Uniform Inference in High-Dimensional Threshold Regression Models

Hongqiang Yan*†

October 18, 2023

Abstract

This paper addresses statistical inference for high-dimensional threshold regression parameters. I establish oracle inequalities for the scaled Lasso estimator proposed by Lee, Seo, and Shin, assuming only non-subgaussian error terms and covariates. Subsequently, I desparsify (or debias) the scaled Lasso estimator and derive the asymptotic distribution of tests involving an increasing number of slope parameters in the sense of van de Geer et al. (2014). Utilizing these results, I construct asymptotically valid confidence intervals for the components of the threshold regression slope coefficients. To complement the asymptotic theory in this paper, I conduct simulation studies to demonstrate the performance of our method in finite samples.
JEL classification: C12, C13, C24.

1 Introduction

Threshold models are a popular way to characterize nonlinearities in economic relationships. Hansen (1996) and Hansen (2000) show how the least squares estimation of threshold models is possible and feasible in fixed-dimensional settings, where the number of observations is much larger than the number of variables. These two papers develop a non-standard asymptotic theory of inference which allows for the construction of confidence intervals for the regression estimates, as well as testing of hypotheses for the presence of a threshold. Later, Caner and Hansen (2004) developed instrumental variable estimation techniques that allow for the covariates to be endogenous.

Let $\{(Y_i, X_i, Q_i) : i = 1, \dots, n\}$ be a sample of independent observations such that

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

where for each i, X_i is a $p \times 1$ vector, Q_i is a scalar, U_i is error terms, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. The scalar variable Q_i is the threshold variable determining regime switching and τ_0 is the unknown threshold parameter.

*I am grateful to my advisor, Mehmet Caner, for his guidance and advice. I thank, Barry Goodwin, Zheng Li and Ilze Kalnina for their helpful comments and suggestions. I also thank participants at the N.C. State University econometrics students workshop for discussion and comments. All remaining errors are my own.

†Nelson Hall 4331, North Carolina State University, 2801 Founders Drive, Raleigh, NC 27607. Email: hyan6@ncsu.edu

Threshold models have been well studied and applied in econometrics. In empirical studies, threshold models have been used to investigate the non-linearity in the threshold effect of government debt on economic output(e.g. Chudik et al. (2017), Afonso and Jalles (2013), Grennes et al. (2010)). Recently, there has been a growing interest in panel threshold models. Seo and Shin (2016) propose a two-step GMM estimator for the dynamic panel threshold model, which also allows for the endogeneity of either the covariates or the threshold variables. Miao et al. (2020) study estimation and inference in a panel threshold model in the presence of interactive fixed effects. Miao et al. (2020) consider latent group structures in a panel threshold regression model, which allows for the slope coefficients and threshold parameters to vary across individual units.

Interest in high-dimensional data has motivated much recent research on Lasso for threshold regression. Lee et al. (2016) establish sparsity oracle inequalities for the prediction norm and estimation error of the scaled Lasso applied to (1.1) in the case of fixed regressors and Gaussian error terms for both the no threshold effect case and the threshold effect case. In their simulation section, they also extended their results to random regressors with Gaussian errors. Callot et al. (2017) develop sup-norm oracle inequalities for the estimation error of the Lasso of Lee et al. (2016). Then they propose a thresholded scaled Lasso estimator based on the sup-norm bound to provide threshold selection consistency or even model selection consistency.

In the era of Big Data, we are interested in inference in high-dimensional settings where the number of parameters is much greater than the sample size. Our approach is an adaptation of the desparsifying a Lasso estimator introduced in van de Geer et al. (2014). Specifically, van de Geer et al. (2014) propose a desparsified Lasso estimator and construct asymptotically valid confidence bands for the estimated parameter. Similar advancements were made in the papers by Zhang and Zhang (2014) and Javanmard and Montanari (2014). The idea is to remove the bias introduced by shrinkage by desparsifying the estimator with a constructed approximate inverse of a singular sample covariance matrix for estimating high-dimensional regression models. Two approaches are widely used to construct the approximate inverse matrix: the nodewise regression introduced by Meinshausen and Bühlmann (2006) and the CLIME estimator of Cai et al. (2011).

Much of the present work is devoted to solving the inference problem in a high-dimensional linear model by desparsifying Lasso-type estimators to construct asymptotically valid confidence bands for the parameters of interest. Caner and Kock (2018) propose the conservative Lasso estimator allowing for non-identically distributed or non-sub-Gaussian error terms and develop the asymptotic distribution of tests involving an increasing number of parameters. Gold et al. (2020) propose a desparsified Lasso based on a two-stage least squares estimator with sub-Gaussian data and homoskedastic errors for a high-dimensional instrumental variables regression. They allow both the number of instruments and the number of regressors to be greater than the sample size. Another relevant paper is by Caner and Kock (2019), which develops a desparsified GMM estimator for estimating a high-dimensional instrumental variables regression that has many more endogenous regressors than observations. In their simulations, they compare it to the estimator in Gold et al. (2020). Belloni et al. (2019) provide a new way of handling linear and nonlinear instrumental variables regression as well as relaxing the sparsity assumption.

Present work by Belloni et al. (2014), Semenova et al. (2021) propose estimation methods for desparsification in treatment effects. In the context of generalized linear models, we refer to the articles by Belloni et al. (2016) and Caner (2021). Also, note that high-dimensional panel data models are considered in papers by Kock (2016) and Kock and Tang (2019).

Overall, we contribute to the literature in two ways. Our primary contribution is to develop a desparsified Lasso estimator for the threshold regression in the high-dimensional regime: $p >> n$—that is, if p, the number of variables is much larger than n, the number of observations. The estimator in Lee et al. (2016) may be desparsified in the sense of van de Geer et al. (2014) in order to construct asymptotically uniform confidence intervals for the parameters of interest and hypothesis tests under a sparse setting. Despite the considerable progress that has been made for inference in linear high-dimensional regression, only a few papers provide theoretical insights into more complex models, such as nonlinear models. Another contribution involves extending oracle inequalities for the Lasso estimator for high-dimensional threshold regression to non-subgaussian error terms and regressors, using the maximal inequalities by Chernozhukov et al. (2017). Strengthening our assumption of sub-Gaussianity could deliver even stronger results.

The rest of the paper is organized as follows. Section 2 recalls the Lasso estimator of Lee et al. (2016). In Section 3 we develop oracle inequalities for the Lasso estimator of regression slopes as well as the threshold estimator only assuming non-sub-Gaussian error terms and regressors. In section 5 we propose a desparsified Lasso estimator for the high-dimensional threshold regression model and derive the asymptotic distribution of hypothesis tests for slope parameters based on an adaptation of the work in van de Geer et al. (2014). In section 6 we investigate the finite sample properties of the desparsified Lasso for threshold models and compare it to the desparsified Lasso estimator for linear models of van de Geer et al. (2014). All proofs are deferred to the Appendix.

2 The Model

Notation

For $\{(Y_i, X_i, Q_i) : i = 1, \ldots, n\}$ following (1.1), let bold font $\mathbf{X}_i(\tau)$ denote the $(2p \times 1)$ vector such that $\mathbf{X}_i(\tau) = (X_i', X_i'\mathbf{1}\{Q_i < \tau\})'$ and let $\mathbf{X}(\tau)$ denote the $(n \times 2p)$ matrix whose i-th row is $\mathbf{X}_i(\tau)'$. Let X and $X(\tau)$ denote the first and last p columns of $\mathbf{X}(\tau)$, respectively.

For any $L \times 1$ real vector a, let $\|a\|_q$ denote the ℓ_q norm of a. Particularly, if $a = (a_1, \ldots, a_n)'$, n-dimensional vector, the prediction norm is defined as $\|a\|_n := \sqrt{\frac{1}{n}\sum_{i=1}^n a_i^2}$.

Also, let $J(a) := \{j \in \{1, \ldots, L\} : a_j \neq 0\}$ and let $|J(a)|$ denote the cardinality of $J(a)$. Let $\mathcal{M}(a)$ denote the number of nonzero elements of a, i.e. $\mathcal{M}(a) = |J(a)|$. Then we let $b_J \in \mathbb{R}^L$ denote the vector has the same coordinates as a on J and zero coordinates on the complement J^C. Let the superscript $^{(j)}$ denote the j-th element of a vector or the j-th column of a matrix depending on the context.

