^2 - \mu_1 \lambda}} s_0 \lambda.$$

*Proof of Lemma 9.* Following the proof of Lemma 9 in Lee et al. (2016), conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, we have

$$\left\| \widehat{f} - f_0 \right\|_n^2 + (1-\mu)\lambda \left| \widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0) \right|_1 \leq 2\lambda \left| \widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0} \right|_1, \tag{7.12}$$

which implies that

$$\left| \widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0^c} \right|_1 \leq \frac{1+\mu}{1-\mu} \left| \widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0} \right|_1. \tag{7.13}$$

And we have,

$$
\begin{aligned}
&\kappa^2 \left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2^2 \le 2\widehat{\kappa}^2 \left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_2^2 \le \frac{2}{n}\left|\boldsymbol{X}(\widehat{\tau})\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right|_2^2 \\
&= \frac{2}{n}(\widehat{\alpha}-\alpha_0)'\widehat{\boldsymbol{D}}\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0) \le \frac{2\max(\widehat{\boldsymbol{D}})^2}{n}(\widehat{\alpha}-\alpha_0)'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})(\widehat{\alpha}-\alpha_0) \\
&= 2\max(\widehat{\boldsymbol{D}})^2\left|\widehat{f}-f_0\right|_n^2.
\end{aligned} \tag{7.14}
$$

Combining (7.12) with (7.14) yields

$$
\left\|\widehat{f}-f_0\right\|_n^2 \le 2\lambda\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1 \le 2\lambda\sqrt{s_0}\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_2 \le \frac{2\sqrt{2}\lambda}{\kappa}\sqrt{s_0}\max(\widehat{\boldsymbol{D}})\left\|\widehat{f}-f_0\right\|_n,
$$

then conditional on $\mathbb{A}_1$, we obtain

$$
\left\|\widehat{f}-f_0\right\|_n \le \frac{2\sqrt{2}}{\kappa}\left(\sqrt{C_2^2+\mu_1\lambda}\right)\sqrt{s_0}\lambda.
$$

Next, conditional on $\mathbb{A}_1$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, by (7.13) and (7.14),

$$
\begin{aligned}
&\left|\widehat{\boldsymbol{D}}\left(\widehat{\alpha}-\alpha_0\right)\right|_1 = \left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1 + \left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0^c}\right|_1 \le 2\left(1-\mu\right)^{-1}\sqrt{s_0}\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_2 \\
&\le \frac{2}{\kappa\left(1-\mu\right)}\sqrt{s_0}\max(\widehat{\boldsymbol{D}})\left\|\widehat{f}-f_0\right\|_n \le \frac{4\sqrt{2}\lambda}{\left(1-\mu\right)\kappa^2}s_0\max(\widehat{\boldsymbol{D}})^2 \le \frac{4\sqrt{2}\lambda}{\left(1-\mu\right)\kappa^2}s_0(C_2^2+\mu_1\lambda),
\end{aligned} \tag{7.15}
$$

conditional on $\mathbb{A}_2'$,

$$
\left|\widehat{\boldsymbol{D}}\left(\widehat{\alpha}-\alpha_0\right)\right|_1 \ge \min(\widehat{\boldsymbol{D}})\left|\widehat{\alpha}-\alpha_0\right|_1 \ge \sqrt{C_3^2-\mu_1\lambda}\left|\widehat{\alpha}-\alpha_0\right|_1, \tag{7.16}
$$

we thus have

$$
\left|\widehat{\alpha}-\alpha_0\right|_1 \le \frac{4\sqrt{2}}{\left(1-\mu\right)\kappa^2}\frac{C_2^2+\mu_1\lambda}{\sqrt{C_3^2-\mu_1\lambda}}s_0\lambda. \tag{7.17}
$$

□

*Proof of Theorem 1.* The proof follows immediately from combining Assumptions 1 and 2 with Lemma 9. Specially, $P(\mathbb{A}_1\cap\mathbb{A}_2\cap\mathbb{A}_3\cap\mathbb{A}_4\cap\mathbb{A}_5)\ge 1-C(logn)^{-1}$. □

## 7.3 Proofs for Section 2.2.2

The following lemma provides an upper bound for $|\widehat{\tau} - \tau_0|$.

**Lemma 10.** *Suppose that Assumption 3 holds. Let*

$$\eta^* = \max\left\{\min_i |Q_i - \tau_0|, \frac{1}{C_4}\Big(2C_1(3+\mu_2)\left(C_2^2 + \mu_1\lambda\right)^{\frac{1}{2}} s_0\lambda\Big)\right\},$$

*then, conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, we have*

$$|\widehat{\tau} - \tau_0| \leq \eta^*.$$

*Proof of Lemma 10.* Following the proof of Lemma 11 in Lee et al. (2016), conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, we have

$$\widehat{S}_n - S_n(\alpha_0, \tau_0) \geq \left\|\widehat{f} - f_0\right\|_n^2 - \mu\lambda\left|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)\right|_1 - R_n, \tag{7.18}$$

and

$$\begin{aligned}
&\left[\widehat{S}_n + \lambda\left|\widehat{\boldsymbol{D}}\widehat{\alpha}\right|_1\right] - \left[S_n(\alpha_0, \tau_0) + \lambda\left|\boldsymbol{D}\alpha_0\right|_1\right] \\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - 2\lambda\left|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0}\right|_1 - \lambda\left|\left|\boldsymbol{D}\alpha_0\right|_1 - \left|\widehat{\boldsymbol{D}}\widehat{\alpha}\right|_1\right| - R_n \\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - \left(6\lambda\sqrt{C_2^2 + \mu_1\lambda}C_1 s_0 + 2\mu_2\lambda\sqrt{C_2^2 + \mu_1\lambda}C_1 s_0\right) \\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - \Big(2C_1(3+\mu_2)\left(C_2^2 + \mu_1\lambda\right)^{\frac{1}{2}} s_0\lambda\Big) \geq 0,
\end{aligned} \tag{7.19}$$

the second inequality in (7.19) is obtained by applying (7.4), (7.5) and (7.6) to bound the last three terms, and the last inequality follows from Lemma 1.
Now suppose that $|\widehat{\tau} - \tau_0| > \eta^*$, then Assumption 3 and (7.19) together imply that

$$\left[\widehat{S}_n + \lambda\left|\widehat{\boldsymbol{D}}\widehat{\alpha}\right|_1\right] - \left[S_n(\alpha_0, \tau_0) + \lambda\left|\boldsymbol{D}\alpha_0\right|_1\right] \geq \left\|\widehat{f} - f_0\right\|_n^2 - C_4\eta^* > 0,$$

which leads to contradiction as $\widehat{\tau}$ is the minimizer of (2.4). Therefore, we have proved the lemma. □

**Lemma 11.** *Suppose that Assumption 1 holds, then for any $\eta > C\log p/n > 0$, with $C > 0$, there exists a finite constant $C_5$, such that with probability at least $1 -$*

$C(logn)^{-1}$,

$$\sup_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|<\eta}\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}X_i^{(l)}\right|\left|1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right|\leq C_5\eta, \tag{7.20}$$

$$\sup_{|\tau-\tau_0|<\eta}\left|\frac{1}{n}\sum_{i=1}^{n}U_iX_i'\delta_0\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|\leq\frac{\lambda\sqrt{\eta}}{2}. \tag{7.21}$$

*Proof of Lemma 11.* Under Assumption 1, $Q_i$ is continuously distributed, and $E\left(X_i^{(j)}X_i^{(l)}|Q_i=\tau\right)$ is continuous and bounded in a neighborhood of $\tau_0$ for all $1\leq j,l\leq p$, and by (7.7), (7.20) holds immediately.

To show that (7.21) holds, we sort $\{X_i,U_i,Q_i\}_{i=1}^n$ by $(Q_1,...,Q_n)$ in ascending order, and obtain

$$\begin{aligned}
&\mathbb{P}\left(\sup_{|\tau-\tau_0|<\eta}\left|\frac{1}{n}\sum_{i=1}^{n}U_iX_i'\delta_0\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|\leq\frac{\lambda\sqrt{\eta}}{2}\right)\\
&\geq\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=\min\{1,[n(\tau_0-\eta)]\}}^{\max\{[n(\tau_0+\eta)],n\}}U_i^2\right|^{1/2}\leq\frac{\lambda}{2\sqrt{2}h_n(\eta)}\right)\\
&=\mathbb{P}\left(\sup_{|\tau-\tau_0|<\eta}\left|\frac{1}{n}\sum_{i=1}^{n}U_i^2\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|^{1/2}\leq\frac{\lambda}{2\sqrt{2}h_n(\eta)}\right),
\end{aligned}$$

under Assumption 1 (iv), with $\eta>Clogp/n>0$, by lemma 3, (7.21) holds with probability at least $1-C(logn)^{-1}$. □

We now provide a lemma for bounding the prediction loss and the $l_1$ estimation loss for $\alpha_0$. To do so, we define the following $G_1$, $G_2$ and $G_3$:

$$\begin{aligned}
G_1=&\sqrt{c_\tau}+\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5\|\delta_0\|_1c_\tau,\quad G_2=\frac{12\left(C_2^2+\mu_1\lambda\right)}{\kappa^2},\\
G_3=&\frac{2\sqrt{2}\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{\kappa}\left(c_\alpha c_\tau\right)^{1/2}.
\end{aligned}$$

**Lemma 12.** *Suppose that* $|\widehat{\tau}-\tau_0|\leq c_\tau$ *and* $|\widehat{\alpha}-\alpha_0|_1\leq c_\alpha$ *for some* $(c_\tau,c_\alpha)$. *Suppose that Assumptions 2 and 4 hold with* $\mathbb{S}=\{|\tau-\tau_0|\leq c_\tau\}$, $\kappa=\kappa\left(s_0,\frac{2+\mu}{1-\mu},\mathbb{S},\mathbf{\Sigma}\right)$ *for*

$0 < \mu < 1$. *Then, conditional on* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, *and* $\mathbb{A}_5$, *we have*

$$\begin{aligned}
\left\|\widehat{f} - f_0\right\|_n^2 &\leq 3\lambda \cdot \left\{G_1 \vee G_2\lambda s_0 \vee G_3\sqrt{s_0|\delta_0|_1}\right\}, \\
|\widehat{\alpha} - \alpha_0|_1 &\leq \frac{3}{(1-\mu)\sqrt{C_3^2 - \mu_1\lambda}} \cdot \left\{G_1 \vee G_2\lambda s_0 \vee G_3\sqrt{s_0|\delta_0|_1}\right\}.
\end{aligned}$$

*Proof of Lemma 12.* The proof follows that of Lemma 12 in Lee et al. (2016). We have

$$|R_n| = \left|2n^{-1}\sum_{i=1}^{n} U_i X_i' \delta_0 \left\{1(Q_i < \widehat{\tau}) - 1(Q_i < \tau_0)\right\}\right| \leq \lambda\sqrt{c_\tau}, \tag{7.22}$$

by (7.21). Conditioning on $\mathbb{A}_4$, the triangular inequality implies that

$$\begin{aligned}
\left|\left|\widehat{\boldsymbol{D}}\alpha_0\right|_1 - |\boldsymbol{D}\alpha_0|_1\right| &\leq \left|\sum_{j=1}^{p}\left(\left\|X^{(j)}(\widehat{\tau})\right\|_n - \left\|X^{(j)}(\tau_0)\right\|_n\right)\left|\delta_0^{(j)}\right|\right| \\
&\leq \sum_{j=1}^{p}\left(2\left\|X^{(j)}(t_0)\right\|_n\right)^{-1}\left|\left\|X^{(j)}(\widehat{\tau})\right\|_n^2 - \left\|X^{(j)}(\tau_0)\right\|_n^2\right|\left|\delta_0^{(j)}\right| \\
&\leq \sum_{j=1}^{p}\left(2\left\|X^{(j)}(t_0)\right\|_n\right)^{-1}\left|\delta_0^{(j)}\right|\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}\right|^2\left|\mathbf{1}\{Q_i < \widehat{\tau}\} - \mathbf{1}\{Q_i < \tau_0\}\right| \\
&\leq \left(2\sqrt{C_3^2 - \mu_1\lambda}\right)^{-1}|\delta_0|_1 C_5 c_\tau,
\end{aligned} \tag{7.23}$$

where the last inequality is by Assumption 4. We now consider two cases:
(i) $\left|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0}\right|_1 > \sqrt{c_\tau} + \left(2\sqrt{C_3^2 - \mu_1\lambda}\right)^{-1} C_5|\delta_0|_1 c_\tau$ and
(ii) $\left|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0}\right|_1 \leq \sqrt{c_\tau} + \left(2\sqrt{C_3^2 - \mu_1\lambda}\right)^{-1} C_5|\delta_0|c_\tau$.
**Case (i)**: Combining (7.22) and (7.23) yields

$$\begin{aligned}
\lambda\left|\left|\widehat{\boldsymbol{D}}\alpha_0\right|_1 - |\boldsymbol{D}\alpha_0|_1\right| + R_n &< \lambda\left(2\sqrt{C_3^2 - \mu_1\lambda}\right)^{-1}|\delta_0|_1 C_5(c_\tau + \lambda\sqrt{c_\tau}) + \lambda\sqrt{c_\tau} \\
&< \lambda\left|\widehat{\boldsymbol{D}}\left(\widehat{\alpha} - \alpha_0\right)_{J_0}\right|_1.
\end{aligned}$$

Along with (7.2), we have

$$\left|\widehat{f}-f_0\right|_n^2+(1-\mu)\,\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1\le 3\lambda\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1, \tag{7.24}$$

which implies

$$(1-\mu)\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right|_1\le 3\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1.$$

Then, subtracting $(1-\mu)\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1$ from both sides yields

$$\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0^c}\right|_1\le\frac{2+\mu}{1-\mu}\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_1. \tag{7.25}$$

In this case, we are applying Assumption 2 with adaptive restricted eigenvalue condition $\kappa\left(s_0,\frac{2+\mu}{1-\mu},\mathbb{S},\boldsymbol{\Sigma}\right)$. Since $|\widehat{\tau}-\tau_0|\le c_\tau$, Assumption 2 only requires to hold with $\mathbb{S}$ in the $c_\tau$ neighborhood of $\tau_0$. As $\delta_0\neq 0$, (7.14) now includes an extra term

$$\begin{aligned}
&\kappa^2\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_2^2\le 2\widehat{\kappa}\left(\frac{2+\mu}{1-\mu},\mathbb{S},\widehat{\Sigma}\right)^2\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right|_2^2\le\frac{2}{n}\left|\boldsymbol{X}(\widehat{\tau})\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right|_2^2\\
&\le 2\left|\widehat{\boldsymbol{D}}\right|_\infty^2\left(\left\|\widehat{f}-f_0\right\|_n^2+2c_\alpha|\delta_0|_1\sup_j\frac{1}{n}\sum_{i=1}^n\left|X_i^{(j)}\right|^2|1(Q_i<\tau_0)-1(Q_i<\widehat{\tau})|\right)\\
&\le 2\left(C_2^2+\mu_1\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2+2C_5|\delta_0|_1c_\alpha c_\tau\right),
\end{aligned}$$

where the last inequality is due to conditioning on events $\mathbb{A}_1$ and Assumption 4. Combining this result with (7.24) yields

$$\begin{aligned}
\left\|\widehat{f}-f_0\right\|_n^2&\le 3\lambda\left|\widehat{\boldsymbol{D}}\left(\widehat{\alpha}-\alpha_0\right)_{J_0}\right|_1\le 3\lambda\sqrt{s_0}\left|\widehat{\boldsymbol{D}}\left(\widehat{\alpha}-\alpha_0\right)_{J_0}\right|_2\\
&\le 3\lambda\sqrt{s_0}\left(2\kappa^{-2}\left(C_2^2+\mu_1\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2+2C_5|\delta_0|_1c_\alpha c_\tau\right)\right)^{1/2}.
\end{aligned}$$

Applying $a+b\le 2a\vee 2b$, we get the upper bound of $\left\|\widehat{f}-f_0\right\|_n$ on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, and $\mathbb{A}_5$,

$$\left\|\widehat{f}-f_0\right\|_n^2\le\frac{36\left(C_2^2+\mu_1\lambda\right)}{\kappa^2}\lambda^2s_0\vee\frac{6\sqrt{2}\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{\kappa}\lambda\sqrt{s_0|\delta_0|_1}\left(c_\alpha c_\tau\right)^{1/2}. \tag{7.26}$$

We next derive the upper bound for $\|\widehat{\alpha}-\alpha_0\|_1$, using (7.25),

$$\min(\widehat{\boldsymbol{D}}) \left|\widehat{\alpha}-\alpha_0\right|_1 \leq \frac{3}{1-\mu}\sqrt{s_0}\left(2\kappa^{-2}\left(C_2^2+\mu_1\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2+2c_\alpha c_\tau C_5|\delta_0|_1\right)\right)^{1/2}$$
$$=\frac{3\sqrt{2}}{(1-\mu)\kappa}\sqrt{s_0}\left(\left(C_2^2+\mu_1\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2+2C_5|\delta_0|_1 c_\alpha c_\tau\right)\right)^{1/2},$$

where the last inequality is due to conditioning on $\mathbb{A}_3$. Then using the inequality that $a+b \leq 2a \vee 2b$ with (7.17) and (7.26) yields

$$\left|\widehat{\alpha}-\alpha_0\right|_1 \leq \frac{36}{(1-\mu)\kappa^2}\frac{\left(C_2^2+\mu_1\lambda\right)}{\sqrt{C_3^2-\mu_1\lambda}}\lambda s_0 \vee \frac{6\sqrt{2}}{(1-\mu)\kappa}\frac{\sqrt{C_2^2+\mu_1\lambda}\sqrt{C_5}}{\sqrt{C_3^2-\mu_1\lambda}}\sqrt{s_0|\delta_0|_1}\left(c_\alpha c_\tau\right)^{1/2}.$$

**Case (ii)**: In this case, (7.2) shows

$$\left\|\widehat{f}-f_0\right\|_n^2 \leq 3\lambda\left(\sqrt{c_\tau}+\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5|\delta_0|_1 c_\tau\right),$$
$$\left|\widehat{\alpha}-\alpha_0\right|_1 \leq \frac{3}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}\left(\sqrt{c_\tau}+\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5|\delta_0|_1 c_\tau\right),$$

which provides the result. □

We further tighten the bound for $|\widehat{\tau}-\tau_0|$ in the following lemma using Lemmas 10 and 12.

**Lemma 13.** *Suppose that* $|\widehat{\tau}-\tau_0| \leq c_\tau$ *and* $\left|\widehat{\alpha}-\alpha_0\right|_1 \leq c_\alpha$ *for some* $(c_\tau, c_\alpha)$. *Let* $\widetilde{\eta}=C_4^{-1}\lambda\left((1+\mu)\sqrt{C_2^2+\mu\lambda}c_\alpha+G_1\right)$. *If Assumption 3 holds, then conditional on the events* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, *and* $\mathbb{A}_4$, *we have,*

$$|\widehat{\tau}-\tau_0| \leq \widetilde{\eta}.$$

*Proof of Lemma 13.* The proof follows that of Lemma 13 in Lee et al. (2016). Conditioning on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ and (7.21), we derive

$$\left|\frac{2}{n}\sum_{i=1}^{n}\left[U_i X_i'\left(\widehat{\beta}-\beta_0\right)+U_i X_i' 1\left(Q_i<\widehat{\tau}\right)\left(\widehat{\delta}-\delta_0\right)\right]\right| \leq \mu\lambda\left(\sqrt{C_2^2+\mu_1\lambda}\right)c_\alpha,$$

and

$$\left|\frac{2}{n}\sum_{i=1}^{n}U_iX_i'\delta_0\left[1\left(Q_i<\widehat{\tau}\right)-1\left(Q_i<\tau_0\right)\right]\right|\leq\lambda\sqrt{c_\tau}.$$

Suppose $\widetilde{\eta}<|\widehat{\tau}-\tau_0|\leq c_\tau$. As in (7.18),

$$\widehat{S}_n-S_n(\alpha_0,\tau_0)\geq\left\|\widehat{f}-f_0\right\|_n^2-\mu\lambda\left(\sqrt{C_2^2+\mu_1\lambda}c_\alpha\right)-\lambda\sqrt{c_\tau}.$$

Additionally,

$$\begin{aligned}
&\left[\widehat{S}_n+\lambda\left|\widehat{\boldsymbol{D}}\widehat{\alpha}\right|_1\right]-\left[S_n(\alpha_0,\tau_0)+\lambda\left|\boldsymbol{D}\alpha_0\right|_1\right]\\
&\geq\left\|\widehat{f}-f_0\right\|_n^2-\mu\lambda\left(\sqrt{C_2^2+\mu_1\lambda}c_\alpha\right)-2|\delta_0|_1\lambda\sqrt{c_\tau}-\lambda\left(\left|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right|_1+\left|(\widehat{\boldsymbol{D}}-\boldsymbol{D})\alpha_0\right|_1\right)\\
&>C_4\widetilde{\eta}-\left((1+\mu)\left(\sqrt{C_2^2+\mu_1\lambda}c_\alpha\right)+G_1\right)\lambda,
\end{aligned}$$

where the last inequality is due to Assumption 3, Hölder's inequality and (7.23).

Since $C_4\widetilde{\eta}=\left((1+\mu)\sqrt{C_2^2+\mu_1\lambda}c_\alpha+G_1\right)\lambda$ by definitation, similarly to the proof of Lemma 10, contradiction yields the result. □

There are three different bounds for $|\alpha-\alpha_0|_1$ in Lemma 12 and the two terms $G_1$ and $G_3$ are functions of $c_\tau$ and $c_\alpha$. We thus apply Lemmas 12 and 13 iteratively to tighten up the bounds. We start the iteration with $c_\tau^{(0)}=\frac{2C_1(3+\mu_2)\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}}{C_4}s_0\lambda$ from the results of Lemma 10 and $c_\alpha^{(0)}=\frac{\left(2C_1(3+\mu_2)\right)\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}}{(1-\mu)\left(C_3^2-\mu_1\lambda\right)^{\frac{1}{2}}}s_0$ from (2.9) in Lemma 1.

**Lemma 14.** *Suppose that Assumptions 1 to 4 hold with* $\mathbb{S}=\{|\tau-\tau_0|\leq\eta^*\}$, $\kappa=\kappa\left(s_0,\frac{2+\mu}{1-\mu},\mathbb{S},\boldsymbol{\Sigma}\right)$ *for* $0<\mu<1$. *Let* $(\widehat{\alpha},\widehat{\tau})$ *be the Lasso estimator defined by* (2.4) *with* $\lambda$ *given by* (2.8). *In addition, there exists a sequence of constants* $\eta_1,...,\eta_{m^*}$ *for some finite* $m^*$. *With probability at least* $1-C(logn)^{-1}$, *we have*

$$
\begin{aligned}
&\left\|\widehat{f}-f_0\right\|_n^2 \le 3G_2\lambda^2 s_0,\\
&|\widehat{\alpha}-\alpha_0|_1 \le \frac{3}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}G_2\lambda s_0,\\
&|\widehat{\tau}-\tau_0| \le \left(\frac{3\,(1+\mu)\,\sqrt{(C_2^2+\mu_1\lambda)}}{(1-\mu)\sqrt{(C_3^2-\mu_1\lambda)}}+1\right)\frac{1}{C_4}G_2\lambda^2 s_0.
\end{aligned}
$$

*Proof of Lemma 14.* The proof follows that of Lemma 14 in Lee et al. (2016). The iteration to implement is as follows:

**Step 1**: Starting values $c_\tau^{(0)} = \frac{2C_1(3+\mu_2)\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}}{C_4}s_0\lambda$ and $c_\alpha^{(0)} = \frac{\left(2C_1(3+\mu_2)\right)\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}}{(1-\mu_1)\left(C_3^2-\mu\lambda\right)^{\frac{1}{2}}}s_0$.

**Step 2**: When $m \ge 1$, define

$$
\begin{aligned}
G_1^{(m-1)} &= \sqrt{c_\tau^{(m-1)}} + \left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1} C_5|\delta_0|_1 c_\tau^{(m-1)},\\
G_3^{(m-1)} &= \frac{2\sqrt{2}\,(C_2^2+\mu_1\lambda)^{\frac{1}{2}}\,\sqrt{C_5C_1}}{\kappa}\sqrt{c_\alpha^{(m-1)}c_\tau^{(m-1)}},\\
c_\alpha^{(m)} &= \frac{3}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}\cdot\left\{G_1^{(m-1)}\vee G_2\lambda s_0 \vee G_3^{(m-1)}\sqrt{s_0|\delta_0|_1}\right\},\\
c_\tau^{(m)} &= \frac{\lambda}{C_4}\left((1+\mu)\sqrt{C_2^2+\mu_1\lambda}c_\alpha^{(m)}+G_1^{(m-1)}\right).
\end{aligned}
$$

**Step 3**: We stop the iteration if $\left\{G_1^{(m)}\vee G_2\lambda s_0\vee G_3^{(m)}\sqrt{s_0|\delta_0|_1}\right\}$ keeps the same.

Suppose the rule in step 3 is met when $\left\{G_1^{(m)}\vee G_2\lambda s_0\vee G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}\right\} = G_2\lambda s_0$, then the bound is reached within $m^*$, a finite number of iterative applications. We have

$$
\begin{aligned}
c_\tau^{(m)} =& \frac{\lambda}{C_4}\left((1+\mu)\sqrt{C_2^2+\mu_1\lambda}c_\alpha^{(m)}+G_1^{(m-1)}\right) \ge \frac{\lambda}{C_4}\left(\frac{3\,(1+\mu)\,\sqrt{C_2^2+\mu_1\lambda}}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}G_2\lambda s_0+G_1^{(m-1)}\right)\\
\ge& \frac{1}{C_4}\left(\frac{3\,(1+\mu)\,\sqrt{C_2^2+\mu_1\lambda}}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}+\frac{G_1^{(m-1)}}{G_2\lambda s_0}\right)G_2\lambda^2 s_0 > \frac{1}{C_4}\left(\frac{3\,(1+\mu)\,\sqrt{C_2^2+\mu_1\lambda}}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}\right)G_2\lambda^2 s_0,
\end{aligned}
\tag{7.27}
$$

as $G_1^{(m-1)} > 0$, $G_2\lambda s_0 > 0$ and $c_\alpha^{(m)} \ge \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}G_2\lambda s_0$.

Note that (7.27) shows $c_\tau^{(m)} \geq Cs_0 \frac{\log p}{2n}$, which is a necessary condition to apply Lemma 10 through Lemma 13. Then $c_\alpha^{(m^*+1)}$ is the bound for $|\widehat{\alpha} - \alpha_0|_1$ . Then,

$$
\begin{aligned}
&c_\tau^{(m^*+1)} = \frac{\lambda}{C_4}\left((1+\mu)\sqrt{C_2^2+\mu_1\lambda}c_\alpha^{(m^*+1)} + G_1^{(m^*)}\right) \\
&\leq \frac{\lambda}{C_4}\left(\frac{3(1+\mu)\sqrt{C_2^2+\mu_1\lambda}}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}G_2\lambda s_0 + G_2\lambda s_0\right) = \left(\frac{3(1+\mu)\sqrt{(C_2^2+\mu_1\lambda)}}{(1-\mu)\sqrt{(C_3^2-\mu_1\lambda)}}+1\right)\frac{G_2}{C_4}\lambda^2 s_0,
\end{aligned}
$$

which is the bound for $|\widehat{\tau} - \tau_0|$.

Next, we turn to prove the existence of $m^*$. First, by induction, we can show that $G_1^{(m-1)}$, $G_1^{(m-1)}$, $c_\alpha^{(m)}$ and $c_\tau^{(m)}$ are decreasing as $m$ increases. We start the iteration with $c_\tau^{(0)}$ and $c_\alpha^{(0)}$ in step 1. In step 2, as long as $n$, $p$, $s_0$ and $|\delta_0|_1$ are large enough, we obtain [21]

$$
\begin{aligned}
G_1^{(0)} =& \sqrt{c_\tau^{(0)}} + \left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1} C_5|\delta_0|_1 c_\tau^{(0)} = \widetilde{C}\sqrt{s_0\lambda} + \widetilde{C}|\delta_0|_1 s_0\lambda, \\
G_3^{(0)} =& \frac{2\sqrt{2}\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{\kappa}\sqrt{c_\alpha^{(0)}}\sqrt{c_\tau^{(0)}} = \widetilde{C}\sqrt{s_0^2\lambda},
\end{aligned}
$$

Then, as $|\delta_0|_1 s_0\lambda = o_p(1)$,

$$
\left\{G_1^{(0)} \vee G_2\lambda s_0 \vee G_3^{(0)}\sqrt{s_0|\delta_0|_1}\right\} = G_3^{(0)}\sqrt{s_0|\delta_0|_1}.
$$

We derive

$$
\begin{aligned}
c_\alpha^{(1)} =& \frac{3}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}} \cdot \left\{G_1^{(0)} \vee G_2\lambda s_0 \vee G_3^{(0)}\sqrt{s_0\|\delta_0\|_1}\right\} = \widetilde{C}s_0\sqrt{s_0\|\delta_0\|_1\lambda}, \\
c_\tau^{(1)} =& \frac{\lambda}{C_4}\left((1+\mu)\sqrt{C_2^2+\mu_1\lambda}c_\alpha^{(1)} + G_1^{(0)}\right) = \widetilde{C}s_0\lambda\sqrt{s_0\|\delta_0\|_1\lambda} + \widetilde{C}\lambda\sqrt{s_0\lambda} + \widetilde{C}\|\delta_0\|_1 s_0\lambda^2.
\end{aligned}
$$

Thus, we have $c_\alpha^{(0)} > c_\alpha^{(1)}$ and $c_\tau^{(0)} > c_\tau^{(1)}$. If we assume $c_\alpha^{(m)} > c_\alpha^{(m+1)}$ and $c_\tau^{(m)} > c_\tau^{(m+1)}$, it is easy to show $G_1^{(m)} > G_1^{(m+1)}$ and $G_3^{(m)} > G_3^{(m+1)}$, then $c_\alpha^{(m+1)} > c_\alpha^{(m+2)}$ and $c_\tau^{(m+1)} > c_\tau^{(m+2)}$, which means that applying the iteration can tighten up the bounds.