For any $m \times n$ matrix A, we define $\|A\|_\infty := \max_{1\le i\le m, 1\le j\le n} |A_{ij}|$. $\|A\|_{l_\infty} := \max_{1\le i\le m} \sum_{j=1}^n |A_{ij}|$ denotes the induced l_∞-norm of A. Similarly, $\|A\|_{l_1} := \max_{1\le j\le n} \sum_{i=1}^m |A_{ij}|$ denotes the induced

l_1-norm of A.

Finally, define $f_{(\alpha,\tau)}(x,q) := x'\beta + x'\delta 1\{q < \tau\}$, $f_0(x,q) := x'\beta_0 + x'\delta_0 1\{q < \tau_0\}$, and $\widehat{f}(x,q) := x'\widehat{\beta} + x'\widehat{\delta}1\{q < \widehat{\tau}\}$. Then, we define the prediction norm as $\left\|\widehat{f} - f_0\right\|_n := \sqrt{\frac{1}{n}\sum_{i=1}^n \left(\widehat{f}(X_i,Q_i) - f_0(X_i,Q_i)\right)^2}$. Throughout the paper, we use the superscript zero to signify the true parameter value.

2.1 Lasso Estimation

We consider the model in (1.1). It can be written 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 in the above model is used to split the sample into two groups. When $Q_i < \tau_0$, the regression function becomes $X_i'(\beta_0 + \delta_0) + U_i$; if $Q_i \geq \tau_0$, the regression function reduces to $X_i'\beta_0 + U_i$. As δ_0 is the change of regression coefficients between two regimes, the model in (1.1) captures a regime switch based on an observable scalar variable Q_i with a scalar unknown parameter τ_0. The case of $\delta_0 = 0$ corresponds to the linear model. If $\widehat{\delta} = 0$, then this case amounts to selecting the linear model.

Recall the model in Lee et al. (2016). Further assumptions in the model are detailed in Section 3. Let $\alpha_0 = (\beta_0', \delta_0')'$. Then, using notation defined above, we can rewrite (1.1) as

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

α_0 is the $2p \times 1$ population vector of coefficients, which we shall assume to be sparse. However, the location of the non-zero coefficients is unknown and potentially $2p$ could be much greater than n. We assume that the explanatory variables are exogenous and precise assumptions will be made in Assumption 1 below. Let $J_0 = J(\alpha_0)$, denote the set of non-zero coefficients and $s_0 = |J_0|$, the cardinality. In this paper, we study the high-dimensional case where p is much greater than n.

Let $\mathbf{Y} := (Y_1, \ldots, Y_n)'$. For any $\tau \in \mathbb{T}$, where $\mathbb{T} := [t_0, t_1]$ is a parameter space for τ_0, consider the residual sum of squares

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

where $\alpha = (\beta', \delta')'$.

The scaled Lasso for threshold regression is defined as the one-step minimizer such that:

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

where $\mathcal{A}$ is a parameter space for α_0 and λ is a tuning parameter chosen by the researcher which we discuss further in Section 3. The $(2p \times 2p)$ diagonal weighting matrix is denoted as follows:

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

where $\mathbf{X}^{(j)}(\tau)$ denotes the j-th column of $\mathbf{X}(\tau)$. Furthermore, we can rewrite the ℓ_1 penalty as

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

To be more exact, $(\widehat{\alpha}, \widehat{\tau})$ in (2.4) can be regarded as a two-step minimizer such that:

Step 1.

For each $\tau \in \mathbb{T}$, $\widehat{\alpha}(\tau)$ is defined as

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

Step 2.

Define $\widehat{\tau}$ as the estimator of τ_0 such that:

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

It is worth mentioning that $\widehat{\alpha}(\tau)$ is the weighted Lasso that uses a data-dependent ℓ_1 penalty to balance covariates adequately. Additionally, $\widehat{\tau}$ is an interval and in accordance with Lee et al. (2016), we define the maximum of the interval as the estimator $\widehat{\tau}$. For any n, it suffices in practice to search over $Q_1, \cdots Q_n$ as candidates for $\widehat{\tau}$, as these are the points where $1\{Q_i > \tau\}, i = 1, \cdots n$ will change. To put it another way, we think the parameter space $\mathbb{T}$ is divided into n intervals depending on $Q_1, \cdots Q_n$.

3 Oracle inequalities

In this section, we establish the oracle inequality for the scaled Lasso estimator in (2.4). As we are considering a random design as opposed to a fixed regressor design in Lee et al. (2016), our assumptions are imposed in a slightly different form. Note Lee et al. (2016) have already argued how some of their assumptions could be valid in a random design.

Assumption 1. *Let $\{X_i, U_i, Q_i\}_{i=1}^n$ be an i.i.d. sample;*

(i)For the parameter space $\mathcal{A}$ for α_0, any $\alpha \equiv (\alpha_1, \cdots, \alpha_{2p}) \in \mathcal{A} \subset \mathbb{R}^{2p}$,including α_0, satisfies $\|\alpha\|_\infty \leq C_1$, for some constant $C_1 > 0$. $\mathcal{M}(\alpha_0) \leq s_0$ and $\frac{s_0^2 \|\delta_0\|_1^2 \log p}{n} = o_p(1)$.

(ii) Marginal distribution of Q_i is uniform $(0, 1)$. $\tau_0 \in \mathbb{T} = [t_0, t_1]$ with $0 < t_0 < t_1 < 1$.

(iii) $\max_{1\le j\le p} E\left[(X_i^{(j)})^4\right] \le C_2^4$ *and* $\min_{1\le j\le p} E\left[(X_i^{(j)}(t_0))^2\right] \ge C_3^2$ *uniformly in* n *for some universal constants* C_2 *and* C_3. $E\left[X_i^{(j)}X_i^{(l)}|Q_i=\tau\right]$ *is continuous and bounded when* τ *is in a neighborhood of* τ_0 *for all* $1\le j,l\le p$. *(iv) The error terms* $E(U_i|X_i,Q_i)=0$ *and* $E(U_i^2)=\sigma^2$.

(v) $\frac{\sqrt{EM_{UX}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$ *where* $M_{UX} = \max_{1\le i\le n}\max_{1\le j\le p}|U_iX_i^{(j)}|$.

(vi) $\frac{\sqrt{EM_{XX}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$ *where* $M_{XX} = \max_{1\le i\le n}\max_{1\le j,l\le p}|X_i^{(j)}X_i^{(l)} - E[X_i^{(j)}X_i^{(l)}]|$.

(vii) $\frac{\sqrt{EM_{Xt_0}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$, *where* $M_{Xt_0} = \max_{1\le i\le n}\max_{1\le j\le p}|(X_i^{(j)}(t_0))^2 - E(X_i^{(j)}(t_0))^2|$.

Assumption 1 (i) imposes restrictions for each component of the slope parameter vector. The second part of Assumption 1 (i) implies that s_0 and $\|\delta_0\|_1$ can increase with n.

Next, we describe how to solve the problem where the distribution of the threshold variable is not uniform. This technique is based on empirically transforming the distribution of the threshold variables to a uniform distribution. Suppose that the threshold variable $\{\tilde{Q}\}$ has a continuous distribution for which the cumulative distribution function is $F_{\tilde{Q}}$. The probability integral transform implies that the random variable Q has a standard uniform distribution where Q is defined as $Q = F_{\tilde{Q}}(\tilde{Q})$. To transform the marginals, we compute $Q_i = \widehat{F}_{\tilde{Q}}(\tilde{Q}_i) = \frac{\text{rank of } \tilde{Q}_i \text{ among } \left\{\tilde{Q}_i\right\}_{i=1}^n}{n}$, where $\widehat{F}_{\tilde{Q}}$ denotes the empirical distribution functions of the data $\left\{\tilde{Q}_i\right\}_{i=1}^n$. In particular, as a result of a continuous distribution, there is no tie among $\left\{\tilde{Q}_i\right\}_{i=1}^n$. We will show that the performance of our estimator does not depend on whether the threshold variable(Q_i) is part of the set of covariates(X_i) or correlated with the covariates in Section 6.

Assumption 1 (iii) to (vii) states restrictions on the covariates as well as the error terms in the random design setup studied in this article. Compared to Assumption 1 in Lee et al. (2016), we only assume the covariates and error terms are independently and identically distributed with uniformly bounded certain moments instead of sub-Gaussian data (Callot et al. (2017)) due to Chernozhukov et al. (2017). That is a much stronger assumption than the one imposed here and rules out data with heavy tails. Assumption 1 (ii) implies that $\min_{i=1,\cdots n} Q_i < t_0$. Intuitively, we assume that $\min_{1\le j\le p} E\left[(X_i^{(j)}(t_0))^2\right]$ is bounded away from 0.

Assumption 1 (iii) is a much stronger assumption than necessary conditions for the maximal inequality due to Chernozhukov et al. (2017). Apply the Cauchy-Schwarz Inequality to obtain the following:

(i) $\max\limits_{1\le j,l\le p} E\left[X_i^{(j)}X_i^{(l)}\right] \le C_2^2$ uniformly in n;

(ii) $\max\limits_{1\le j\le p} var(U_iX_i^{(j)})$, $\max\limits_{1\le j\le p} var(U_iX_i^{(j)}1(Q_i<\tau))$, $\max\limits_{1\le j,l\le p} var(X_i^{(j)}X_i^{(l)})$, $\max\limits_{1\le j,l\le p} var(X_i^{(j)}X_i^{(l)})$, $\max\limits_{1\le j,l\le p} var(X_i^{(j)}X_i^{(l)}1(Q_i<\tau))$, $\max\limits_{1\le j,l\le p} var(X_i^{(j)}X_i^{(l)}1(Q_i<\tau))$ and $\max\limits_{1\le j\le p} var(X_i^{(j)}(t_0))^2$are bounded away from infinity uniformly in n.

Assumption 1 is used to establish the oracle inequality in Lemma 1, Theorem 1 and 2.

Now define

$$\lambda = \frac{C}{\mu}\frac{\sqrt{\log p}}{\sqrt{n}} \tag{3.1}$$

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

Lemma 1. *Under Assumption 1, let $(\widehat{\alpha}, \widehat{\tau})$ be the Lasso estimator defined by (2.4) with $\lambda = \frac{C}{\mu}\frac{\sqrt{\log p}}{\sqrt{n}}$ for a constant C and a fixed constant $\mu \in (0,1)$. Then, with probability approaching 1* [1] *we have*

$$\left\| \widehat{f} - f_0 \right\|_n \le \sqrt{(6+2\mu) C_1 \sqrt{C_2^2 + \mu\lambda}}\sqrt{s_0 \lambda}. \tag{3.2}$$

Lemma 1 states that regardless of the linearity of the model, the prediction norm of the scaled Lasso estimator defined by (2.4) converges to 0, provided that $n \to \infty$, $p \to \infty$ and $s_0\lambda \to 0$. This, in turn, plays an important role for proving the oracle inequality in Theorem 1 for the case of linear models and Theorem 2 for nonlinear models.

Next, we turn towards the standard assumptions in high-dimensional regression models. To this end, define the population covariance matrix $\mathbf{\Sigma}(\tau) := E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]$, $\mathbf{M} := E({X_i}'X_i)$, $\mathbf{M}(\tau) := E\left[X_i(\tau)'X_i(\tau)\right]$ and $\mathbf{N}(\tau) := \mathbf{M} - \mathbf{M}(\tau)$. Then, $\mathbf{\Sigma}(\tau)$ can be represented by a 4-block matrix, i.e.

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

The population uniform adaptive restricted eigenvalue is denoted by

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

or

$$\kappa(s_0, c_0, \mathbb{S}, \mathbf{M}) = \min_{\tau \in \mathbb{S}} \min_{J_0 \subset \{1,\dots,2p\}, |J_0| \le s_0} \min_{\gamma \neq 0, \|\gamma_{J_0^c}\|_1 \le c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{\left(\gamma' E\left[X_i(\tau)'X_i(\tau)\right]\gamma\right)^{1/2}}{\|\gamma_{J_0}\|_2}.$$

or

$$\kappa(s_0, c_0, \mathbf{M}) = \min_{J_0 \subset \{1,\dots,2p\}, |J_0| \le s_0} \min_{\gamma \neq 0, \|\gamma_{J_0^c}\|_1 \le c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{\left(\gamma' E\left[{X_i}'X_i\right]\gamma\right)^{1/2}}{\|\gamma_{J_0}\|_2}.$$

depending on the context.

In the literature on high-dimensional econometrics and statistics, it is common to add an adaptive restricted eigenvalue condition. Additionally, we consider that an adaptive restricted eigenvalue is of the same magnitude uniformly over $\tau \in \mathbb{T}$ as follows

[1] at least $1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{EM_{X^2}^2}{n \log p}\right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{EM_{Xt_0}^2}{n \log p}\right) - \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM_{UX}^2}{n \log p}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM_{UX}^2}{n \log (pn)}\right)$, for some universal positive constants $\tilde{C}_1 \cdots \tilde{C}_8$.

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

$$\kappa(s_0, c_0, \mathbb{S}, \mathbf{\Sigma}) > 0. \tag{3.3}$$

Assumption 2 (i) is a standard assumption in regression models. One can provide sufficient conditions for Assumption 2 (ii) by imposing the condition that the population covariance matrix $\mathbf{\Sigma}(\tau)$ have full rank. Hence, we are interested in property of $\mathbf{\Sigma}(\tau)$. To solve the inverse of the population covariance matrix, we do the Gaussian elimination to get

$$\mathbf{\Sigma}(\tau)^{-1} = \begin{bmatrix} \mathbf{N}(\tau)^{-1} & -\mathbf{N}(\tau)^{-1} \\ -\mathbf{N}(\tau)^{-1} & \mathbf{M}(\tau)^{-1} + \mathbf{N}(\tau)^{-1} \end{bmatrix}, \tag{3.4}$$

provided that $\mathbf{M}$, $\mathbf{M}(\tau)$ and $\mathbf{N}(\tau)$ are non-singular. Therefore, $\mathbf{\Sigma}(\tau)$ has full rank as long as $\mathbf{M}$, $\mathbf{M}(\tau)$ and $\mathbf{N}(\tau)$ are invertible. Thus, Assumption 2 (ii) is almost automatic under non-singularity conditions for $\mathbf{M}$, $\mathbf{M}(\tau)$ and $\mathbf{N}(\tau)$.

We will show that $\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)$ uniformly converges to $\mathbf{\Sigma}(\tau)$ under Assumption 1 in Lemma 6 in the Appendix. Thus the empirical adaptive restricted eigenvalue condition can hold under the population eigenvalue condition imposed here, which can be seen in Lemma 7 in the Appendix.

Considering τ_0 is unknown, we impose that the restricted eigenvalue condition holds uniformly over τ. Intuitively, $\delta_0 \neq 0$ is a necessary condition of identifiability of τ_0. If $\delta_0 = 0$, we have to assume Assumption 2 holds with $\mathbb{S} = \mathbb{T}$, the whole parameter space for τ_0. By contrast, it suffices to impose the Adaptive Restricted Eigenvalue Condition holding uniformly on a neighborhood of τ_0, when $\delta_0 \neq 0$.

The Uniform Adaptive Restricted Eigenvalue Condition is crucial for us to update the boundness in Lemma 1. Lemma 1 states that the prediction norm is bounded by a factor of $s_0\lambda$. This bound is larger than what is desired for an oracle inequality. Depending on the UARE condition, the prediction norm as well as the ℓ_1 estimation error will be further tightened in the next section. Lee et al. (2016) proposed a type of slope estimator that is not affected by the presence of a threshold effect. That is to say, we can make predictions and estimate α_0 even if $\delta_0 = 0$ does not hold. However, we can derive oracle inequalities in terms of the prediction error and the ℓ_1 estimation error for unknown parameters α_0 separately in two cases depending on whether $\delta_0 = 0$ or not.

3.1 Case I: no threshold

First, we consider the situation where $\delta_0 = 0$. In this case, the true model is a linear model $Y_i = X_i'\beta_0 + U_i$, but we estimate it using the method defined by (2.4). Our estimated model is much more over-parametrized than the true one, but we shall obtain relatively precise estimates for the slope parameter vector α_0.

Theorem 1. *Supposed that $\delta_0 = 0$, let Assumptions 1-2 hold with $\kappa = \kappa(s_0, \frac{1+\mu}{1-\mu}, \mathbb{T}, \Sigma)$ for $0 < \mu < 1$. Let $(\widehat{\alpha}, \widehat{\tau})$ be the Lasso estimator defined by (2.4) with λ given by (3.1). Then for all sufficiently large n, with probability approaching 1 [2] we have*

$$\left\| \widehat{f} - f_0 \right\|_n \leq \frac{2\sqrt{2}}{\kappa} \left(\sqrt{C_2^2 + \mu\lambda} \right) \sqrt{s_0}\lambda,$$

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

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

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

It is worth noting that the bound of the prediction norm here is much smaller than in Lemma 1. Compared to Theorem 2 in Lee et al. (2016) or the oracle inequality in the literature on high-dimensional linear models (Bickel et al. (2009), van de Geer et al. (2014) etc.), Theorem 1 delivers results of the same magnitude. Although our model is much more overparametrized, our estimation method can accommodate the linear model. Nonetheless, there is a variable selection problem on δ_0. Our estimation method can find more nonzero coefficients than the true number. In particular,, some $\delta(\widehat{\tau})_j$ is incorrectly estimated as nonzero. We shall discuss this in more detail in Section 4.