We then use proof by contradiction method to show that there exists an $m^*$ such

[21] $\widetilde{C}$ is positive, finite and varies for each term.

that $\left\{G_1^{(m^*)} \vee G_2\lambda s_0 \vee G_3^{(m^*)}\sqrt{s_0|\delta_0|_1}\right\} = G_2\lambda s_0$. Suppose for all $m>1$, $\left\{G_1^{(m)} \vee G_3^{(m)}\sqrt{s_0|\delta_0|_1}\right\}$ $> G_2\lambda s_0$. As $G_1^{(m-1)}$ and $G_3^{(m-1)}$ are decreasing as $m$ increases, and $\left\{G_1^{(m)} \vee G_3^{(m)}\sqrt{s_0|\delta_0|_1}\right\}$ is bounded, we consider the following two cases:

**Case (1)**: For sufficiently large $m$, assume $G_1^{(m)} \le G_3^{(m)}\sqrt{s_0|\delta_0|_1}$. Let $G_3^{(m)}$ converge to $G_3^{(\infty)}$ and $G_3^{(\infty)} > G_2\lambda s_0$. We have

$$c_\alpha^{(\infty)} = \frac{3}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}G_3\sqrt{s_0\|\delta_0\|_1} =: H_1\sqrt{s_0\|\delta_0\|_1}\sqrt{c_\alpha^{(\infty)}}\sqrt{c_\tau^{(\infty)}},$$

where $H_1 = \frac{6\sqrt{2}\left(C_2^2+\mu_1\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda\kappa}}$, then $c_\alpha^{(\infty)} = H_1^2 s_0|\delta_0|_1 c_\tau^\infty$; and

$$\begin{aligned} c_\tau^\infty =& C_4^{-1}\lambda\left((1+\mu)\sqrt{(C_2^2+\mu_1\lambda)c_\alpha^\infty} + \sqrt{c_\tau^\infty} + \left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5|\delta_0|_1 c_\tau^\infty\right) \\ =& C_4^{-1}(1+\mu)\sqrt{(C_2^2+\mu_1\lambda)}\lambda c_\alpha^\infty + C_4^{-1}\lambda\sqrt{c_\tau^\infty} + C_4^{-1}\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5|\delta_0|_1\lambda c_\tau^\infty \\ =:& H_2\lambda c_\alpha^\infty + H_3\lambda\sqrt{c_\tau^\infty} + H_4|\delta_0|_1\lambda c_\tau^\infty, \end{aligned}$$

where $H_2 = C_4^{-1}(1+\mu)\sqrt{(C_2^2+\mu_1\lambda)}$, $H_3 = C_4^{-1}$ and $H_4 = C_4^{-1}\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5$.

To solve the above equation system, as $n$, $p$ are sufficiently large, $\sqrt{C_3^2-\mu_1\lambda}$ and $\sqrt{C_2^2+\mu_1\lambda}$ converge to constants; $s_0\|\delta\|_1\lambda$ and $\|\delta_0\|_1\lambda$ converge to 0. Therefore,

$$\begin{aligned} c_\tau^\infty =& \left(\frac{H_1^2H_2s_0|\delta|_1\lambda^2 + H_3\lambda}{1-H_1^2H_2s_0|\delta|_1\lambda - H_4\lambda|\delta|_1}\right)^2 = O_p(\lambda^2), \\ c_\alpha^{(\infty)} =& H_1^2 s_0|\delta_0|_1 c_\tau^\infty = O_p(s_0\|\delta_0\|_1\lambda^2). \end{aligned}$$

Then,

$$G_3^{(\infty)}\sqrt{s_0|\delta_0|} = \frac{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}{3}c_\alpha^{(\infty)} = O_p(s_0|\delta_0|_1\lambda^2).$$

Obviously, it leads to contradiction, because $c_\tau^\infty < s_0\lambda^2$ and $G_3^{(\infty)}\sqrt{s_0|\delta_0|_1} < G_2\lambda s_0$.

**Case (2)**: For sufficiently large $m$, assume $G_1^{(m)} > G_3^{(m)}\sqrt{s_0|\delta_0|_1}$. Let $G_1^{(m)}$ converge to $G_1^{(\infty)}$ and $G_1^{(\infty)} > G_2\lambda s_0$. We have

$$c_\alpha^{(\infty)} = G_1\frac{3}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}},$$

$$
\begin{aligned}
c_\tau^{(\infty)} &= C_4^{-1}\lambda\left((1+\mu)\sqrt{(C_2^2+\mu_1\lambda)}c_\alpha^{(\infty)}+G_1^{(\infty)}\right)\\
&= C_4^{-1}\lambda\left((1+\mu)\sqrt{(C_2^2+\mu_1\lambda)}\frac{3}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}+1\right)G_1^{(\infty)}\\
&= C_4^{-1}\left(\frac{3\,(1+\mu)\sqrt{(C_2^2+\mu_1\lambda)}}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}+1\right)\lambda\sqrt{c_\tau^{(\infty)}}\\
&+ C_4^{-1}\left(\frac{3\,(1+\mu)\sqrt{(C_2^2+\mu_1\lambda)}}{(1-\mu)\sqrt{C_3^2-\mu_1\lambda}}+1\right)\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5\|\delta_0\|_1\lambda c_\tau^{(\infty)}\\
&=: H_5\lambda\sqrt{c_\tau^{(\infty)}}+H_6|\delta_0|_1\lambda c_\tau^{(\infty)},
\end{aligned}
$$

where $H_5$ and $H_6$ are defined accordingly. Furthermore, as $n$, $p$ are sufficiently large, $\sqrt{C_3^2-\mu_1\lambda}$ and $\sqrt{C_2^2+\mu_1\lambda}$ converge to constants , $|\delta_0|_1\lambda$ converges to 0. Therefore,

$$
c_\tau^\infty = \left(\frac{H_5\lambda}{1-H_6|\delta_0|_1\lambda}\right)^2 = O_p(\lambda^2).
$$

Then

$$
\begin{aligned}
G_1^{(\infty)} &= \left(1+\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}\lambda\|\delta_0\|_1C_5\right)\sqrt{c_\tau^{(\infty)}}+\left(2\sqrt{C_3^2-\mu_1\lambda}\right)^{-1}C_5\|\delta_0\|_1c_\tau^{(\infty)}\\
&= O_p(\lambda+\lambda^2),
\end{aligned}
$$

which leads to the contradiction because $c_\tau^\infty < s_0\lambda^2$ and $G_1^{(\infty)} < G_2\lambda s_0$.

Finally, Lemma 12 yields $\left\|\widehat{f}-f_0\right\|_n^2 \leq 3G_2\lambda^2 s_0$. □

*Proof of Theorem 2.* The proof follows immediately from Lemma 14 under Assumptions 1 to 5. Specially, $P(\mathbb{A}_1\cap\mathbb{A}_2\cap\mathbb{A}_3\cap\mathbb{A}_4\cap\mathbb{A}_5)\geq 1-C(logn)^{-1}$. □

## 7.4 Proofs for the Asymptotic Properties of Nodewise Regression Estimator

The proof is similar to that of Lemma A.9 in the Appendix of Caner and Kock (2018). Define the following events:

$$\mathbb{A}_{node} = \left\{ \max_{j+p\in H} \sup_{\tau\in\mathbb{T}} \left| X^{(-j)}(\tau)'\upsilon^{(j)}/n \right|_\infty \leq \frac{\mu\lambda_{node}}{2} \right\},$$

$$\mathbb{A}_{EV}^{(j)} = \left\{ \frac{\kappa\left(s_j, c_0, \mathbb{T}, M_{-j,-j}\right)^2}{2} \leq \widehat{\kappa}\left(s_j, c_0, \mathbb{T}, \widehat{M}_{-j,-j}\right)^2 \right\},$$

$$\mathbb{B}_{node} = \left\{ \max_{j\in H \text{or} j+p\in H} \sup_{\tau\in\mathbb{T}} \left| \widetilde{X}^{(-j)}(\tau)'\widetilde{\upsilon}^{(j)}/n \right|_\infty \leq \frac{\mu\lambda_{node}}{2} \right\},$$

$$\mathbb{B}_{EV}^{(j)} = \left\{ \frac{\kappa\left(s_j, c_0, \mathbb{T}, N_{-j,-j}\right)^2}{2} \leq \widehat{\kappa}\left(s_j, c_0, \mathbb{T}, \widehat{N}_{-j,-j}\right)^2 \right\}.$$

**Lemma 15.** *Suppose that Assumptions 1-6 hold and that $\widehat{\delta}(\widehat{\tau}) \neq 0$ via (2.4). Set $\lambda_{node} = \frac{C}{\mu}\sqrt{\frac{\log p}{n}}$. Then*

$$P\left(\mathbb{A}_{node} \cap \left(\cap_{j+p\in H}\mathbb{A}_{EV}^{(j)}\right) \cap \mathbb{B}_{node} \cap \left(\cap_{j\in H or j+p\in H}\mathbb{B}_{EV}^{(j)}\right)\right) \geq 1 - C(logn)^{-1}.$$

*Proof of Lemma 15.* We start with $\mathbb{A}_{node}$, $\max_{j+p\in H}\sup_{\tau\in\mathbb{T}} \|X^{(-j)}(\tau)'\upsilon^{(j)}/n\|_\infty \leq \frac{\mu\lambda_{node}}{2}$ is equivalent to $\max_{j+p\in H}\max_{1\leq l\leq p-1}\sup_{\tau\in\mathbb{T}} \frac{1}{n}\sum_{i=1}^{n} X_i^{(-j,l)}(\tau)\upsilon_i^{(j)} \leq \frac{\mu\lambda_{node}}{2}$. Then sort $\{X_i, \upsilon_i, Q_i\}_{i=1}^n$ by $(Q_1,\ldots,Q_n)$ in ascending order, it is equivalent to $\max_{j+p\in H}\max_{1\leq k\leq n}$ $\max_{1\leq l\leq p-1}\frac{1}{n}\sum_{i=1}^{k} X_i^{(-j,l)}\upsilon_i^{(j)} \leq \frac{\mu\lambda_{node}}{2}$. Following directly from the proof of Lemma 5, we obtain that $\mathbb{A}_{node}$ holds with probability at least $1 - C(logn)^{-1}$.

Similarly, consider the transpose of $\Xi_{n,n}$ in (7.1) and let $\widetilde{\xi}_i^{(l)}$ be the element in the $i$-th row and $l-th$ column of the transpose of $\Xi_{n,n}$, we can then obtain that $\mathbb{B}_{node}$ holds with probability at least $1 - C(logn)^{-1}$.

Next, for each $j+p\in H$, by Lemma 8

$$(1+c_0)^2 s_j \sup_{\tau\in\mathbb{T}} \|\widehat{M}_{-j,-j}(\tau) - M_{-j,-j}(\tau)\|_\infty \leq (1+c_0)^2 \bar{s} \sup_{\tau\in\mathbb{T}} \|\widehat{M}(\tau) - M(\tau)\|_\infty$$
$$\leq \frac{\kappa(\bar{s}, c_0, \mathbb{T}, \boldsymbol{M})}{2} \leq \frac{\kappa(s_j, c_0, \mathbb{T}, \boldsymbol{M})}{2}$$

implies that

$$\left\{(1+c_0)^2 s_j \sup_{\tau\in\mathbb{T}} \|\widehat{M}_{-j,-j}(\tau) - M_{-j,-j}(\tau)\|_\infty \leq \frac{\kappa(s_j, c_0, \mathbb{T}, \boldsymbol{M})}{2}\right\} \subset \mathbb{A}_{EV}^{(j)}.$$

Then we have,

$$\left\{(1+c_0)^2\bar{s}\sup_{\tau\in\mathbb{T}}\|\widehat{M}(\tau)-M(\tau)\|_\infty\le\frac{\kappa(\bar{s},c_0,\mathbb{T},\boldsymbol{M})}{2}\right\}\subset\cap_{j+p\in H}\mathbb{A}_{EV}^{(j)}.$$

We thus obtain, by Lemma 8, that $\cap_{j+p\in H}\mathbb{A}_{EV}^{(j)}$ holds with probability at least $1-C(logn)^{-1}$ provided that $\kappa(s_j,c_0,\mathbb{T},M)>0$. Similarly, we can show that $\cap_{j\in H\text{or}j+p\in H}\mathbb{B}_{EV}^{(j)}$ holds with probability at least $1-C(logn)^{-1}$.

Therefore, by $P(A\cap B)\ge 1-P(A^c)-P(B_2^c)$, we derive

$$P\left(\mathbb{A}_{node}\cap\left(\cap_{j+p\in H}\mathbb{A}_{EV}^{(j)}\right)\cap\mathbb{B}_{node}\cap\left(\cap_{j\in H\text{or}j+p\in H}\mathbb{B}_{EV}^{(j)}\right)\right)\ge 1-C(logn)^{-1}.$$

□

**Lemma 16.** *Suppose that Assumptions 1-6 hold and set* $\lambda_{node}=\frac{C}{\mu}\sqrt{\frac{\log p}{n}}$. *Then,*

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)-\Theta_j(\tau)\right|_1=O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right)\tag{7.28}$$

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)-\Theta_j(\tau)\right|_2=O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right)\tag{7.29}$$

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)\right|_1=O_p\left(\sqrt{\bar{s}}\right)\tag{7.30}$$

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)'\widehat{\boldsymbol{\Sigma}}(\tau)-e_j'\right|_\infty=O_p\left(\sqrt{\frac{\log p}{n}}\right)\tag{7.31}$$

*Proof of Lemma 16.* Given $\forall\,\tau\in\mathbb{T}$ and each $j\in H$ or $j+p\in H$, (3.6) is a loss function for linear model, the pointwise oracle inequalities for linear model have been proved in Theorem 2.4 of van de Geer et al. (2014). Since the uniform oracle inequalities only involve the noise conditions $\mathbb{A}_{node}$ and $\mathbb{B}_{node}$, and adaptive restricted eigenvalue conditions $\cap_{j+p\in H}\mathbb{A}_{EV}^{(j)}$ and $\cap_{j\in H\text{or}j+p\in H}\mathbb{B}_{EV}^{(j)}$, by Lemma 15, we obtain that the following results hold uniformly in $\mathbb{T}$ and $H$,

$$\sup_{\tau\in\mathbb{T}}\max_{j+p\in H}|X^{(-j)}(\tau)'\gamma_j(\tau)-X^{(-j)}(\tau)'\widehat{\gamma}_j(\tau)|_n\le\frac{C}{\kappa(\bar{s},c_0,\mathbb{T},\boldsymbol{M})}\sqrt{\bar{s}},\lambda_{node}\tag{7.32}$$

$$\sup_{\tau\in\mathbb{T}}\max_{j+p\in H}|\gamma_j(\tau)-\widehat{\gamma}_j(\tau)|_1\le\frac{C}{\kappa(\bar{s},c_0,\mathbb{T},\boldsymbol{M})^2},\bar{s},\lambda_{node}\tag{7.33}$$

with probability at least $1 - (logn)^{-1}$.

In line with the inequalities presented in Lemma A.9 in the Appendix of Caner and Kock (2018), we can thus establish the following set of inequalities:

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{A}_j(\tau) - A_j(\tau)\right|_1 = O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{7.34}$$

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{A}_j(\tau) - A_j(\tau)\right|_2 = O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{7.35}$$

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{A}_j(\tau)\right|_1 = O_p\left(\sqrt{\bar{s}}\right) \tag{7.36}$$

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\frac{1}{\widehat{z}_j(\tau)^2} = O_p\left(1\right) \tag{7.37}$$

$$\max_{j\in H \mathrm{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{B}_j(\tau) - B_j(\tau)\right|_1 = O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{7.38}$$

$$\max_{j\in H \mathrm{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{B}_j(\tau) - B_j(\tau)\right|_2 = O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{7.39}$$

$$\max_{j\in H \mathrm{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{B}_j(\tau)\right|_1 = O_p\left(\sqrt{\bar{s}}\right) \tag{7.40}$$

$$\max_{j\in H \mathrm{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\frac{1}{\widehat{\widehat{z}}_j(\tau)^2} = O_p\left(1\right) \tag{7.41}$$

Now consider (2.11) and (3.13),

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau) - \Theta_j(\tau)\right|_1 \leq$$
$$\max_{j\in H \mathrm{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\max\left\{2\left|\widehat{B}_j(\tau) - B_j(\tau)\right|_1, 2\left|\widehat{B}_j(\tau) - B_j(\tau)\right|_1 + \left|\widehat{A}_j(\tau) - A_j(\tau)\right|_1\right\},$$

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau) - \Theta_j(\tau)\right|_2 \leq$$
$$\max_{j\in H \mathrm{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\max\left\{2\left|\widehat{B}_j(\tau) - B_j(\tau)\right|_2, 2\left|\widehat{B}_j(\tau) - B_j(\tau)\right|_2 + \left|\widehat{A}_j(\tau) - A_j(\tau)\right|_2\right\},$$

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)\right|_1 \leq \max_{j\in H \mathrm{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\max\left\{2\left|\widehat{B}_j(\tau)\right|_1, 2\left|\widehat{B}_j(\tau)\right|_1 + \left|\widehat{A}_j(\tau)\right|_1\right\}.$$

We thus have proved the first 3 inequalities in Lemma 16.