3.2 Case II: fixed threshold

In this subsection, we construct oracle inequalities when $\delta_0 \neq 0$. More explicitly, the true model has a well-identified and discontinuous threshold effect.

Assumption 3 (Identifiability under Sparsity and Discontinuity of Regression)**.** *For a given $s_0 \geq \mathcal{M}(\alpha_0)$, and for any η and τ 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 wpa1*

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

Assumption 3 states identifiability of τ_0. Lee et al. (2016) have already discussed in Appendix B.1. (page. A7–A8) that Assumption 3 is valid in a random design under Assumption 1 above. As mentioned before, we need Assumption 2 to hold uniformly in a neighborhood of τ_0. Lemma 9 shows how we can get an upper bound of $\widehat{\tau} - \tau_0$ only under Assumption 1 and 3.

Given Lemma 9 in the appendix, we define

$$\eta^* = \frac{2(3+\mu)C_1}{C_4} \sqrt{C_2^2 + \mu\lambda} s_0\lambda$$

[2] at least $1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{EM_{X^2}^2}{n \log p} \right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{EM_{Xt_0}^2}{n \log p} \right) - \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM_{UX}^2}{n \log p} \right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM_{UX}^2}{n \log(pn)} \right) - \left(\frac{1}{p^{2\tilde{C}_9}} + \tilde{C}_{10} \frac{EM_{XX}^2}{n \log p^2} \right) - \left(\frac{1}{(p^2 n)^{\tilde{C}_{11}}} + \tilde{C}_{12} \frac{EM_{XX}^2}{n \log(p^2 n)} \right)$, for some universal positive constants $\tilde{C}_1 \cdots \tilde{C}_{12}$.

and

$$\mathbb{S} = \{|\tau - \tau_0| \leq \eta^*\},$$

where $\mathbb{S}$ can be inserted into Assumption 2. Note that we omit the restriction, $\eta \geq \min_i |Q_i - \tau_0|$ which is imposed in Lee et al. (2016). The reason is that $\eta \geq \min_i |Q_i - \tau_0|$ never binds for sufficiently large n. Intuitively, $\min_i |Q_i - \tau_0|$ will be small enough in the random design.

Assumption 4 (Smoothness of Design)**.** *For any $\eta > 0$, there exists a constant $C_5 < \infty$ such that wpa1*

$$\sup_{1\leq j,l\leq p} \sup_{|\tau-\tau_0|<\eta} \frac{1}{n}\sum_{i=1}^{n} \left|X_i^{(j)}X_i^{(l)}\right| \left|1\left(Q_i<\tau_0\right) - 1\left(Q_i<\tau\right)\right| \leq C_5\eta, \tag{3.5}$$

$$\sup_{1\leq j,l\leq p} \sup_{|\tau-\tau_0|<\eta} \|\delta_0\|_1 \left|\frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)} \left[1\left(Q_i<\tau_0\right) - 1\left(Q_i<\tau\right)\right]\right| \leq \frac{\lambda\sqrt{\eta}}{2}, \tag{3.6}$$

$$\sup_{|\tau-\tau_0|<\eta} \left|\frac{1}{n}\sum_{i=1}^{n} U_i X_i'\delta_0 \left[1\left(Q_i<\tau_0\right) - 1\left(Q_i<\tau\right)\right]\right| \leq \frac{\lambda\sqrt{\eta}}{2}. \tag{3.7}$$

Lemma 4 shows that $\sup_{1\leq j\leq p} \frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)}$ is bounded by λ wpa1. Correspondingly, Lemma 6 reveals that $\sup_{1\leq j,l\leq p} |\frac{1}{n}\sum_{i=1}^{n} X_i^{(j)}X_i^{(l)}|$ is bounded from above wpa1. The supremum in Assumption 4 above is bounded in a neighborhood of τ_0 for all $1 \leq j, l \leq p$. This strengthening is essential to establish oracle inequalities where a threshold is present. Note that (3.7) implies (3.6) almost automatically.

Theorem 2. *Suppose that $\delta_0 \neq 0$, let Assumption1 to 2 hold with $\kappa = \kappa(s_0, \frac{2+\mu}{1-\mu}, \mathbb{S}, \mathbf{\Sigma})$ for $0 < \mu < 1$. Furthermore, Assumptions 3 and 4 hold. Let $(\widehat{\alpha}, \widehat{\tau})$ be the Lasso estimator defined by (2.4) with λ given by (3.1).*

Then for all sufficiently large n, with probability approaching 1 [3] we have

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

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

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

[3] at least $1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2\frac{EM_{X^2}^2}{n\log p}\right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4\frac{EM_{Xt_0}^2}{n\log p}\right) - \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6\frac{EM_{UX}^2}{n\log p}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8\frac{EM_{UX}^2}{n\log(pn)}\right) - \left(\frac{1}{p^{2\tilde{C}_9}} + \tilde{C}_{10}\frac{EM_{XX}^2}{n\log p^2}\right) - \left(\frac{1}{(p^2n)^{\tilde{C}_{11}}} + \tilde{C}_{12}\frac{EM_{XX}^2}{n\log(p^2n)}\right)$, for some universal positive constants $\tilde{C}_1 \cdots \tilde{C}_{12}$.

Theorem 2 provides the same oracle inequalities (ignoring the constant terms) as those in Theorem 1 in terms of prediction norm and ℓ_1 errors for estimates. For the super-consistency result of $\widehat{\tau}$, Lee et al. (2016) argued that the least squares objective function behaving locally linearly around the true threshold parameter value is the key to achieving the super-consistency for the threshold parameter.

The main contribution of this section is that we have extended the oracle inequality to non-sub-Gaussian random regressors with non-sub-Gaussian errors for both the prediction norm and ℓ_1 errors for estimates.

4 Model Selection Consistency

So far we have established oracle inequalities for the prediction norm and ℓ_1 errors for estimates. Before we turn towards desparsifying the estimator in order to construct tests and confidence intervals, we consider the variable selection issue. If the true model is a linear model, van de Geer et al. (2014) has solved the desparsified Lasso estimation. Otherwise, we need to propose the desparsified Lasso when the threshold effect is well-identified and discontinuous. To this end, we impose some assumptions under which the Lasso estimator defined by (2.4) correctly estimates $\delta_0 = 0$ when the true model is linear.

The situation where $\delta_0 = 0$ is non-trivial since the consistency of an estimator does not provide selection consistency. Suppose $\delta_0 = 0$, Theorem 1 shows that

$$\widehat{\delta}^{(j)}(\widehat{\tau}) \xrightarrow{p} 0,$$

for each $j \in \{1, \cdots p\}$. However, this does not imply that we will correctly estimate zero coefficients as zero. The consistency implies that for all $\varepsilon > 0$,

$$\mathbb{P}\left\{|\widehat{\delta}^{(j)}(\widehat{\tau})| \geq \varepsilon\right\} \to 0$$

But as we need to control the correct model, we instead require

$$\mathbb{P}\left\{\widehat{\delta}^{(j)}(\widehat{\tau}) = 0\right\} \to 1. \tag{4.1}$$

(4.1) states that, with a consistent estimator, selection consistency comes from (4.1). In particular, Lasso has a tendency to overshoot the correct model, finding more nonzero coefficients than the true number. Strictly speaking, if the estimated number of nonzero coefficients is $\widehat{s}$, then in finite samples Lasso has a tendency to obtain $\widehat{s} > s_0$. To our scaled threshold model, the Lasso estimator defined in (2.4) may be much more over-parameterized in that τ and δ are added to β as parameters.

We next turn to variable selection by means of thresholding. For this purpose, we follow Callot et al. (2017) to define the thresholded Lasso estimator[4] as

$$\tilde{\delta}^{(j)}(\hat{\tau}) = \begin{cases} \hat{\delta}^{(j)}(\hat{\tau}), & \text{if } |\hat{\delta}^{(j)}(\hat{\tau})| \geq H, \\ 0, & \text{if } |\hat{\delta}^{(j)}(\hat{\tau})| < H. \end{cases} \tag{4.2}$$

where H is the threshold determining whether a coefficient should be classified as zero or nonzero and $\hat{\delta}^{(j)}(\hat{\tau})$ are elements of the Lasso estimator defined by (2.4). In particular, we shall see that choosing $H = 2C\lambda$ results in consistent model selection.

Theorem 3 (Consistency in threshold selection)**.** *Suppose that $\delta_0 = 0$, i.e., the true model is linear. Let Assumptions 1 hold. Let $\tilde{\delta}^{(j)}(\hat{\tau})$ be the thresholded Lasso estimator defined by (4.2) with λ given by (3.1). Then, there exists a C such that for $H = 2C\lambda$*

$$\mathbb{P}\left\{\tilde{\delta}^{(j)}(\hat{\tau}) = 0, \forall j \in \{1 \cdots p\}\right\} \to 1$$

as $n \to \infty$.

Theorem 3 mentioned above is derived from Theorem 4 in Callot et al. (2017). Consistency in threshold selection, while less demanding than model selection consistency, ensures the estimation of δ_0 as 0 under the linear model. Nevertheless, it provides a direct response to whether the presence or absence of a threshold exists (i.e., whether the model is linear or nonlinear). The discussion on choosing the thresholding parameter C through the Bayesian Information Criterion (BIC) is omitted, as it is similarly implemented in the simulation section of Callot et al. (2017).

5 The Desparsified Lasso

We now introduce the desparsified Lassso estimator that we use to construct tests and confidence intervals.

5.1 The Desparsified Lasso

Suppose we estimated some $\tilde{\delta}^{(j)}(\hat{\tau}) \neq 0$, Theorem 3 reveals that our true model is a nonlinear threshold model($\delta_0 \neq 0$).

In connection with Theorem 2 we note that $|\hat{\tau} - \tau_0|$ is bounded by $\eta_0 = \left(\frac{3(1+\mu)\sqrt{(C_2^2+\mu\lambda)}}{(1-\mu)\sqrt{(C_3^2-\mu\lambda)}} + 1\right) \frac{12(C_2^2+\mu\lambda)}{\kappa^2 C_4} s_0\lambda^2$, i.e. a factor of $\frac{s_0 \log p}{n}$. We define

$$\hat{\mathbb{T}} = \{\tau \in \mathbb{T} \mid \tau_0 - \tilde{C}\frac{s_0 \log p}{n} \leq \tau \leq \tau_0 + \tilde{C}\frac{s_0 \log p}{n}\} \subseteq \mathbb{S},$$

where $\hat{\mathbb{T}}$ is the neighborhood of τ_0 and constant $\tilde{C}$ can be chosen depending on η_0. The estimated $\hat{\tau}$ falls in $\hat{\mathbb{T}}$ w.p.a.1.

[4]Note that since we are only interested in finding out whether δ_0 is nonzero or not, one can simply threshold $\hat{\delta}$.

Next, we turn to desparsify the Lasso estimator defined by (2.4) and derive the asymptotic distribution of the slope estimator $\widehat{\alpha}(\widehat{\tau})$. Considering any $\widehat{\tau} \in \widehat{\mathbb{T}}$ we will show how the desparsification proposed in van de Geer et al. (2014) works in our context. The idea is that the shrinkage bias introduced due to the presence of penalization in (2.6) will show up in the properly scaled limiting distribution of $\widehat{a}^{(j)}(\widehat{\tau})$. Hence, we remove this bias prior to conducting statistical inference. Recall that $\mathbf{D}(\tau)$ is defined in (2.5). By the minimizing property of $\widehat{\alpha}(\widehat{\tau})$, it is followed by the first order condition of (2.6):

$$-\mathbf{X}(\widehat{\tau})'(Y - \mathbf{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}))/n + \lambda \mathbf{D}(\widehat{\tau})\hat{\rho} = 0, \tag{5.1}$$

where $\hat{\rho}$ is a $2p$ by 1 vector, arising from the subdifferential of $\|\widehat{\alpha}(\widehat{\tau})\|_1 . \|\hat{\rho}\|_\infty \leq 1$ and $\hat{\rho}_j = \text{sign}(\widehat{a}^{(j)}(\widehat{\tau}))$ if $\widehat{a}^{(j)}(\widehat{\tau}) \neq 0$, where "sign()" is the function which maps positive entries to 1, negative entries to -1.

Defining $\widehat{\Sigma}(\widehat{\tau}) = \mathbf{X}'(\widehat{\tau})\mathbf{X}(\widehat{\tau})/n$ and inserting $Y = \mathbf{X}(\tau_0)\alpha_0 + U$, then the above expression yields

$$\lambda \mathbf{D}(\widehat{\tau})\hat{\rho} + \widehat{\Sigma}(\widehat{\tau})\hat{\alpha}(\widehat{\tau}) - \mathbf{X}'(\widehat{\tau})\mathbf{X}(\tau_0)\alpha_0/n = \mathbf{X}'(\widehat{\tau})U/n.$$

In order to separate $\hat{\alpha}(\widehat{\tau})$ from $\widehat{\Sigma}(\widehat{\tau})\hat{\alpha}(\widehat{\tau})$, we have to left-multiply $\widehat{\Sigma}(\widehat{\tau})\hat{\alpha}(\widehat{\tau})$ by the inverse of $\widehat{\Sigma}(\widehat{\tau})$. However, this is not feasible. $\widehat{\Sigma}(\tau)$ is of reduced rank provided that $2p > n$. Thus, the idea is to construct an approximate inverse, $\hat{\Theta}(\hat{\tau})$, to $\hat{\Sigma}(\widehat{\tau})$ and control the error term resulting from this approximation. Our construction of this approximate inverse relies on nodewise regression of Yuan (2010), which will be introduced in the next subsection. Multiplying all terms in the above equation by $\widehat{\Theta}(\widehat{\tau})$, and adding $\hat{\alpha}(\widehat{\tau}) - \alpha_0 + \widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})\alpha_0$ to both sides to get,

$$\begin{aligned}
&\hat{\alpha}(\widehat{\tau}) - \alpha_0 + \widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})\alpha_0 + \hat{\Theta}(\widehat{\tau})\lambda\mathbf{D}(\widehat{\tau})\hat{\rho} + \hat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})\hat{\alpha}(\widehat{\tau}) - \hat{\Theta}(\widehat{\tau})\mathbf{X}'(\widehat{\tau})\mathbf{X}(\tau_0)\alpha_0/n \\
=&\hat{\Theta}(\widehat{\tau})\mathbf{X}'(\widehat{\tau})U/n + \hat{\alpha}(\widehat{\tau}) - \alpha_0 + \widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})\alpha_0.
\end{aligned}$$

The above equation can be rewritten as

$$\hat{\alpha}(\widehat{\tau}) = \alpha_0 + \widehat{\Theta}(\widehat{\tau})\mathbf{X}'(\widehat{\tau})\left(\mathbf{X}(\tau_0)\alpha_0 - \mathbf{X}(\widehat{\tau})\alpha_0\right)/n - \hat{\Theta}(\widehat{\tau})\lambda\mathbf{D}(\widehat{\tau})\hat{\rho} + \hat{\Theta}(\widehat{\tau})\mathbf{X}'(\widehat{\tau})U/n - \Delta(\widehat{\tau})/n^{1/2}, \tag{5.2}$$

where

$$\Delta(\tau) = \sqrt{n}(\hat{\Theta}(\tau)\hat{\Sigma}(\tau) - I_{2p})(\hat{\alpha}(\tau) - \alpha_0),$$

which can be interpreted as the approximation error due to using an approximate inverse instead of an exact inverse. The details of asymptotic negligibility of $\Delta(\widehat{\tau})$ are given in Lemma 17 in Appendix.

In general, we need our estimate to be decomposable into terms that are either asymptotically negligible or asymptotically convergent. $\hat{\Theta}(\widehat{\tau})\lambda\mathbf{D}(\widehat{\tau})\hat{\rho}$ in (5.2) is the shrinkage bias introduced due to the presence of penalization in (2.6).