Next, we will bound $\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)'\widehat{\boldsymbol{\Sigma}}(\tau) - e_j'\right|_\infty$. We can show that $\widehat{\boldsymbol{A}}(\tau)$

is an approximate inverse matrix of $\widehat{\boldsymbol{M}}(\tau)$. Let $\widehat{A}_j(\tau)$ denote the $j$-th row of $\widehat{\boldsymbol{A}}(\tau)$, we then have $\widehat{A_j}(\tau) = \widehat{C_j}(\tau)/\widehat{z_j}(\tau)^2$. Denoting by $\widetilde{e}_j$ the $j$-th unit vector, the KKT conditions imply that

$$\left|\widehat{A_j}(\tau)'\widehat{\boldsymbol{M}}(\tau) - \widetilde{e}_j'\right|_\infty \leq \left|\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}(\tau)\right| \frac{\lambda_{node}}{\widehat{z_j}(\tau)^2}. \tag{7.42}$$

Similarly, we have

$$\left|\widehat{B}_j(\tau)'\widehat{\boldsymbol{N}}(\tau) - \widetilde{e}_j'\right|_\infty \leq \left|\widehat{\widehat{\boldsymbol{\Gamma}}}_{\boldsymbol{j}}(\tau)\right| \frac{\lambda_{node}}{\widehat{\widehat{z}}_j(\tau)^2}. \tag{7.43}$$

Therefore, we obtain

$$\max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)'\widehat{\boldsymbol{\Sigma}}(\tau) - e_j'\right|_\infty = \max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\left|\begin{bmatrix}\widehat{B}_j(\tau) & -\widehat{B}_j(\tau)\end{bmatrix}\begin{bmatrix}\widehat{\boldsymbol{M}} & \widehat{\boldsymbol{M}}(\tau)\\ \widehat{\boldsymbol{M}}(\tau) & \widehat{\boldsymbol{M}}(\tau)\end{bmatrix} - e_j'\right|_\infty$$
$$= \max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\left|\begin{bmatrix}\widehat{B}_j(\tau)\widehat{\boldsymbol{N}}(\tau) & 0\end{bmatrix} - e_j'\right|_\infty \leq \max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\left|\widehat{B}_j(\tau)'\widehat{\boldsymbol{N}}(\tau) - \widetilde{e}_j'\right|_\infty \leq \max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\frac{\lambda_{node}}{\widehat{\widehat{z}}_j(\tau)^2}.$$

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\tau)'\widehat{\boldsymbol{\Sigma}}(\tau) - e_j'\right|_\infty = \max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\begin{bmatrix}-\widehat{B}_j(\tau) & \widehat{B}_j(\tau)+\widehat{A}_j(\tau)\end{bmatrix}\begin{bmatrix}\widehat{\boldsymbol{M}} & \widehat{\boldsymbol{M}}(\tau)\\ \widehat{\boldsymbol{M}}(\tau) & \widehat{\boldsymbol{M}}(\tau)\end{bmatrix} - e_j'\right|_\infty$$
$$= \max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\left|\begin{bmatrix}\widehat{A}_j(\tau)\widehat{\boldsymbol{M}}(\tau) - \widehat{B}_j(\tau)\widehat{\boldsymbol{N}}(\tau) & \widehat{A}_j(\tau)\widehat{\boldsymbol{M}}(\tau)\end{bmatrix} - \begin{bmatrix}0 & \widetilde{e}_j'\end{bmatrix}\right|_\infty$$
$$\leq \max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\max\left\{\left|\widehat{A}_j(\tau)'\widehat{\boldsymbol{M}}(\tau) - \widetilde{e}_j'\right|_\infty + \left|\widehat{B}_j(\tau)'\widehat{\boldsymbol{N}}(\tau) - \widetilde{e}_j'\right|_\infty, \left|\widehat{A}_j(\tau)'\widehat{\boldsymbol{M}}(\tau) - \widetilde{e}_j'\right|_\infty\right\}$$
$$\leq \max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\frac{\lambda_{node}}{\widehat{z_j}(\tau)^2} + \frac{\lambda_{node}}{\widehat{\widehat{z}}_j(\tau)^2}.$$

□

## 7.5 Proofs for Theorem 3

### 7.5.1 No Threshold Effect

We first prove the case with no threshold effect, i.e. the true model is linear.

To show that the ratio

$$t = \frac{\sqrt{n} g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g' \widehat{\mathbf{\Theta}}(\widehat{\tau}) \widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) \widehat{\mathbf{\Theta}}(\widehat{\tau})' g}} \tag{7.44}$$

is asymptotically standard normal. First, by (3.3), we have $t = t_1 + t_2$, where

$$t_1 = \frac{g' \widehat{\mathbf{\Theta}}(\widehat{\tau}) \boldsymbol{X}(\widehat{\tau})' U / n^{1/2}}{\sqrt{g' \widehat{\mathbf{\Theta}}(\widehat{\tau}) \widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) \widehat{\mathbf{\Theta}}(\widehat{\tau})' g}}, \text{ and } t_2 = \frac{g' \Delta(\widehat{\tau})}{\sqrt{g' \widehat{\mathbf{\Theta}}(\widehat{\tau}) \widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) \widehat{\mathbf{\Theta}}(\widehat{\tau})' g}},$$

which suffices to show that $t_1$ is asymptotically standard normal and $t_2 = o_p(1)$.

**Lemma 17.** *Suppose that Assumptions 1, 2, 6 and 7 hold, conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, we have $g' \Delta(\widehat{\tau}) = O_p\left(\frac{s_0 \sqrt{h} \log p}{\sqrt{n}}\right)$.*

*Proof of Lemma 17.* By Hölder's inequality, Theorem 1, and Lemma 16, we obtain

$$\begin{aligned}
g' \Delta(\widehat{\tau}) &\le \max_{j \in H} |\Delta_j(\widehat{\tau})| \sum_{j \in H} |g_j| = \max_{j \in H} \left| \left( \widehat{\Theta}_j(\widehat{\tau}) \widehat{\mathbf{\Sigma}}(\widehat{\tau}) - e_j' \right) \sqrt{n} (\widehat{\alpha}(\widehat{\tau}) - \alpha_0) \right| \sum_{j \in H} |g_j| \\
&\le \max_{1 \le j \le 2p} \left| \widehat{\Theta}_j(\widehat{\tau}) \widehat{\mathbf{\Sigma}}(\widehat{\tau}) - e_j' \right|_\infty \sqrt{n} \left| \widehat{\alpha}(\widehat{\tau}) - \alpha_0 \right|_1 \sum_{j \in H} |g_j| \\
&\le C \left( \frac{\lambda_{node}}{\widehat{z}_{1j}^2(\widehat{\tau})} + \frac{\lambda_{node}}{\widehat{z}_{2j}^2(\widehat{\tau})} \right) \cdot \sqrt{n} \cdot \lambda s_0 \sqrt{h} = O_p\left( \frac{s_0 \sqrt{h} \log p}{\sqrt{n}} \right).
\end{aligned}$$

□

**Lemma 18.** *Suppose that Assumption 7 hold, then*

$$\max_{1 \le k,l,j \le p} \left| \frac{1}{n} \sum_{i=1}^n \left( X_i^{(k)} X_i^{(l)} X_i^{(j)} \right)^2 - \frac{1}{n} \sum_{i=1}^n E\left[ \left( X_i^{(k)} X_i^{(l)} X_i^{(j)} \right)^2 \right] \right| = O_p\left( \sqrt{\frac{\log p}{n}} \right),$$

$$\max_{1 \le k,l \le p} \left| \frac{1}{n} \sum_{i=1}^n \left( X_i^{(k)} X_i^{(l)} U_i \right)^2 - \frac{1}{n} \sum_{i=1}^n E\left[ \left( X_i^{(k)} X_i^{(l)} U_i \right)^2 \right] \right| = O_p\left( \sqrt{\frac{\log p}{n}} \right),$$

$$\max_{1 \le l,k \le 2p} \sup_{\tau \in \mathbb{T}} \left| \frac{1}{n} \sum_{i=1}^n \boldsymbol{X}_i^{(k)}(\tau) \boldsymbol{X}_i^{(l)}(\tau) U_i^2 - \frac{1}{n} \sum_{i=1}^n E\left[ \boldsymbol{X}_i^{(k)}(\tau) \boldsymbol{X}_i^{(l)}(\tau) U_i^2 \right] \right| = O_p\left( \frac{\sqrt{\log p}}{\sqrt{n}} \right).$$

*Proof of Lemma 18.* Under Assumption 7, by applying Lemmas 2 and 3, we can obtain the results using similar proofs as in Lemmas 5 and 7; therefore, the proof is omitted. □

**Lemma 19.** *Suppose that Assumptions 1 to 7 hold, conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, then we have*

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g\right| = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).$$

*Proof of Lemma 19.* Recall no threshold effect case, we have $\mathbf{\Sigma}_{xu}(\widehat{\tau})$ $= E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i'(\widehat{\tau})U_i^2\right]$, $\widehat{U}_i(\widehat{\tau}) = Y_i - \boldsymbol{X}_i'(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}) = U_i + \boldsymbol{X}_i(\widehat{\tau})'\alpha_0 - \boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})$, $\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{U}_i(\widehat{\tau})^2$, and define $\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2$. Then we will follow the proof (part b) of Theorem 2 in Caner and Kock (2018) to derive our results. First, to prove this lemma, we need to prove the followings, as (A.62), (A.63) and (A.64) in Caner and Kock (2018),

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| = o_p(1), \tag{7.45}$$

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| = o_p(1), \tag{7.46}$$

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g\right| = o_p(1). \tag{7.47}$$

To prove (7.45), we write

$$\begin{aligned}&\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| \\ &\leq \left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\left(\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) - \widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\right)\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| \leq \left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\right|_1^2 \|\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) - \widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\|_\infty.\end{aligned}$$

Then we obtain,

$$\begin{aligned}\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) - \widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) &= \frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{U}_i^2(\widehat{\tau}) - \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2\right) \\ =&\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'(U_i + \boldsymbol{X}_i(\widehat{\tau})'\alpha_0 - \boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau}))^2 - \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2\right)\end{aligned}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\alpha_0'\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\alpha_0\right)+\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}'(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)$$
$$-\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\alpha_0'\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)+\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\alpha_0'\boldsymbol{X}_i(\widehat{\tau})U_i\right)$$
$$-\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})'\boldsymbol{X}_i(\widehat{\tau})U_i\right)$$
$$=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\alpha_0\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\left(\alpha_0-\widehat{\alpha}(\widehat{\tau})\right)+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\left(\widehat{\alpha}(\widehat{\tau})'-\alpha_0'\right)\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})$$
$$+\frac{2}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\left(\alpha_0'-\widehat{\alpha}(\widehat{\tau})'\right)\boldsymbol{X}_i(\widehat{\tau})U_i.$$

By Cauchy-Schwarz inequality and Hölder's inequality

$$\max_{1\le k,l\le 2p}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\alpha_0'\boldsymbol{X}_i(\widehat{\tau})\left(\boldsymbol{X}_i(\widehat{\tau})'\alpha_0-\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)\right)\right|$$
$$\le\sqrt{\max_{1\le k,l\le 2p}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)^2\left(\boldsymbol{X}_i(\widehat{\tau})'\alpha_0\right)^2}\|\boldsymbol{X}(\widehat{\tau})\alpha_0-\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_n$$
$$\le\sqrt{\max_{1\le k,l\le 2p}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)^2\left(\max_{1\le k\le 2p}\boldsymbol{X}_i^{(k)}(\tau_0)\right)^2|\alpha_0|_1^2}\|\boldsymbol{X}(\widehat{\tau})\alpha_0-\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_n$$
$$\le\sqrt{\max_{1\le k,l,j\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(k)}X_i^{(l)}X_i^{(j)}\right)^2|\alpha_0|_1^2\,\mathbf{1}\left(Q_i<\widehat{\tau}\right)}\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\widehat{\tau})\alpha_0\|_n=O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right),$$

the last equality follows from Lemma 18, and $|\widehat{\alpha}(\widehat{\tau})|_1\le|\alpha_0|_1+O_p\left(s_0\sqrt{\frac{\log p}{n}}\right)$, $\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\widehat{\tau})\alpha_0\|_n=O_p\left(\sqrt{s_0}\sqrt{\frac{\log p}{n}}\right)$ by Theorem 1, and $|\alpha_0|_1=O_p(s_0)$ under Assumption 1. Also, we have

$$\max_{1\le k,l\le 2p}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\left(\widehat{\alpha}(\widehat{\tau})'-\alpha_0'\right)\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)\right|$$

$$\le \max_{1\le k,l\le 2p}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)^2\left(\widehat{\alpha}'(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})\right)^2}\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\widehat{\tau})\alpha_0\|_n$$

$$\le\sqrt{\max_{1\le k,l,j\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(k)}X_i^{(l)}X_i^{(j)}\right)^2|\widehat{\alpha}(\widehat{\tau})|_1^2\,\mathbf{1}\left(Q_i<\widehat{\tau}\right)}\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\widehat{\tau})\alpha_0\|_n=O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right),$$

and

$$\max_{1\le k,l\le 2p}\left|\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})'\alpha_0-\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)U_i\right|$$

$$\le 2\sqrt{\max_{1\le k,l\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(k)}X_i^{(l)}U_i\right)^2\mathbf{1}\left(Q_i<\widehat{\tau}\right)}\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\widehat{\tau})\alpha_0\|_n=O_p\left(\sqrt{s_0}\sqrt{\frac{\log p}{n}}\right).$$

We then obtain

$$\left\|\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})-\widetilde{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\right\|_\infty=O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).$$

Therefore,

$$\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g-g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widetilde{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g\right|\le\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\right|_1^2\|\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})-\widetilde{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\|_\infty$$

$$\le\left(\sum_{j\in H}|g_j|\max_{j\in H}\sup_{\tau\in\mathbb{T}}\left\|\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\right\|_1\right)^2\left\|\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})-\widetilde{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\right\|_\infty$$

$$=O_p\left(h\bar{s}\right)O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right)=O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).$$

To prove (7.46), we have

$$\widetilde{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})-\boldsymbol{\Sigma}_{xu}(\widehat{\tau})=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2-\frac{1}{n}\sum_{i=1}^{n}E\left[\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2\right].$$

We thus derive

$$
\begin{aligned}
&\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| \leq \left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\left(\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) - \mathbf{\Sigma}_{xu}(\widehat{\tau})\right)\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| \\
\leq & \left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\right|_1^2 \left\|\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) - \mathbf{\Sigma}_{xu}(\widehat{\tau})\right\|_\infty \leq \left(\sum_{j\in H}|g_j| \max_{j\in H}\sup_{\tau\in\mathbb{T}}\left|\widehat{\Theta}_j(\widehat{\tau})\right|_1\right)^2 \left\|\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) - \mathbf{\Sigma}_{xu}(\widehat{\tau})\right\|_\infty \\
\leq & O_p\left(h\bar{s}\right)O_p\left(\sqrt{\frac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{\frac{\log p}{n}}\right).
\end{aligned}
$$