By adding $\hat{\Theta}(\widehat{\tau})\lambda\mathbf{D}(\widehat{\tau})\hat{\rho}$ to both sides of (5.2),

(5.3)
$$\hat{\alpha}(\widehat{\tau}) + \hat{\Theta}(\widehat{\tau})\lambda\mathbf{D}(\widehat{\tau})\hat{\rho} = \alpha_0 + \hat{\Theta}(\widehat{\tau})\mathbf{X}'(\widehat{\tau})U/n - \Delta(\widehat{\tau})/n^{1/2} + \hat{\Theta}(\widehat{\tau})(\mathbf{X}'(\widehat{\tau})\mathbf{X}(\tau_0) - \mathbf{X}'(\widehat{\tau})\mathbf{X}(\widehat{\tau}))\alpha_0/n.$$

We define our desparsified Lasso for threshold regression as

$$\hat{a}(\widehat{\tau}) \quad = \quad \hat{\alpha}(\hat{\tau}) + \hat{\Theta}(\hat{\tau})\lambda\mathbf{D}(\hat{\tau})\hat{\rho} = \hat{\alpha}(\hat{\tau}) + \hat{\Theta}(\hat{\tau})\mathbf{X}'(\tau)(Y - \mathbf{X}(\hat{\tau})\widehat{\alpha}(\hat{\tau}))/n. \tag{5.4}$$

In order 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 $H = \{j = 1, \cdots 2p \mid g_j \neq 0\}$ with cardinality $|H| = h < p$. H contains the indices of the coefficients involved. This implies $\|g\|_1 \leq \sqrt{h}$ by Cauchy–Schwarz inequality. In particular, $g = e_j$ is the case where we only consider a single coefficient, where e_j is the j-th $2p \times 1$ unit vector.

Considering

(5.5)
$$\begin{aligned}\sqrt{n}g'(\hat{a}(\hat{\tau}) - \alpha_0) &= g'\hat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)U//n^{1/2} \\ &+ g'(\hat{\Theta}(\hat{\tau})\mathbf{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)U)/n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\mathbf{X}'(\hat{\tau})\mathbf{X}(\tau_0) - \mathbf{X}'(\hat{\tau})\mathbf{X}(\hat{\tau}))\alpha_0/n^{1/2} - g'\Delta(\hat{\tau}),\end{aligned}$$

a central limit theorem for $g'\widehat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)U//n^{1/2}$ and a verification of asymptotic negligibility of $g'(\hat{\Theta}(\hat{\tau})\mathbf{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)U)//n^{1/2}$, $g'\Delta(\hat{\tau})$ and $g'\hat{\Theta}(\hat{\tau})(\mathbf{X}'(\hat{\tau})\mathbf{X}(\tau_0) - \mathbf{X}'(\hat{\tau})\mathbf{X}(\hat{\tau}))\alpha_0/n^{1/2}$ will achieve the desired convergence and yield asymptotically Gaussian inference.

5.2 Constructing the Approximate Inverse $\widehat{\Theta}(\tau)$

In this section, we formalize the approximate inverse $\widehat{\Theta}(\tau)$ used in our threshold model. The approach is as in van de Geer et al. (2014) but we must also verify that our required conditions are satisfied.

For the purpose discussed in the above (5.5), we need to find a well-behaved $\widehat{\Theta}(\tau)$ and investigate the asymptotic properties of $\widehat{\Theta}(\tau)$ uniformly on $\tau \in \widehat{\mathbb{T}}$. To do so we relate $\widehat{\Theta}(\tau)$ to $\Theta(\tau) := \Sigma(\tau)^{-1}$. Recall that $\Theta(\tau) := \Sigma(\tau)^{-1}$ defined in (3.4),

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

Define $\widehat{\mathbf{M}}(\tau) = \frac{1}{n}\sum_{i=1}^{n} X_i{}'X_i 1\{Q_i < \tau\}$ and $\widehat{\mathbf{N}}(\tau) = \frac{1}{n}\sum_{i=1}^{n} X_i{}'X_i 1\{Q_i \geq \tau\}$. We construct the approximate inverse $\widehat{\Theta}(\tau)$ because $2p > n$. To be precise, the threshold variable Q_i is used to split the sample into two groups. As long as either sample covariance matrix $\widehat{\mathbf{M}}(\tau)$ or $\widehat{\mathbf{N}}(\tau)$ is of reduced rank, we have to construct their respective approximate inverses.

Then we construct approximate inverse $\widehat{\mathbf{A}}(\tau)$ of $\widehat{\mathbf{M}}(\tau)$ and $\widehat{\mathbf{B}}(\tau)$ of $\widehat{\mathbf{N}}(\tau)$ and we relate $\widehat{\mathbf{A}}(\tau)$ to $\mathbf{A}(\tau) := \mathbf{M}(\tau)^{-1}$ and $\widehat{\mathbf{B}}(\tau)$ to $\mathbf{B}(\tau) := \mathbf{N}(\tau)^{-1}$.

Let $X^{(-j)}(\tau)$ denote all columns of $X(\tau)$ except for the j-th one and let $\tilde{X}^{(j)}(\tau)$ denote the $(n \times 1)$ vector such that $\tilde{X}_i^{(j)}(\tau) = X_i^{(j)} 1\{Q_i \geq \tau\}$. Then, $\tilde{X}^{(-j)}(\tau)$ denotes a $(n \times (p-1))$ matrix except for the j-th column of $\tilde{X}(\tau)$. Along Section 2.1 of Yuan (2010) we can rewrite the following regression models with covariates orthogonal in L_2 to the error terms for all $j = 1 \cdots p$,

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

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

The details of the covariance matrix's relation to the regression coefficients are given in Appendix B of Caner and Kock (2018). $\xi^{(j)}$ and $\tilde{\xi}^{(j)}$ are independent of Q_i, so they are not a function of τ.

Define $s_j(\tau) = \max\{|\{k = 1 \dots p : \mathbf{A}(\tau)_{j,k} \neq 0\}|, |\{k = 1 \dots p : \mathbf{B}(\tau)_{j,k} \neq 0\}|\}$ i.e. the greater cardinality of $\mathbf{A}(\tau)_j$ or $\mathbf{B}(\tau)_j$. And let $\bar{s} = sup_{\tau \in \widehat{\mathbb{T}}} \max_{j \in H} s_j(\tau)$.

We put forward the following assumptions:

Assumption 5. *(i) For the parameter space* $\max_{1 \leq j \leq p} \|\gamma_j\|_\infty \leq C$ *, for some constant* $C > 0$*;*
(ii)$E(\xi_i^{(j)}|X_i, Q_i) = 0$ *and* $E[(\xi_i^{(j)})^2]$ *is uniformly bounded over* $j = 1, \cdots p$*;* $E(\tilde{\xi}_i^{(j)}|X_i, Q_i) = 0$ *and* $E[(\tilde{\xi}_i^{(j)})^2]$ *is uniformly bounded over* $j = 1, \cdots p$*;*
(iii) $\frac{\sqrt{EM_{\xi X}^2}\sqrt{\log p}}{\sqrt{n}} < \infty$*, where* $M_{\xi X} = \max_{1 \leq i \leq n} \max_{1 \leq l \leq p} |\xi_i^{(j)} X_i^{(l)}|$.

Assumption 5 controls the tail distribution of $|\xi_i^{(j)} X_i^{(l)}|$ and $|\tilde{\xi}_i^{(j)} X_i^{(l)}|$, in order to apply the oracle inequality proved in previous work.

Given any $\tau \in \widehat{\mathbb{T}}$, the Lasso nodewise regression for $\widehat{\mathbf{A}}(\tau)$ is defined as follows:

$$\hat{\gamma}_j(\tau) = \text{argmin}_{\gamma \in \mathbb{R}^{p-1}} \|X^{(j)}(\tau) - X^{(-j)}(\tau)\gamma\|_n^2 + \lambda_{node}\|\widehat{\mathbf{\Gamma}}_{\mathbf{j}}(\tau)\gamma\|_1, \tag{5.6}$$

where

$$\widehat{\mathbf{\Gamma}}_{\mathbf{j}}(\tau) := \text{diag}\left\{\left\|\mathbf{X}^{(l)}(\tau)\right\|_n, l = 1, ..., p, l \neq j\right\},$$

with components of $\hat{\gamma}_j(\tau) = \{\widehat{\gamma_j}^{(k)}(\tau);\ k = 1, ..., p,\ k \neq j\}$. The $(2p \times 2p)$ diagonal weighting matrix is denoted as follows: It is noteworthy that we choose λ to be the same in all of the nodewise regressions. The nodewise Lasso runs p times as an intermediate step to construct $\widehat{\mathbf{A}}(\tau)$. Let

$$\widehat{C}(\tau) = \begin{pmatrix} 1 & -\widehat{\gamma_1}^{(2)}(\tau) & \cdots & -\widehat{\gamma_1}^{(p)}(\tau) \\ -\widehat{\gamma_2}^{(1)}(\tau) & 1 & \cdots & -\widehat{\gamma_2}^{(p)}(\tau) \\ \dots & \dots & \ddots & \dots \\ -\widehat{\gamma_p}^{(1)}(\tau) & -\widehat{\gamma_p}^{(2)}(\tau) & \cdots & 1 \end{pmatrix}. \tag{5.7}$$

Then, model (5.6) will be sparse with $\hat{\gamma}_j(\tau)$ possessing $s_j(\tau)$ non-zero entries. To define $\widehat{\mathbf{A}}(\tau)$ we introduce a $p \times p$ diagonal matrix $\hat{Z}(\tau)^2 = diag(\hat{z_1}(\tau)^2, \cdots, \hat{z_p}(\tau)^2)$, where

$$\hat{z_j}(\tau)^2 = \|X^{(j)}(\tau) - X^{(-j)}(\tau)\hat{\gamma}_j(\tau)\|_n^2 + \lambda_{node}\|\widehat{\mathbf{\Gamma}}_{\mathbf{j}}(\tau)\widehat{\gamma}_j(\tau)\|_1, \tag{5.8}$$

for all $j = 1, ..., p$. Hence, we may define

$$\widehat{\mathbf{A}}(\tau) = \hat{Z}(\tau)^{-2}\hat{C}(\tau). \tag{5.9}$$

It remains to be shown that this $\widehat{\mathbf{A}}(\tau)$ is close to the inverse of $\widehat{\mathbf{M}}(\tau)$. We define $\hat{A}_j(\tau)$ as the j'th row of $\widehat{\mathbf{A}}(\tau)$. Thus, $\hat{A}_j(\tau) = \hat{C}_j(\tau)/\hat{z}_j(\tau)^2$. Denoting by $\tilde{e}_j$ the j'th $p \times 1$ unit vector, the KKT conditions also imply that

$$\|\hat{A}_j(\tau)'\widehat{\mathbf{M}}(\tau) - \tilde{e}_j'\|_\infty \leq \frac{\lambda_{node}}{\hat{z}_j(\tau)^2}. \tag{5.10}$$

Parallel to construction of $\widehat{\mathbf{A}}(\tau)$ above, we define

$$\widehat{\mathbf{B}}(\tau) = \widehat{\hat{Z}}(\tau)^{-2}\widehat{\hat{C}}(\tau). \tag{5.11}$$

We define

$$\widehat{\hat{\gamma}}(\tau)_j = \text{argmin}_{\gamma \in \mathbb{R}^{p-1}} \|\tilde{X}^{(j)}(\tau) - \tilde{X}^{(-j)}(\tau)'\gamma\|_n^2 + \lambda_{node}\|\widehat{\tilde{\mathbf{\Gamma}}}_{\mathbf{j}}(\tau)\gamma\|_1 \tag{5.12}$$

where

$$\widehat{\tilde{\mathbf{\Gamma}}}_{\mathbf{j}}(\tau) := \text{diag}\left\{ \left\| \tilde{\mathbf{X}}^{(l)}(\tau) \right\|_n, l = 1, ..., p, l \neq j \right\},$$

with components of $\widehat{\hat{\gamma}}(\tau)_j = \{\widehat{\hat{\gamma}}_j^{(k)}(\tau);\ k = 1, ..., p,\ k \neq j\}$. Denote by $\widehat{\hat{Z}}(\tau)^2 = diag(\widehat{\hat{z}}_1(\tau)^2, \cdots, \widehat{\hat{z}}_p(\tau)^2)$ which is a $p \times p$ diagonal matrix with

$$\widehat{\hat{z}}_j(\tau)^2 = \|\tilde{X}^{(j)}(\tau) - \tilde{X}^{(-j)}(\tau)'\widehat{\hat{\gamma}}(\tau)_j\|_n^2 + \lambda_{node}\|\widehat{\tilde{\mathbf{\Gamma}}}_{\mathbf{j}}(\tau)\widehat{\hat{\gamma}}(\tau)_j\|_1, \tag{5.13}$$

for all $j = 1, ..., p$. We let

$$\widehat{\hat{C}}(\tau) = \begin{pmatrix} 1 & -\widehat{\hat{\gamma}}_1^{(2)}(\tau) & \cdots & -\widehat{\hat{\gamma}}_1^{(p)}(\tau) \\ -\widehat{\hat{\gamma}}_2^{(1)}(\tau) & 1 & \cdots & -\widehat{\hat{\gamma}}_2^{(p)}(\tau) \\ \cdots & \cdots & \ddots & \cdots \\ -\widehat{\hat{\gamma}}_p^{(1)}(\tau) & -\widehat{\hat{\gamma}}_p^{(2)}(\tau) & \cdots & 1 \end{pmatrix}. \tag{5.14}$$

We also get the following inequality:

$$\|\hat{B}(\tau)_j'\widehat{\mathbf{N}}(\tau) - \tilde{e}_j'\|_\infty \leq \frac{\lambda_{node}}{\widehat{\hat{z}}_j(\tau)^2}. \tag{5.15}$$

Thus

$$\widehat{\mathbf{\Theta}}(\tau) = \begin{bmatrix} \widehat{\mathbf{B}}(\tau) & -\widehat{\mathbf{B}}(\tau) \\ -\widehat{\mathbf{B}}(\tau) & \widehat{\mathbf{A}}(\tau) + \widehat{\mathbf{B}}(\tau) \end{bmatrix}. \tag{5.16}$$

Denoting by e_j the j'th $2p \times 1$ unit vector,

$$\max_{j \in H} \sup_{\tau \in \widehat{\mathbb{T}}} \|\widehat{\boldsymbol{\Theta}}(\tau)'_j \widehat{\boldsymbol{\Sigma}}(\tau) - e'_j\|_\infty \leq \max_{j \in H} \sup_{\tau \in \mathbb{T}} \frac{\lambda_{node}}{\widehat{z}_j(\tau)^2} + \max_{j \in H} \sup_{\tau \in \mathbb{T}} \frac{\lambda_{node}}{\widehat{\widetilde{z}}_j(\tau)^2}. \tag{5.17}$$

(Formal proof is given in the Appendix.)

Hence, we get the error term resulting from this approximation, i.e. the upper bound on the maximal absolute entry of the j'th row of $\widehat{\boldsymbol{\Theta}}(\tau)'\widehat{\boldsymbol{\Sigma}}(\tau) - \mathbf{I}_{2p}$. This provides the sufficient conditions to show that $g'\Delta(\tau)$ in (5.5) is asymptotically negligible. We then have the following result.

Lemma 2. *Let Assumptions 1-5 be satisfied and set* $\lambda_{node} = \frac{C}{\mu}\sqrt{\frac{\log p}{n}}$. *Suppose that* $\tilde{\delta}^{(j)}(\widehat{\tau}) \neq 0$,*estimated via* (4.2). *Then,*

$$\max_{j \in H} \sup_{\tau \in \widehat{\mathbb{T}}} \|\widehat{\boldsymbol{\Theta}}(\tau)_j - \boldsymbol{\Theta}(\tau)_j\|_1 = O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{5.18}$$

$$\max_{j \in H} \sup_{\tau \in \widehat{\mathbb{T}}} \|\widehat{\boldsymbol{\Theta}}(\tau)_j - \boldsymbol{\Theta}(\tau)_j\|_2 = O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{5.19}$$

$$\max_{j \in H} \sup_{\tau \in \widehat{\mathbb{T}}} \|\widehat{\boldsymbol{\Theta}}(\tau)_j\|_1 = O_p\left(\sqrt{\bar{s}}\right) \tag{5.20}$$

$$\max_{j \in H} \sup_{\tau \in \widehat{\mathbb{T}}} \|\widehat{\boldsymbol{\Theta}}(\tau)'_j \widehat{\boldsymbol{\Sigma}}(\tau) - e'_j\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right) \tag{5.21}$$

5.3 Inference

In this section, we derive asymptotic normality under high-level conditions which allows us to establish joint inference on a linear combination of the entries of the desparsified Lasso $\widehat{a}(\widehat{\tau})$.

To this end, we define

$$\widehat{\Sigma}_{xu}(\hat{\tau}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i(\hat{\tau})'\mathbf{X}_i(\hat{\tau})(\widehat{U}_i(\hat{\tau}))^2,$$

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

$$\tilde{\kappa}(s_0, c_0, \widehat{\mathbb{T}}, \Sigma) = \max_{\tau \in \widehat{\mathbb{T}}} \max_{J_0 \subset \{1,\ldots,2p\}, |J_0| \leq s_0} \max_{\gamma \neq 0, \|\gamma_{J_0^c}\|_1 \leq c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{(\gamma' E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\gamma)^{1/2}}{\|\gamma_{J_0}\|_2},$$

and

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

The following assumptions are imposed to establish a limiting distribution for an increasing number of coefficients.