To prove (7.47), we write

$$
\begin{aligned}
&\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g\right| \\
\leq & |\mathbf{\Sigma}_{xu}(\widehat{\tau})\|_\infty \left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_1^2 + 2\left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_2 |\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g|_2 \\
= & \|\mathbf{\Sigma}_{xu}(\widehat{\tau})\|_\infty \left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_1^2 + 2\widetilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \mathbf{\Sigma}_{\boldsymbol{xu}})\left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_2 |\mathbf{\Theta}(\widehat{\tau})'g|_2 \\
\leq & \|\mathbf{\Sigma}_{xu}(\widehat{\tau})\|_\infty \left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_1^2 + 2\widetilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \mathbf{\Sigma}_{\boldsymbol{xu}})\left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_2 \widetilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \mathbf{\Theta})\,|g|_2 .
\end{aligned}
$$

As $\|\mathbf{\Sigma}_{xu}(\widehat{\tau})\|_\infty = \max_{1\leq l,k\leq 2p} E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})u_i^2\right]$, $\widetilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \mathbf{\Sigma}_{\boldsymbol{xu}})$ and $\widetilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \mathbf{\Theta})$ are bounded under Assumption 7, we obtain

$$
\begin{aligned}
&\left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_1 = \sum_{j\in H}\left(|g_j|\left|\widehat{\Theta}_j(\widehat{\tau}) - \Theta_j(\widehat{\tau})\right|_1\right) \leq \sum_{j\in H}|g_j|\sup_{\tau\in\mathbb{T}}\max_{j\in H}\left|\widehat{\Theta}_j(\tau) - \Theta_j(\tau)\right|_1 \\
\leq & \sqrt{h}\sup_{\tau\in\mathbb{T}}\max_{j\in H}|\Theta_j(\tau) - \Theta_j(\tau)|_1 = O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\left(\widehat{\mathbf{\Theta}}(\widehat{\tau}) - \mathbf{\Theta}(\widehat{\tau})\right)'g\right|_2 = \left|\sum_{j\in H}\left(\Theta_j(\widehat{\tau}) - \Theta_j(\tau_0)\right)|g_j|\right|_2 \leq \max_{j\in H}|\Theta_j(\widehat{\tau}) - \Theta_j(\tau_0)|_2\sum_{j\in H}|g_j| \\
\leq & \sqrt{h}\sup_{\tau\in\mathbb{T}}\max_{j\in H}|\Theta_j(\tau) - \Theta_j(\tau_0)|_2 = O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
&\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g - g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g\right| \\
\leq & \|\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\|_\infty \left|\left(\widehat{\boldsymbol{\Theta}}(\widehat{\tau}) - \boldsymbol{\Theta}(\widehat{\tau})\right)' g\right|_1^2 + 2\widetilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \boldsymbol{\Sigma}_{\boldsymbol{xu}}) \left|\left(\widehat{\boldsymbol{\Theta}}(\widehat{\tau}) - \boldsymbol{\Theta}(\widehat{\tau})\right)' g\right|_2 \widetilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \boldsymbol{\Theta}) \left|g\right|_2 \\
\leq & O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right)^2 + O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right) = O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right).
\end{aligned}
$$

Finally, under Assumption 7 (ii),

$$
\begin{aligned}
&\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g - g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g\right| \\
= & O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right) + O_p\left(h\bar{s}\sqrt{\frac{\log p}{n}}\right) + O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).
\end{aligned}
$$

□

*Proof of Theorem 3 in no threshold effect case.* **Step 1**.

**Step 1.1)** Given that $\tau_0$ is undefined and unknown in the current setup, it is necessary to show the asymptotic standard normality of $t_1'(\tau) = \frac{g'\boldsymbol{\Theta}(\tau)\boldsymbol{X}'(\tau)U/n^{1/2}}{\sqrt{g'\boldsymbol{\Theta}(\tau)\boldsymbol{\Sigma}(\tau)_{xu}\boldsymbol{\Theta}(\tau)'g}}$ uniformly over $\tau \in \mathbb{T}$. Subsequently, for any $\widehat{\tau}$ obtained from (2.4), we need to show get $t_1'(\widehat{\tau})$ and $t_1$ are asymptotically equivalent.

We will follow the proof (part a) of Theorem 2 in Caner and Kock (2018) to derive our results. As $E(U_i|X_i) = 0$ for all $i = 1, ..., n$, we have

$$
E\left[t_1'(\tau)\right] = E\left[\frac{g'\boldsymbol{\Theta}(\tau)\sum_{i=1}^n \boldsymbol{X}_i(\tau)U_i/n^{1/2}}{\sqrt{g'\boldsymbol{\Theta}(\tau)\boldsymbol{\Sigma}_{xu}(\tau)\boldsymbol{\Theta}(\tau)'g}}\right] = 0, \tag{7.48}
$$

and

$$
E\left[\left(t_1'(\tau)\right)^2\right] = E\left[\left(\frac{g'\boldsymbol{\Theta}(\tau)\sum_{i=1}^n \boldsymbol{X}_i(\tau)U_i/n^{1/2}}{\sqrt{g'\boldsymbol{\Theta}(\tau)\boldsymbol{\Sigma}(\tau)_{xu}\boldsymbol{\Theta}(\tau)'g}}\right)^2\right] = 1.
$$

Next, we will apply Lyapounov's central limit theorem for a sequence of independent random variables, we thus need to show that for some $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E\left[|g'\mathbf{\Theta}(\tau)\boldsymbol{X}_i(\tau)'U_i/n^{1/2}|\right]^{2+\varepsilon}}{(g'\mathbf{\Theta}(\tau)\mathbf{\Sigma}_{xu}(\tau)\mathbf{\Theta}(\tau)'g)^{1+\varepsilon/2}} \to 0.$$

Let $\widetilde{S}(\tau) = \cup_{j\in H} S_j(\tau)$, then the cardinality $\sup_{\tau\in\mathbb{T}} |\widetilde{S}(\tau)| = 2p \wedge h\bar{s}$. We then have

$$\begin{aligned}
&E\left[\left|g'\mathbf{\Theta}(\tau)\boldsymbol{X}_i'(\tau)U_i/n^{1/2}\right|^{2+\varepsilon}\right] \le E\left[\left|\left|g'\mathbf{\Theta}(\tau)/n^{1/2}\right|_1 \max_{j\in\widetilde{S}(\tau)} \left(\boldsymbol{X}_i^{(j)}(\tau)U_i\right)\right|^{2+\varepsilon}\right] \\
&\le E\left[\left|g'\mathbf{\Theta}(\tau)/n^{1/2}\right|_1^{2+\varepsilon} \max_{j\in\widetilde{S}(\tau)} \left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \le \left|g'\mathbf{\Theta}(\tau)/n^{1/2}\right|_1^{2+\varepsilon} E\left[\max_{j\in\widetilde{S}(\tau)} \left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \\
&\le \left|g'\mathbf{\Theta}(\tau)/n^{1/2}\right|_1^{2+\varepsilon} E\left[\sum_{j\in\widetilde{S}(\tau)} \left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \le \left|g'\mathbf{\Theta}(\tau)/n^{1/2}\right|_1^{2+\varepsilon} (p\wedge h\bar{s}) \max_{j\in\widetilde{S}(\tau)} E\left[\left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \\
&\le \left|g'\mathbf{\Theta}(\tau)/n^{1/2}\right|_1^{2+\varepsilon} (p\wedge h\bar{s}) \max_{1\le j\le p} E\left[\left|X_i^{(j)}U_i\right|^{2+\varepsilon}\right] \\
&= O_p\left(\frac{(h\bar{s})^{2+\varepsilon/2}}{n^{1+\varepsilon/2}}\right) \max_{1\le j\le p} E\left[\left(X_i^{(j)}U_i\right)^{2+\varepsilon}\right] \wedge O_p\left(\frac{(h\bar{s})^{1+\varepsilon/2}p}{n^{1+\varepsilon/2}}\right) \max_{1\le j\le p} E\left[\left(X_i^{(j)}U_i\right)^{2+\varepsilon}\right],
\end{aligned}$$

where the first inequality follows from the Holder's inequality.

$E\left[\left(X_i^{(j)}U_i\right)^4\right] \le \sqrt{E\left[\left(X_i^{(j)}\right)^8\right] E\left[(U_i)^8\right]}$ is bounded by Cauchy–Schwarz inequality under assumption 7 (i). We thus take $\varepsilon = 2$, $\sum_{i=1}^{n} E\left[\left|g'\mathbf{\Theta}(\tau)\boldsymbol{X}_i(\tau)U_i/n^{1/2}\right|^4\right] = O_p\left(\frac{(h\bar{s})^3}{n}\right) \wedge O_p\left(\frac{(h\bar{s})^2 p}{n}\right) = o_p(1)$ under Assumption 7 (iv)

Next, we show that $g'\mathbf{\Theta}(\tau)\Sigma_{xu}(\tau)\mathbf{\Theta}(\tau)'g$ is asymptotically bounded away from zero. We have,

$$\begin{aligned}
&g'\mathbf{\Theta}(\tau)\mathbf{\Sigma}_{xu}(\tau)\mathbf{\Theta}(\tau)'g \ge \kappa(\bar{s}, c_0, \mathbb{T}, \mathbf{\Sigma}_{xu}) \left|g'\mathbf{\Theta}(\tau)\right|_2^2 \\
&\ge \kappa(\bar{s}, c_0, \mathbb{T}, \mathbf{\Sigma}_{xu}) \left|g'\right|_2^2 \kappa(\bar{s}, c_0, \mathbb{T}, \mathbf{\Theta})^2 = \kappa(\bar{s}, c_0, \mathbb{T}, \mathbf{\Sigma}_{xu})\kappa(\bar{s}, c_0, \mathbb{T}, \mathbf{\Theta})^2,
\end{aligned} \tag{7.49}$$

which is bounded away from zero since $\kappa(\bar{s}, c_0, \mathbb{T}, \mathbf{\Sigma}_{xu})$ and $\kappa(\bar{s}, c_0, \mathbb{T}, \mathbf{\Theta})$ are bounded away from zero under Assumption 7 (iv). The Lyapunov condition is thus satisfied. For $\forall\, \tau \in \mathbb{T}, t_1'(\tau)$ converges in distribution to a standard normal distribution.

**Step 1.2)**.

Let

$$t_1''(\widehat{\tau}) = \frac{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})g}}.$$

We have

$$\begin{aligned}&\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}\right| \leq \left|g'\left(\widehat{\boldsymbol{\Theta}}(\widehat{\tau}) - \boldsymbol{\Theta}(\widehat{\tau})\right)\right|_1 \left|\boldsymbol{X}(\widehat{\tau})U/n^{1/2}\right|_\infty \\ =&O_p\left(\sqrt{h}\bar{s}\frac{\sqrt{\log p}}{\sqrt{n}}\right) O_p\left(\sqrt{\log p}\right) = O_p\left(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}\right) = o_p(1),\end{aligned} \tag{7.50}$$

where the first equality holds by conditioning on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$ and by Lemma 16, and the last equality holds under Assumption 7 (ii). In addition, we can show

$$|t_1''(\widehat{\tau}) - t_1| = \frac{g'\left(\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - \widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}\right)}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}} = o_p(1),$$

$$\begin{aligned}|t_1'(\widehat{\tau}) - t_1''(\widehat{\tau})| =& \frac{\left(\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g} - \sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}\right) g'\boldsymbol{\Theta}(\widehat{\tau})X(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}\sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}} \\
=&\frac{\left(g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g - g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g\right) g'\boldsymbol{\Theta}(\widehat{\tau})X(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}\sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}\left(\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g} + \sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}\right)} \\
\leq&\frac{\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g - g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g\right| g'\boldsymbol{\Theta}(\widehat{\tau})X(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}\sqrt{g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\tau_0)\boldsymbol{\Theta}(\tau_0)'g}\left(\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g} + \sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}\right)} \\
\leq&\frac{O_p\left(h\sqrt{s_0^3\bar{s}^2}\sqrt{\frac{\log p}{n}}\right) O_p\left(\sqrt{h\bar{s}\log p}\right)}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}\sqrt{g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\tau_0)\boldsymbol{\Theta}(\tau_0)'g}\left(\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g} + \sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}\right)} \\
=&o_p(1)\end{aligned}$$

by Lemma 19. Thus, $|t_1 - t_1'(\widehat{\tau})| = o_p(1)$.

**Step 2**. In addition, we have

$$t_2 = \frac{g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}} = o_p(1)$$

by Lemma 17. Finally, by Slutsky's theorem, $t = o_p(1) + t_1'(\widehat{\tau}) \xrightarrow{d} N(0,1)$. □

### 7.5.2 Fixed Threshold Effect

This subsection proves the case where the threshold effect is well-identified and discontinuous. To show that the ratio

$$t = \frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}} \tag{7.51}$$

is asymptotically standard normal. Now, by (3.5), $t = t_1 + t_2$, where

$$t_1 = \frac{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}} \text{ and } t_2 = \frac{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2} - g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}},$$

which still suffices to show that $t_1$ is asymptotically standard normal and $t_2 = o_p(1)$.

**Lemma 20.** *Suppose that Assumptions 1 to 7 hold, conditional on events* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, *and* $\mathbb{A}_5$, *then we have* $g'\Delta(\widehat{\tau}) = O_p\left(\frac{s_0\sqrt{h}\log p}{\sqrt{n}}\right)$.

*Proof of Lemma 20.* Recall that $\Delta(\tau) = \sqrt{n}\left(\widehat{\boldsymbol{\Theta}}(\tau)\widehat{\boldsymbol{\Sigma}}(\tau) - I_{2p}\right)(\widehat{\alpha}(\widehat{\tau}) - \alpha_0)$. Then, by Holder's inequality, Lemma 16 and Theorem 2, we have,

$$\begin{aligned} g'\Delta(\widehat{\tau}) &\le \max_{j\in H}|\Delta_j(\widehat{\tau})|\sum_{j\in H}|g_j| \le \max_{j\in H}\left|\left(\widehat{\boldsymbol{\Theta}}_j(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}(\widehat{\tau}) - e_j'\right)\sqrt{n}(\widehat{\alpha}(\widehat{\tau}) - \alpha_0)\right|\sum_{j\in H}|g_j| \\ &\le \max_{1\le j\le 2p}\left|\widehat{\boldsymbol{\Theta}}_j(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}(\widehat{\tau}) - e_j'\right|_\infty \sqrt{n}\,|\widehat{\alpha}(\widehat{\tau}) - \alpha_0|_1\sum_{j\in H}|g_j| \\ &\le C\left(\frac{\lambda_{node}}{\widehat{z}_j^2(\widehat{\tau})} + \frac{\lambda_{node}}{\widehat{\widetilde{z}}_j^2(\widehat{\tau})}\right)\sqrt{n}\lambda s_0\sqrt{h} = O_p\left(\frac{s_0\sqrt{h}\log p}{\sqrt{n}}\right). \end{aligned}$$

□

The result of Lemma 20 is similar to that of Lemma 17 but is derived under different conditions.

**Lemma 21.** *Suppose that Assumptions 1 to 7 hold and let $g$ be $2p \times 1$ vector satisfying $|g|_2 = 1$. Then, conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, and $\mathbb{A}_5$, we have*

$$\left| g' \left( \widehat{\mathbf{\Theta}}(\widehat{\tau}) - \widehat{\mathbf{\Theta}}(\tau_0) \right) \right|_1 = O_p \left( \sqrt{h} \bar{s} \sqrt{\frac{\log p}{n}} \right).$$

*Proof of Lemma 21.* As $Q_i$ is continuously distributed and $E\left[ \left| X_i^{(j)} X_i^{(l)} \right| |Q_i = \tau \right]$ is continuous and bounded in a neighborhood of $\tau_0$, such that conditions for Lemma A.1 in Hansen (2000) hold. Then, we have

$$\begin{aligned}
\|\mathbf{\Sigma}(\tau_0) - \mathbf{\Sigma}(\widehat{\tau})\|_\infty = & \left\| \left[ \begin{array}{cc} 0 & \boldsymbol{M}(\tau_0) - \boldsymbol{M}(\widehat{\tau}) \\ \boldsymbol{M}(\tau_0) - \boldsymbol{M}(\widehat{\tau}) & \boldsymbol{M}(\tau_0) - \boldsymbol{M}(\widehat{\tau}) \end{array} \right] \right\|_\infty \\
\leq & \|\boldsymbol{M}(\tau_0) - \boldsymbol{M}(\widehat{\tau})\|_\infty = \max_{1 \leq j,l \leq p} E\left[ \left| X_i^{(j)} X_i^{(l)} \right| |1(Q_i < \tau_0) - 1(Q_i < \widehat{\tau})| \right] \\
\leq & C |\tau_0 - \widehat{\tau}| = O_p \left( \frac{(\log p) s_0}{n} \right)
\end{aligned}$$

where the last inequality is by Lemma A.1 in Hansen (2000) and the last equality is due to Theorem 2. Next, consider

$$\begin{aligned}
& \left| \Theta_j(\widehat{\tau}) - \Theta_j(\tau_0) \right|_1 = \left| \Theta_j(\widehat{\tau}) \left( \Sigma_j(\tau_0) - \Sigma_j(\widehat{\tau}) \right)' \Theta_j(\tau_0) \right|_1 \\
\leq & \left| \Theta_j(\widehat{\tau}) \|_1 \| \left( \Sigma_j(\tau_0) - \Sigma_j(\widehat{\tau}) \right)' \Theta_j(\tau_0) \right|_\infty \leq \left| \Theta_j(\widehat{\tau}) \|_1 \| \Theta_j(\tau_0) \|_1 \| \left( \Sigma_j(\tau_0) - \Sigma_j(\widehat{\tau}) \right)' \right|_\infty .
\end{aligned}$$

Then, by Lemma 16, we obtain

$$\begin{aligned}
& \left| g' \left( \mathbf{\Theta}(\widehat{\tau}) - \mathbf{\Theta}(\tau_0) \right) \right|_1 = \sum_{j \in H} \left( |g_j| \| \Theta_j(\widehat{\tau}) - \Theta_j(\tau_0) \|_1 \right) \leq \sum_{j \in H} |g_j| \max_{j \in H} \left| \Theta_j(\widehat{\tau}) - \Theta_j(\tau_0) \right|_1 \\
\leq & \sqrt{h} \max_{j \in H} \left| \Theta_j(\tau) \right|_1 \max_{j \in H} \left| \Theta_j(\tau_0) \right|_1 \left| \left( \Sigma_j(\tau_0) - \Sigma_j(\widehat{\tau}) \right)' \right|_\infty = O_p \left( \sqrt{h} \bar{s} s_0 \frac{\log p}{n} \right).
\end{aligned} \tag{7.52}$$

We thus have,

$$
\begin{aligned}
&\left|g'\left(\widehat{\mathbf{\Theta}}(\widehat{\tau})-\widehat{\mathbf{\Theta}}(\tau_0)\right)\right|_1\\
\leq&\sum_{j\in H}|g_j|\max_{j\in H}\left|\widehat{\Theta}_j(\widehat{\tau})-\Theta_j(\widehat{\tau})\right|_1+\sum_{j\in H}|g_j|\sup_{\tau\in\mathbb{T}}\left|\Theta_j(\widehat{\tau})-\Theta_j(\tau_0)\right|_1+\sum_{j\in H}|g_j|\max_{j\in H}\left|\widehat{\Theta}_j(\tau_0)-\Theta_j(\tau_0)\right|_1\\
=&O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right)+O_p\left(\sqrt{h\bar{s}}s_0\frac{\log p}{n}\right)=O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right),
\end{aligned}
$$

as $s_0\sqrt{\frac{\log p}{n}}=o_p(1)$ under Assumption 1. □

**Lemma 22.** *Suppose that Assumptions 1 to 7 hold, conditional on events* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, *and* $\mathbb{A}_5$, *then we have*

$$
\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)-\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}\right|=O_p\left(\frac{s_0^2\sqrt{h\bar{s}}\log p}{\sqrt{n}}\right).
$$

*Proof of Lemma 22.* There are only two cases for $\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)$: $\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)=\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\tau_0)$ or $\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)=\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})$, thus.

$$
\begin{aligned}
&\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\left(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)-\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\right)\alpha_0/n^{1/2}\right|\\
\leq&\sqrt{n}\sum_{j\in H}|g_j|\left|\widehat{\Theta}_j(\widehat{\tau})\right|_1\left\|\begin{bmatrix}0 & \widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\widehat{\tau})\\ 0 & \widehat{\boldsymbol{M}}(\min\{\tau_0,\widehat{\tau}\})-\widehat{\boldsymbol{M}}(\widehat{\tau})\end{bmatrix}\begin{bmatrix}\beta_0' & \delta_0'\end{bmatrix}'\right\|_\infty\\
\leq&\sqrt{n}\max_{j\in H}\left|\widehat{\Theta}_j(\widehat{\tau})\right|_1\sum_{j\in H}|g_j|\left|\widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\widehat{\tau})\right|_\infty|\delta_0|_1.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\left\|\widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\widehat{\tau})\right\|_\infty\leq\max_{1\leq j,l\leq p}\left|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\widehat{\tau}\right)\right]\right|\\
\leq&\max_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|\leq|\tau_0-\widehat{\tau}|}\left|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|\\
\leq&C_5|\tau_0-\widehat{\tau}|=O_p\left(\frac{s_0\log p}{n}\right),
\end{aligned}
$$

where the last equality follows from Assumption 4, and we know that $\sup_{\tau\in\mathbb{T}}$

$\max_{j\in H}\left|\widehat{\Theta}_j(\tau)\right|_1 = O_p(\sqrt{\bar{s}})$ by Lemma 16, we thus obtain,

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)-\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}\right| \le \sqrt{n}\max_{j\in H}\left|\widehat{\Theta}_j(\widehat{\tau})\right|_1\sum_{j\in H}|g_j|\left\|\widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\widehat{\tau})\right\|_\infty|\delta_0|_1$$

$$\le \sqrt{n}O_p(\sqrt{h\bar{s}})O_p\left(\frac{s_0\log p}{n}\right)|\delta_0|_1 = O_p\left(\frac{|\delta_0|_1 s_0\sqrt{h\bar{s}}\log p}{\sqrt{n}}\right) = O_p\left(\frac{s_0^2\sqrt{h\bar{s}}\log p}{\sqrt{n}}\right).$$

□

**Lemma 23.** *Suppose that Assumptions 1 to 7 hold, conditional on events* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, *and* $\mathbb{A}_5$, *then we have*

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g\right| = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).$$

*Proof of Lemma 23.* Recall the fixed threshold effect case, we have $\mathbf{\Sigma}_{xu}(\widehat{\tau}) = E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2\right]$, $\widehat{U}_i(\widehat{\tau}) = Y_i - \boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau}) = U_i + \boldsymbol{X}_i(\tau_0)'\alpha_0 - \boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})$, $\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{U}_i(\widehat{\tau})^2$, and define $\widetilde{\mathbf{\Sigma}}(\widehat{\tau})_{xu} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2$. We first need to prove the followings, as in Lemma 19,

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| = o_p(1), \tag{7.53}$$

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g\right| = o_p(1), \tag{7.54}$$

$$\left|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g\right| = o_p(1). \tag{7.55}$$

Proving (7.53) is similar to proving (7.45); the difference is

$$\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) - \widetilde{\mathbf{\Sigma}}_{xu}(\widehat{\tau}) = \frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{U}_i^2(\widehat{\tau}) - \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2\right)$$

$$=\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'(U_i + \boldsymbol{X}_i(\tau_0)'\alpha_0 - \boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau}))^2 - \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i^2\right)$$

$$=\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\alpha_0\boldsymbol{X}_i(\tau_0)'\left(\boldsymbol{X}_i(\tau_0)'\alpha_0 - \boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\left(\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}_i(\tau_0)'\alpha_0\right)\right)$$

$$+\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'U_i\left(\alpha_0'\boldsymbol{X}_i(\tau_0)-\widehat{\alpha}(\widehat{\tau})'\boldsymbol{X}_i(\widehat{\tau})\right)\right),$$

given that $\alpha_0'\boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})=\widehat{\alpha}'(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i(\tau_0)\alpha_0'$.

By Cauchy-Schwarz inequality and Hölder's inequality

$$\max_{1\le k,l\le 2p}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\alpha_0'\boldsymbol{X}_i(\tau_0)\left(\boldsymbol{X}_i(\tau_0)'\alpha_0-\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)\right)\right|$$

$$\le\sqrt{\max_{1\le k,l\le 2p}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)^2(\boldsymbol{X}_i(\tau_0)\alpha_0)^2}\|\boldsymbol{X}(\tau_0)'\alpha_0-\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_n$$

$$\le\sqrt{\max_{1\le k,l\le 2p}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)^2\left(\max_{1\le k\le 2p}\boldsymbol{X}_i^{(k)}(\tau_0)\right)^2|\alpha_0|_1^2}\|\boldsymbol{X}(\tau_0)\alpha_0-\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_n$$

$$\le\sqrt{\max_{1\le k,l,j\le p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(k)}X_i^{(l)}X_i^{(j)}\right)^2|\alpha_0|_1^2\,\mathbf{1}\left(Q_i<\tau_0\right)\mathbf{1}\left(Q_i<\widehat{\tau}\right)}\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\tau_0)\alpha_0\|_n$$

$$=O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right),$$

the last equality follows from Lemma 18, and $|\widehat{\alpha}(\widehat{\tau})|_1\le|\alpha_0|_1+O_p\left(s_0\sqrt{\frac{\log p}{n}}\right)$, $\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\widehat{\tau})\alpha_0\|_n=O_p\left(\sqrt{s_0}\sqrt{\frac{\log p}{n}}\right)$ by Theorem 2, and $|\alpha_0|_1=O_p(s_0)$ under Assumption 1. Also, we have

$$\max_{1\le k,l\le 2p}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau})'\left(\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}_i(\tau_0)'\alpha_0\right)\right)\right|$$

$$\le\max_{1\le k,l\le 2p}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)^2(\widehat{\alpha}'(\widehat{\tau})\boldsymbol{X}_i(\widehat{\tau}))^2}\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})-\boldsymbol{X}(\widehat{\tau})\alpha_0\|_n$$

$$\leq \sqrt{\max_{1\leq k,l,j\leq p} \frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(k)}X_i^{(l)}X_i^{(j)}\right)^2 |\widehat{\alpha}(\widehat{\tau})|_1^2 \mathbf{1}\left(Q_i < \widehat{\tau}\right)} \|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}) - \boldsymbol{X}(\widehat{\tau})\alpha_0\|_n$$

$$= O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right),$$

and

$$\max_{1\leq k,l\leq 2p}\left|\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\tau_0)'\alpha_0 - \boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}(\widehat{\tau})\right)\left(\boldsymbol{X}_i^{(k)}(\widehat{\tau})\boldsymbol{X}_i^{(l)}(\widehat{\tau})\right)U_i\right|$$

$$\leq 2\sqrt{\max_{1\leq k,l\leq p}\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(k)}X_i^{(l)}U_i\right)^2 \mathbf{1}\left(Q_i < \widehat{\tau}\right)}\|\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}) - \boldsymbol{X}(\tau_0)\alpha_0\|_n = O_p\left(\sqrt{s_0}\sqrt{\frac{\log p}{n}}\right).$$

We then obtain

$$\left\|\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau}) - \widetilde{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\right\|_\infty = O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).$$

Therefore,

$$\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g - g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widetilde{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g\right|$$

$$= O_p\left(h\bar{s}\right) O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).$$

Proving (7.54) and (7.55) is the same as deriving (7.46) and (7.47) in Lemma 19. □

**Lemma 24.** *Suppose that Assumptions 1 to 7 hold, conditional on events* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, *and* $\mathbb{A}_5$, *then*

$$\left|g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g - g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\tau_0)\boldsymbol{\Theta}(\tau_0)'g\right| = o_p(1).$$

*Proof of Lemma 24.* To prove this lemma, we require to prove the following,

$$|g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g - g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g| \tag{7.56}$$

$$|g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g - g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\tau_0)'g| \tag{7.57}$$

$$|g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\tau_0)'g - g'\boldsymbol{\Theta}(\tau_0)\boldsymbol{\Sigma}_{xu}(\tau_0)\boldsymbol{\Theta}(\tau_0)'g|. \tag{7.58}$$

Firstly, we prove (7.56). Since $\boldsymbol{\Theta}(\widehat{\tau})$ is symmetric, $|\boldsymbol{\Theta}(\widehat{\tau})'g|_1 = |g'\boldsymbol{\Theta}(\widehat{\tau})|_1$. Additionally,

$\|\mathbf{\Sigma}_{xu}(\widehat{\tau})\|_\infty$ is bounded under Assumption 1. Combining these with (7.52), we obtain

$$
\begin{aligned}
&|g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g - g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g| \le |g'\left(\mathbf{\Theta}(\widehat{\tau}) - \mathbf{\Theta}(\tau_0)\right)|_1 \|\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g\|_\infty \\
\le& |g'\left(\mathbf{\Theta}(\widehat{\tau}) - \mathbf{\Theta}(\tau_0)\right)|_1 \|\mathbf{\Sigma}_{xu}(\widehat{\tau})\|_\infty |g'\mathbf{\Theta}(\widehat{\tau})|_1 = O_p\left(\sqrt{h\bar{s}}s_0\frac{\log p}{n}\right) O_p\left(\sqrt{h\bar{s}}\right) = O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right).
\end{aligned}
$$

To prove (7.57), as $|g'\left(\mathbf{\Theta}(\widehat{\tau}) - \mathbf{\Theta}(\tau_0)\right)|_1 = |\left(\mathbf{\Theta}(\widehat{\tau}) - \mathbf{\Theta}(\tau_0)\right)' g|_1$, we derive

$$
\begin{aligned}
&|g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g - g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\tau_0)'g| \le |g'\mathbf{\Theta}(\tau_0)|_1\|\mathbf{\Sigma}_{xu}(\widehat{\tau})\|_\infty |\left(\mathbf{\Theta}(\widehat{\tau}) - \mathbf{\Theta}(\tau_0)\right)' g|_1 \\
&= O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right).
\end{aligned}
$$

To prove (7.58), we write

$$
\begin{aligned}
&\mathbf{\Sigma}_{xu}(\widehat{\tau}) - \mathbf{\Sigma}_{xu}(\tau_0) = E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i'(\widehat{\tau})u_i^2\right] - E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i'(\tau_0)u_i^2\right] \\
\le& E\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\widehat{\tau})\boldsymbol{X}_i'(\widehat{\tau}) - \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i'(\tau_0)\right] \max_{1\le i\le n} E\left[u_i^2\right] = E\left[\widehat{M}(\widehat{\tau}) - \widehat{M}(\tau_0)\right] \max_{1\le i\le n} E\left[u_i^2\right].
\end{aligned}
$$

Since we have $\left\|(\widehat{M}(\widehat{\tau}) - \widehat{M}(\tau_0)\right\|_\infty = O_p\left(s_0\frac{\log p}{n}\right)$, we obtain,

$$
\begin{aligned}
&|g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\tau_0)'g - g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\tau_0)\mathbf{\Theta}(\tau_0)'g| \le |g'\mathbf{\Theta}(\tau_0)\left(\mathbf{\Sigma}_{xu}(\widehat{\tau}) - \mathbf{\Sigma}_{xu}(\tau_0)\right)\mathbf{\Theta}(\tau_0)'g| \\
\le& |g'\mathbf{\Theta}(\tau_0)|_1^2\|\mathbf{\Sigma}_{xu}(\widehat{\tau}) - \mathbf{\Sigma}_{xu}(\tau_0)\|_\infty = O_p\left(h\bar{s}\right) O_p\left(s_0\frac{\log p}{n}\right) = O_p\left(h\bar{s}s_0\frac{\log p}{n}\right).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
|g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g - g'\mathbf{\Theta}(\tau_0)_{xu}\mathbf{\Sigma}(\tau_0)\mathbf{\Theta}(\tau_0)'g| &= O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right) + O_p\left(h\bar{s}s_0\frac{\log p}{n}\right) \\
&= O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right).
\end{aligned}
$$

By Lemma 23, we obtain

$$|g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g - g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\tau_0)\mathbf{\Theta}(\tau_0)'g| = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right) + O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right)$$
$$= O_p\left(h\sqrt{s_0^3\bar{s}^3}\sqrt{\frac{\log p}{n}}\right).$$

□

*Proof of Theorem 3 in the fixed threshold effect case.* **Step 1**.

This step is the same as Step 1 in the proof of Theorem 3 for the no-threshold case, implying that in the fixed-threshold case, $|t_1 - t_1'(\widehat{\tau})| = o_p(1)$, where $t_1'(\widehat{\tau})$ converges in distribution to a standard normal distribution.

**Step 2**. By Lemma 20 and 22,

$$t_2 = \frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2} - g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} = o_p(1).$$

Finally, by Slutsky's theorem,

$$t = o_p(1) + t_1'(\widehat{\tau}) \xrightarrow{d} N(0,1).$$

Additionally, Lemma 24 implies that $\sup_{\alpha_0 \in \mathcal{A}_{\ell_0}^{(2)}(s_0)} \left|\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})' - \mathbf{\Theta}(\tau_0)\mathbf{\Sigma}_{xu}(\tau_0)\mathbf{\Theta}(\tau_0)'\right|$ $= o_p(1)$. □

## 7.6 Proof of Theorem 4

*Proof of Theorem 4.* We will follow the proof of Theorem 3 in Caner and Kock (2018). For $\varepsilon > 0$, define the following events

$$\mathcal{F}_{1,n} = \left\{\sup_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} |g'\Delta(\widehat{\tau})| < \varepsilon\right\},$$

$$\mathcal{F}_{2,n} = \left\{\sup_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} \left|\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g} - 1\right| < \varepsilon\right\},$$

$$\mathcal{F}_{3,n} = \left\{ \sup_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} |g'\widehat{\mathbf{\Theta}}(\widehat{\tau})X(\widehat{\tau})'U/n^{1/2} - g'\mathbf{\Theta}(\widehat{\tau})X(\widehat{\tau})'U/n^{1/2}| < \varepsilon \right\},$$

$$\mathcal{F}_{4,n} = \left\{ \sup_{\alpha_0 \in \mathcal{A}^{(2)}_{\ell_0}(s_0)} |g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}| < \varepsilon \right\}.$$

By Lemma 17 (and 20), Lemma 19 (and 23), (7.50) from Step 1.2 in the proof of Theorem 3 in no threshold effect case, and Lemma 22, respectively, we obtain that the probabilities of these sets all approach one. Thus, for each $t \in \mathbb{R}$, we have

$$\begin{aligned}
&\left|\mathbb{P}\left\{\frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\} - \Phi(t)\right| \\
=&\left|\mathbb{P}\{\delta_0 \neq 0\}\,\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\}\right. \\
&\left.+\mathbb{P}\{\delta_0 = 0\}\,\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\} - \Phi(t)\right| \\
\le&\mathbb{P}\{\delta_0 \neq 0\}\left|\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\} - \Phi(t)\right| \\
&+\mathbb{P}\{\delta_0 = 0\}\left|\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\} - \Phi(t)\right|,
\end{aligned} \tag{7.59}$$

where $\mathbb{P}\{\delta_0 = 0\} + \mathbb{P}\{\delta_0 \neq 0\} = 1$. Firstly, we consider the second term in the last inequality of (7.59), we write

$$\begin{aligned}
&\left|\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\} - \Phi(t)\right| \\
\le&\left|\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}\right\} - \Phi(t)\right| + \mathbb{P}\left\{\mathcal{F}^c_{1,n} \cup \mathcal{F}^c_{2,n} \cup \mathcal{F}^c_{3,n}\right\}.
\end{aligned} \tag{7.60}$$

As $g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g$ is bounded away from zero, there exists a positive constant

$D_1$ such that

$$
\begin{aligned}
&\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}}\leq t,\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\\
=&\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}\leq t\sqrt{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}},\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\\
\leq&\mathbb{P}\left\{\frac{g'\mathbf{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}\leq t(1+\varepsilon)+\frac{\varepsilon+\varepsilon}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}\right\}\\
\leq&\mathbb{P}\left\{\frac{g'\mathbf{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}\leq t(1+\varepsilon)+D_1\varepsilon\right\}\\
\leq&\Phi(t(1+\varepsilon)+D_1\varepsilon)+\varepsilon,
\end{aligned}
\tag{7.61}
$$

where the last inequality is derived from the proof of Theorem 3, in which we established the asymptotic normality of $\frac{g'\mathbf{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}$. Since the right-hand sides in the last inequality in (7.61) do not depend on $\alpha_0$, we obtain

$$
\sup_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}}\leq t,\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\leq\Phi(t(1+\varepsilon)+D_1\varepsilon)+\varepsilon.
\tag{7.62}
$$

The above arguments hold for all $\varepsilon>0$. By the continuity of $\Phi(\cdot)$, for any $\eta>0$, we can choose $\varepsilon$ to be sufficiently small and derive that

$$
\sup_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}}\leq t,\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\leq\Phi(t)+\eta+\varepsilon.
\tag{7.63}
$$

Next, as $g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g$ is bounded away from zero, there exists a positive constant $D_2$ such that

$$
\begin{aligned}
&\mathbb{P}\left\{\frac{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}}\leq t,\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\\
=&\mathbb{P}\left\{\frac{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}}\leq t\sqrt{\frac{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}},\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\\
\geq&\mathbb{P}\left\{\frac{g'\boldsymbol{\Theta}(\widehat{\tau})X(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}}\leq t(1-\varepsilon)-\frac{\varepsilon+\varepsilon}{\sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}},\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\\
\geq&\mathbb{P}\left\{\frac{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}}\leq t(1-\varepsilon)-D_2\varepsilon\right\}+\mathbb{P}\left\{\mathcal{F}_{1,n}\cap\mathcal{F}_{2,n}\cap\mathcal{F}_{3,n}\right\}-1\\
\geq&\Phi(t(1-\varepsilon)-D_2\varepsilon)-\varepsilon+\mathbb{P}\left\{\mathcal{F}_{1,n}\cap\mathcal{F}_{2,n}\cap\mathcal{F}_{3,n}\right\}-1,
\end{aligned}
\tag{7.64}
$$

where the last inequality is from the asymptotic normality of $\frac{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}}{\sqrt{g'\boldsymbol{\Theta}(\widehat{\tau})\boldsymbol{\Sigma}_{xu}(\widehat{\tau})\boldsymbol{\Theta}(\widehat{\tau})'g}}$.

As $\mathbb{P}\left\{\mathcal{F}_{1,n}\cap\mathcal{F}_{2,n}\cap\mathcal{F}_{3,n}\right\}$ can arbitrarily approach to one by choosing $n$ sufficiently large and $\varepsilon$ sufficiently small, meanwhile, the right-hand sides in the last inequality in (7.64) do not depend on $\alpha_0$, we have

$$
\inf_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}_{xu}(\widehat{\tau})\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}}\leq t,\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\geq\Phi(t(1-\varepsilon)-D_2\varepsilon)-\varepsilon.
\tag{7.65}
$$

By the continuity of $\Phi(\cdot)$, for any $\eta>0$, we can choose $\varepsilon$ to be sufficiently small and obtain

$$
\inf_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2}-g'\Delta(\widehat{\tau})}{\sqrt{g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})\widehat{\boldsymbol{\Sigma}}(\widehat{\tau})_{xu}\widehat{\boldsymbol{\Theta}}(\widehat{\tau})'g}}\leq t,\mathcal{F}_{1,n},\mathcal{F}_{2,n},\mathcal{F}_{3,n}\right\}\geq\Phi(t)-\eta-2\varepsilon.
\tag{7.66}
$$

Combining (7.63) and (7.66), and $\sup_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)}\mathbb{P}\left\{\mathcal{F}_{1,n}^c\cup\mathcal{F}_{2,n}^c\cup\mathcal{F}_{3,n}^c\right\}\to 0$, we thus derive

$$\left| \sup_{\alpha_0 \in \mathcal{A}^{(1)}_{\ell_0}(s_0)} \mathbb{P}\left\{ \frac{\sqrt{n} g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t \right\} - \Phi(t) \right| \to 0. \tag{7.67}$$

We now consider the first term in the last inequality of (7.59) and write

$$\begin{aligned}
&\left| \mathbb{P}\left\{ \frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t \right\} - \Phi(t) \right| \\
\le &\left| \mathbb{P}\left\{ \frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \le t\sqrt{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}, \mathcal{F}_{4,n} \right\} - \Phi(t) \right| \\
&+ \mathbb{P}\left\{ \mathcal{F}^c_{1,n} \cup \mathcal{F}^c_{2,n} \cup \mathcal{F}^c_{3,n} \cup \mathcal{F}^c_{4,n} \right\}.
\end{aligned} \tag{7.68}$$

As $g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}(\widehat{\tau})_{xu}\mathbf{\Theta}(\widehat{\tau})'g$ is bounded away from zero, there exists a positive $D_3$ such that

$$\begin{aligned}
&\mathbb{P}\left\{ \frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \le t\sqrt{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}, \mathcal{F}_{4,n} \right\} \\
\le &\mathbb{P}\left\{ \frac{g'\mathbf{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \le t(1+\varepsilon) + \frac{\varepsilon + \varepsilon + \varepsilon}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \right\} \\
\le &\mathbb{P}\left\{ \frac{g'\mathbf{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \le t(1+\varepsilon) + D_3\varepsilon \right\} \\
\le &\Phi(t(1+\varepsilon) + D_3\varepsilon) + \varepsilon.
\end{aligned} \tag{7.69}$$

Thus, for any $\eta > 0$, we can choose $\varepsilon$ to be sufficiently small and derive that

$$\begin{aligned}
&\sup_{\alpha_0 \in \mathcal{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{ \frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U/n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}, \mathcal{F}_{4,n} \right\} \\
&\le \Phi(t) + \eta + \varepsilon,
\end{aligned} \tag{7.70}$$

by similar arguments of obtaining (7.63).

Next, as $g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}(\widehat{\tau})_{xu}\mathbf{\Theta}(\widehat{\tau})'g$ is bounded away from zero, there exists a positive constant $D_4$,

$$\mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})'U//n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \le t\sqrt{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}_{xu}(\widehat{\tau})\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}, \mathcal{F}_{4,n}\right\}$$
$$\ge \mathbb{P}\left\{\frac{g'\mathbf{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \le t(1+\varepsilon) - \frac{3\varepsilon}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}}, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}, \mathcal{F}_{4,n}\right\}$$
$$\ge \mathbb{P}\left\{\frac{g'\mathbf{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'g}} \le t(1+\varepsilon) - D_4\varepsilon\right\} + \mathbb{P}\left\{\mathcal{F}_{1,n}\cap\mathcal{F}_{2,n}\cap\mathcal{F}_{3,n}\cap\mathcal{F}_{4,n}\right\} - 1. \tag{7.71}$$

As the right-hand sides in the last inequality in (7.71) do not depend on $\alpha_0$, and $\mathbb{P}\left\{\mathcal{F}_{1,n}\cap\mathcal{F}_{2,n}\cap\mathcal{F}_{3,n}\cap\mathcal{F}_{4,n}\right\}$ can be arbitrarily close to one by choosing $n$ sufficiently large and $\varepsilon$ sufficiently small, we have

$$\inf_{\alpha_0\in\mathcal{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}(\widehat{\tau})_{xu}\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}, \mathcal{F}_{4,n}\right\}$$
$$\ge \mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\mathbf{\Theta}(\tau_0)\mathbf{\Sigma}(\tau_0)_{xu}\mathbf{\Theta}(\tau_0)'g}} \le t(1-\varepsilon) - D_4\varepsilon\right\} - \varepsilon. \tag{7.72}$$

Thus, for any $\eta > 0$, we can choose $\varepsilon$ to be sufficiently small and derive

$$\inf_{\alpha_0\in\mathcal{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{\frac{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Delta(\widehat{\tau}) + g'\widehat{\mathbf{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}(\widehat{\tau})_{xu}\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t, \mathcal{F}_{1,n}, \mathcal{F}_{2,n}, \mathcal{F}_{3,n}, \mathcal{F}_{4,n}\right\}$$
$$\ge \Phi(t) - \eta - 2\varepsilon. \tag{7.73}$$

by similar arguments of obtaining (7.66).

Combining (7.70) and (7.73), and $\sup_{\alpha_0\in\mathcal{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{\mathcal{F}^c_{1,n}\cup\mathcal{F}^c_{4,n}\cup\mathcal{F}^c_{5,n}\cup\mathcal{F}^c_{6,n}\cup\mathcal{F}^c_{7,n}\right\} \to 0$, we thus derive

$$\left|\sup_{\alpha_0\in\mathcal{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{\frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}(\widehat{\tau})_{xu}\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\} - \Phi(t)\right| \to 0 \tag{7.74}$$

Therefore, for (7.59), we have

$$\left|\sup_{\alpha_0\in\mathcal{B}_{\ell_0}(s_0)} \mathbb{P}\left\{\frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\mathbf{\Theta}}(\widehat{\tau})\widehat{\mathbf{\Sigma}}(\widehat{\tau})_{xu}\widehat{\mathbf{\Theta}}(\widehat{\tau})'g}} \le t\right\} - \Phi(t)\right| \to 0. \tag{7.75}$$

To obtain (3.19), we write

$$\mathbb{P}\left\{\alpha_0^{(j)} \notin \left[\widehat{a}^{(j)}(\widehat{\tau}) - z_{1-\alpha/2}\frac{\widehat{\Sigma}_j(\widehat{\tau})}{\sqrt{n}}, \widehat{a}^{(j)}(\widehat{\tau}) + z_{1-\alpha/2}\frac{\widehat{\sigma}_j(\widehat{\tau})}{\sqrt{n}}\right]\right\} = \mathbb{P}\left\{\left|\frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}_j(\widehat{\tau})}\right| > z_{1-\alpha/2}\right\}$$
$$= \mathbb{P}\left\{\frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}_j(\widehat{\tau})} > z_{1-\alpha/2}\right\} + \mathbb{P}\left\{\frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}_j(\widehat{\tau})} < -z_{1-\alpha/2}\right\}$$
$$\leq 1 - \mathbb{P}\left\{\frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}_j(\widehat{\tau})} \leq z_{1-\alpha/2}\right\} + \mathbb{P}\left\{\frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}_j(\widehat{\tau})} < -z_{1-\alpha/2}\right\}. \tag{7.76}$$

Thus, taking the supremum over $\mathcal{B}_{\ell_0}(s_0)$ and letting $n$ go to infinity yields (3.19) by (3.18).

Finally, to prove (3.20), let $g = e_j$ and as $\phi_{\max}(\mathbf{\Theta}(\tau))) = 1/\phi_{\min}(\mathbf{\Sigma}(\tau))$, for $\tau \in \mathbb{T}$, we derive

$$\begin{aligned}
&\sup_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} \operatorname{diam}\left[\widehat{a}^{(j)}(\widehat{\tau}) - z_{1-\alpha/2}\frac{\widehat{\sigma}_j(\widehat{\tau})}{\sqrt{n}}, \widehat{a}^{(j)}(\widehat{\tau}) + z_{1-\alpha/2}\frac{\widehat{\sigma}_j(\widehat{\tau})}{\sqrt{n}}\right] \\
&= \sup_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} 2\widehat{\sigma}_j(\widehat{\tau}) z_{1-\alpha/2}/\sqrt{n} \\
&= 2\left(\sup_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} \sqrt{e_j'\mathbf{\Theta}(\widehat{\tau})\mathbf{\Sigma}_{xu}(\widehat{\tau})\mathbf{\Theta}(\widehat{\tau})'e_j} + o_p(1)\right) z_{1-\alpha/2}/\sqrt{n} \\
&\leq 2\left(\sqrt{\phi_{\max}(\mathbf{\Sigma}_{xu}(\widehat{\tau}))}\frac{1}{\phi_{\min}(\mathbf{\Sigma}(\widehat{\tau}))} + o_p(1)\right) z_{1-\alpha/2}/\sqrt{n} = O_p(1/\sqrt{n}),
\end{aligned} \tag{7.77}$$

the last equality is due to the boundedness of $\phi_{\max}(\mathbf{\Sigma}_{xu}(\widehat{\tau}))$ and $\phi_{\min}(\mathbf{\Sigma}(\widehat{\tau}))$ under Assumptions 2 and 7. □

# 8 Appendix B

## 8.1 Proofs for Section 4.1

We first recall the definitions of Near-Epoch Dependence and Mixingale from Davidson (2002), as formulated in Adamek et al. (2023).

**Definition 8.1** (Near-Epoch Dependence, Davidson (2002), ch. 18)**.** *Suppose that there exist non-negative NED constants* $\{c_i\}_{i=-\infty}^{\infty}$*, an NED sequence* $\{\psi_q\}_{q=0}^{\infty}$ *such* *that $\psi_q \to 0$ as $q \to \infty$, and a (possibly vector-valued) stochastic sequence $\{\boldsymbol{s}_i\}_{i=-\infty}^{\infty}$ with $\mathcal{F}_{i-l-q}^{i-l+q} = \sigma\{\boldsymbol{s}_{i-q}, \dots, \boldsymbol{s}_{i+q}\}$, such that $\{\mathcal{F}_{i-l-q}^{i-l+q}\}_{q=0}^{\infty}$ is an increasing sequence of $\sigma$-fields. For $p > 0$, the random variable $\{X_i\}_{i=-\infty}^{\infty}$ is $L_p$-NED on $\boldsymbol{s}_i$ if*

$$\left(\mathrm{E}\left[\left|X_i - \mathrm{E}\left(X_i|\mathcal{F}_{i-l-q}^{i-l+q}\right)\right|^p\right]\right)^{1/p} \le c_i\psi_q.$$

*for all $i$ and $q \ge 0$. Furthermore, we say $\{X_i\}$ is $L_p$-NED of size $-d$ on $\boldsymbol{s}_i$ if $\psi_q = O(q^{-d-\varepsilon})$ for some $\varepsilon > 0$.*

**Definition 8.2** (Mixingale, Davidson (2002), ch. 17)**.** *Suppose that there exist non-negative mixingale constants $\{c_i\}_{i=-\infty}^{\infty}$ and mixingale sequence $\{\psi_q\}_{q=0}^{\infty}$ such that $\psi_q \to 0$ as $q \to \infty$. For $p \ge 1$, the random variable $\{X_i\}_{i=-\infty}^{\infty}$ is an $L_p$-mixingale with respect to the $\sigma$-algebra $\{\mathcal{F}_i\}_{i=-\infty}^{\infty}$ if*

$$\left(\mathrm{E}\left[\left|\mathrm{E}\left(X_i|\mathcal{F}_{i-q}\right)\right|^p\right]\right)^{1/p} \le c_i\psi_q,$$

$$\left(\mathrm{E}\left[\left|X_i - \mathrm{E}\left(X_i|\mathcal{F}_{i+q}\right)\right|^p\right]\right)^{1/p} \le c_i\psi_q,$$

*for all $i$ and $q \ge 0$. Furthermore, we say $\{X_i\}$ is an $L_p$-mixingale of size $-d$ with respect to $\{\mathcal{F}_i\}$ if $\psi_q = O(q^{-d-\varepsilon})$ for some $\varepsilon > 0$. The same notation for the constants $c_i$ and sequence $\psi_q$ used in near-epoch dependence applies, due to the same role in both types of dependence.*

We also recall the properties of NED and mixingale sequences from Davidson (2002).

**Lemma 25.** *Let $\{X_i\}_{i=-\infty}^{\infty}$ be an $L_r$-bounded sequence, for $r > 1$ and $L_p$-NED of size $-b$ on a sequence $\{\boldsymbol{s}_i\}$ for $1 \le p \le r$ with non-negative constants $\{c'_i\}_{i=-\infty}^{\infty}$, if $\{\boldsymbol{s}_i\}$ is $\alpha$-mixing of size $-a$ and $p < r$, then $\{X_i - E[X_i], \mathcal{F}_{-\infty}^i\}$ is an $L_p$- mixingale of size $-\min\{b, a(1/p - 1/r)\}$ with constants $c_i \le \max\{c'_i, |X_i|_r\}$.*

This Lemma is from Theorem 18.6 (i) of Davidson (2002).

**Lemma 26.** *Let $X_i$ and $Y_i$ be $L_p$-NED on a sequence $\boldsymbol{s}_i$ of respective sizes $-d_1$ and $-d_2$. Then $X_i + Y_i$ is $L_p$-NED of size $-\min\{d_1, d_2\}$.*

This Lemma is from Theorem 18.8 of Davidson (2002).

**Lemma 27.** *Let $X_i$ and $Y_i$ be $L_p$-NED on a sequence $\boldsymbol{s}_i$ with $p \ge 2$ of respective sizes $-d_1$ and $-d_2$. Then $X_iY_i$ is $L_{p/2}$-NED of size $-\min\{d_1, d_2\}$.*

This Lemma is from Theorem 18.9 of Davidson (2002).

Due to the existence of the non-zero parameters, we define the weak sparsity index set

$$S_\lambda := \left\{ j : \left| \beta_j^0 \right| > \lambda \right\} \quad \text{with cardinality } |S_\lambda|, \tag{8.1}$$

for $\lambda \geq 0$, and its complement set $S_\lambda^c = \{1, \ldots, N\} \setminus S_\lambda$.

**Lemma 28.** *Suppose that Assumptions 8, 9 and 10 hold, and assume that*

$$\begin{aligned} 0 < r < 1: \quad & \lambda \geq C \ln(\ln(n))^{\frac{d+m-1}{r(dm+m-1)}} \left[ s_r \left( \frac{p^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}}{\sqrt{n}} \right)^{\frac{1}{\left(\frac{1}{d}+\frac{m}{m-1}\right)}} \right]^{\frac{1}{r}} \\ r = 0: \quad & s_0 \leq C \ln(\ln(n))^{-\frac{d+m-1}{dm+m-1}} \left[ \frac{\sqrt{n}}{p^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}} \right]^{\frac{1}{\left(\frac{1}{d}+\frac{m}{m-1}\right)}}, \\ & \lambda \geq C \ln(\ln(n))^{1/m} \frac{p^{1/m}}{\sqrt{n}}, \end{aligned} \tag{8.2}$$

*For* $C > 0$, *with probability at least* $1 - C\ln(\ln(n))^{-1}$, *we have*

$$\max_{1 \leq j \leq p} \max_{1 \leq k \leq n} \frac{1}{n} \sum_{i=1}^{k} \left| u_i X_i^{(j)} \right| \lesssim \frac{\lambda}{4},$$

*and*

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i X_i' - \frac{1}{n} \sum_{i=1}^{n} E\left[ X_i X_i' \right] \right\|_\infty \leq \frac{C}{|S_\lambda|}.$$

This Lemma is from Theorem 1 of Adamek et al. (2023), serving as the concentration inequality for dependent data.

**Lemma 29.** *Let* $\{X_i, \mathcal{F}_i\}$ *be an* $L^r$ *mixingale for some* $r > 1$ *and* $\sum_{q=1}^{\infty} \psi_q < \infty$. *Assume that* $E[X_i] = 0$. *Define* $S_k = \sum_{i=1}^{k} X_i$. *Then there exists a positive constant* $C$ *such that*

$$|| \max_{1 \leq k \leq n} |S_k| ||_r \leq C \left( \sum_{i=1}^{n} c_i^2 \right)^{1/2},$$

where $||X_i||_r = (E|X_i|^r)^{1/r}$.

This mixingale concentration inequality directly follows from Lemma 2 in Hansen (1991).

*Proof of Theorem 5.* The proof of Theorem 5 is similar to those of Theorems 1 and 2. We now apply the concentration inequality from Lemma 29, combined with Triplex inequality (Jiang (2009)), similarly to the proof of Lemma A.3 in Adamek et al. (2023), as a time series analogu of Lemma 2. Additionally, we use the concentration inequality from Lemma 29, combined with Markov's inequality, similarly to the proof of Lemma A.4 in Adamek et al. (2023), as a time series analogu Lemma 3. Meanwhile, we obtain that $\left\{X_i^{(j)}U_i\right\}$ and $X_i^{(j)}X_i^{(l)} - E\left[X_i^{(j)}X_i^{(l)}\right]$ are $L_m$-Mixingale sequences with respect to $\mathcal{F}_i = \sigma\{\boldsymbol{W}_i, \boldsymbol{W}_{i-1}, ...\}$, following Lemma A.1 and Lemma A.2 in Adamek et al. (2023) under Lemmas 25, 26 and 27. Furthermore, the proof of Lemma 28 follows from that of Theorem 1 in Adamek et al. (2023). With the additional Assumptions 3 and 4 for the well-defined threshold effect, we can thus establish the oracle inequalities in Theorem 5. □

*Proof of Theorem 6.* With the condition $s_{r,max}^{3/2} logp/\sqrt{n} \to 0$, we can obtain

$$\left| g'\widehat{\boldsymbol{\Theta}}(\widehat{\tau})(\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0) - \boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2} \right| = o_p(1)$$

by Lemma 22. Then, based on the oracle inequalities in Theorem 5, and combing the proof of Theorem 3 with the proof of Theorem 2 in Adamek et al. (2023), we thus establish the asymptotic normality of the debiased estimator. □

*Proofs of Theorem 7 and Theorem 8.* We can prove Theorem 7 by combiing the proof of Theorem 3 with that of Theorem 3 in Adamek et al. (2023). Additionally, we can prove Theorem 8 by combining the proof of Theorem 4 with Corollary 2 in Adamek et al. (2023). □

# References

Adamek, R., S. Smeekes, and I. Wilms (2023). Lasso inference for high-dimensional time series. *Journal of Econometrics 235*(2), 1114–1143.

Adamek, R., S. Smeekes, and I. Wilms (2024, 04). Local Projection Inference in High Dimensions. *The Econometrics Journal 00*, utae012.

Afonso, A. and J. Jalles (2013). Growth and productivity: The role of government debt. *International Review of Economics & Finance 25*(C), 384–407.

Barro, R. J. and J.-W. Lee (1994). Data set for a panel of 139 countries. Available at `http://admin.nber.org/pub/barro.lee/`.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics 37*(4), 1705 – 1732.

Callot, L., M. Caner, A. B. Kock, and J. A. Riquelme (2017). Sharp threshold detection based on sup-norm error rates in high-dimensional models. *Journal of Business & Economic Statistics 35*(2), 250–264.

Caner, M. and A. B. Kock (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics 203*(1), 143–168.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics 123*(1), 177–218.

Chan, K. S. (1993). Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *The Annals of Statistics 21*(1), 520 – 533.

Chan, K. S. and R. S. Tsay (1998, 06). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika 85*(2), 413–426.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *The annals of statistics 42*, 1564–1597.

Chernozhukov, V., D. Chetverikov, and K. Kato (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields 162*, 47–70.

Chiang, H. D., J. Rodrigue, and Y. Sasaki (2023). Post-selection inference in three-dimensional panel data. *Econometric Theory 39*(3), 623–658.

Chudik, A., K. Mohaddes, M. H. Pesaran, and M. Raissi (2017). Is there a debt-threshold effect on output growth? *The Review of Economics and Statistics 99*(1), 135–150.

Davidson, J. (2002). *Stochastic Limit Theory* (2nd ed.). Oxford: Oxford University Press.

Durlauf, S. N. and P. A. Johnson (1995). Multiple regimes and cross-country growth behaviour. *Journal of applied econometrics 10*(4), 365–384.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

Gold, D., J. Lederer, and J. Tao (2020). Inference for high-dimensional instrumental variables regression. *Journal of Econometrics 217*(1), 79–111.

Grennes, T., M. Caner, and F. Koehler-Geib (2010). *"Finding The Tipping Point – When Sovereign Debt Turns Bad"*. The World Bank.

Hansen, B. E. (1991). Strong laws for dependent heterogeneous processes. *Econometric Theory 7*(2), 213–221.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*(3), 575–603.

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and Its Interface 4*(2), 123–127.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research 15*(82), 2869–2909.

Jiang, W. (2009). On uniform deviations of general empirical risks with unboundedness, dependence, and high dimensionality. *Journal of Machine Learning Research 10*(36), 977–996.

Kock, A. B. (2016). Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models. *Journal of Econometrics 195*(1), 71–85.

Kock, A. B. and H. Tang (2019). Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects. *Econometric Theory 35*(2), 295–359.

Lee, S., M. H. Seo, and Y. Shin (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(1), 193–210.

Lee, Y. and Y. Wang (2023). Threshold regression with nonparametric sample splitting. *Journal of Econometrics 235*(2), 816–842.

Lu, T.-T. and S.-H. Shiou (2002). Inverses of 2× 2 block matrices. *Computers & Mathematics with Applications 43*(1-2), 119–129.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics 34*(3), 1436 – 1462.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Potter, S. M. (1995). A nonlinear approach to us gnp. *Journal of applied econometrics 10*(2), 109–125.

Ramey, V. A. and S. Zubairy (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of political economy 126*(2), 850–901.

Reinhart, C. M. and K. S. Rogoff (2010). Growth in a time of debt. *American Economic Review 100*(2), 573–78.

Semenova, V., M. Goldman, V. Chernozhukov, and M. Taddy (2023). Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence. *Quantitative Economics 14*(2), 471–510.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Tong, H. and K. S. Lim (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological) 42*(3), 245–268.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166 – 1202.

Yu, P. and X. Fan (2021). Threshold regression with a threshold boundary. *Journal of Business & Economic Statistics 39*(4), 953–971.