Assumption 6. *(i)* $\max_{1 \leq j \leq p} E\left[(X_i^{(j)})^{12}\right]$ *and* $E\left[U_i^4\right]$ *are bounded away from infinity uniformly in* n.

$$\frac{\sqrt{EM_{X^6}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1),\ \frac{\sqrt{EM_{X^2U^2}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1) \text{ and } \frac{\sqrt{EM_{X^4U^2}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1), \text{ where}$$

$$M_{X^6} = \max_{1\leq i\leq n}\max_{1\leq k,l,j\leq p} |(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2 - E(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2|,$$

$$M_{X^2U^2} = \max_{1\leq i\leq n}\max_{1\leq j,l\leq p} |X_i^{(j)}X_i^{(l)}U_i^2 - E[X_i^{(j)}X_i^{(l)}U_i^2]|,$$

and

$$M_{X^4U^2} = \max_{1\leq i\leq n}\max_{1\leq j,l\leq p} |(X_i^{(j)}X_i^{(l)}U_i)^2 - E[X_i^{(j)}X_i^{(l)}U_i]^2|.$$

(ii)

$$\frac{\sqrt{h\bar{s}}\log p}{\sqrt{n}} = o_p(1);$$

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

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

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

(iii) $\frac{(h\bar{s})^2}{n^2}(h\bar{s}\wedge p) = o_p(1)$;
$\kappa(\bar{s}, c_0, \widehat{\mathbb{T}}, \Sigma_{xu})$ *and* $\kappa(\bar{s}, c_0, \widehat{\mathbb{T}}, \Sigma)$ *are bounded away from zero;*
$\tilde{\kappa}(\bar{s}, c_0, \widehat{\mathbb{T}}, \Sigma_{xu})$ *and* $\tilde{\kappa}(\bar{s}, c_0, \widehat{\mathbb{T}}, \Sigma)$ *are bounded from above.*

Assumption 6 gives sufficient conditions for a central limit theorem result. Assumption 6 (i) controls the tail behavior of the covariates and the error terms. By Assumption 6(i), $\max_{1\leq j,l\leq p} var(X_i^{(j)}X_i^{(l)}U_i^2)$, $\max_{1\leq k,l,j\leq p} var(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2$ and $\max_{1\leq j,l\leq p} var(X_i^{(j)}X_i^{(l)}U_i)^2$ are bounded away from infinity uniformly in n.

Assumption 6(ii) limits the dimension of the models, the dimension involved in conducting joint inference, the sparsity of the population covariance matrix, and the sparsity of the slope parameter vector. Combining all 4 equations, we obtain

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

Moreover, we can strengthen our assumption to

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

The first part of Assumption 6(iii) is designed to verify the Lyapunov condition. Then the other part restricts the eigenvalues of $\Sigma_{xu}(\tau)$ and $\Sigma(\tau)$.

Hence, we have the following result.

Theorem 4. *Let Assumptions 1, 2, 3, 4, 5 and 6 be satisfied and let g be $2p \times 1$ vector satisfying $\|g\|_2 = 1$. If $\tilde{\delta}^{(j)}(\hat{\tau}) \neq 0$ estimated via (4.2). Then,*

$$\frac{\sqrt{n} g'(\hat{a}(\hat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \xrightarrow{d} N(0,1). \tag{5.22}$$

Furthermore,

$$\sup_{\alpha_0 \in \mathcal{A}_{\ell_0}(s_0)} \sup_{|\tau - \tau_0| \leq C\frac{s_0 \log p}{n}} \|g'\hat{\Theta}(\tau)\hat{\Sigma}_{xu}(\tau)\hat{\Theta}'(\tau)g - g'\Theta(\tau_0)\Sigma_{xu}(\tau_0)\Theta'(\tau_0)g\|_\infty = o_p(1) \tag{5.23}$$

where

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

The first part of Theorem 4 implies convergence to the normal distribution of a sub-vector of $\widehat{a}(\widehat{\tau})$ of increasing dimension uniformly over $\mathcal{A}_{\ell_0}(s_0)$ and uniformly on a neighborhood of τ_0. The number of parameters involved in hypotheses is allowed to grow to infinity at a rate restricted by the above Assumption 6(ii).

The second part shows that we propose a consistent estimator of the covariance matrix uniformly over $\mathcal{A}_{\ell_0}(s_0)$. The uniformity of (5.23) will also be used in the proof of uniform convergence below.

In the case where H is a set of fixed cardinality h, (5.22) implies that

$$\|\left(g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g\right)^{-\frac{1}{2}} \sqrt{n} g'(\hat{a}(\hat{\tau}) - \alpha_0)\|_2^2 \xrightarrow{d} \chi^2(h), \tag{5.24}$$

correspondingly $\|g\|_2 = 1$ and $H = \{j = 1, \cdots 2p \mid g_j \neq 0\}$. Thus, χ^2 test can be conducted with a hypothesis on h parameters simultaneously.

We now establish confidence intervals for our parameters. We refer to the proof of Theorem 3 in Caner and Kock (2018) and its details therefore are omitted.

Let $\Phi(t)$ denote the cumulative distribution function (CDF) of the standard normal distribution and $z_{1-\frac{\varepsilon}{2}}$ is the $1 - \frac{\varepsilon}{2}$ percentile of the standard normal distribution. Denote by $\widehat{\sigma}(\widehat{\tau})_j = \sqrt{e_j'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'e_j}$ for all $j \in \{1, \cdots 2p\}$. Let $\mathrm{diam}([a,b])$ map the length of the interval$[a,b] \subset \mathbb{R}$.

Hence we have the following result.

Theorem 5. *Let Assumptions 1, 2, 3, 4, 5 and 6 be satisfied and let g be $2p \times 1$ vector satisfying $\|g\|_2 = 1$. If $\widehat{\delta}(\widehat{\tau}) \neq 0$ estimated via (2.4). Then,*

$$\sup_{t \in \mathbb{R}} \sup_{\alpha_0 \in \mathcal{A}_{\ell_0}(s_0)} \sup_{|\tau - \tau_0| \leq C\frac{s_0 \log p}{n}} \left| \mathbb{P}\left\{ \frac{\sqrt{n} g'(\hat{a}(\tau) - \alpha_0)}{\sqrt{g'\widehat{\Theta}(\tau)\hat{\Sigma}(\tau)_{xu}\widehat{\Theta}(\tau)'g}} \leq t \right\} - \Phi(t) \right| \to 0. \tag{5.25}$$

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

(5.26)

$$\lim_{n\to\infty} \inf_{\alpha_0 \in \mathcal{A}_{\ell_0}(s_0)} \inf_{|\tau-\tau_0| \le C\frac{s_0 \log p}{n}} \mathbb{P}\left\{ \alpha_0^{(j)} \in \left[\widehat{a}^{(j)}(\tau) - z_{1-\frac{\varepsilon}{2}} \frac{\widehat{\sigma}(\tau)_j}{\sqrt{n}}, \widehat{a}^{(j)}(\tau) + z_{1-\frac{\varepsilon}{2}} \frac{\widehat{\sigma}(\tau)_j}{\sqrt{n}} \right] \right\} = 1 - \varepsilon.$$

Finally,

$$\sup_{\alpha_0 \in \mathcal{A}_{\ell_0}(s_0)} \sup_{|\tau-\tau_0| \le C\frac{s_0 \log p}{n}} diam(\left[\widehat{a}^{(j)}(\tau) - z_{1-\frac{\varepsilon}{2}} \frac{\widehat{\sigma}(\tau)_j}{\sqrt{n}}, \widehat{a}^{(j)}(\tau) + z_{1-\frac{\varepsilon}{2}} \frac{\widehat{\sigma}(\tau)_j}{\sqrt{n}} \right]) = O_p(\frac{1}{\sqrt{n}}). \tag{5.27}$$

(5.25) implies that the convergence to the standard normal distribution is actually valid uniformly over the ℓ_0-ball of radius at most s_0 and uniformly on a neighborhood of τ_0.

6 Monte Carlo Simulation

In this section we investigate the finite sample properties of the desparsified Lasso for threshold regression and compare it to the one of the desparsified Lasso of van de Geer et al. (2014) (for a linear model). All designs are repeated 100 times. Before discussing the results, we explain how the data was generated and the performance measures used to compare the desparsified Lasso for the threshold model to the desparsified Lasso of van de Geer et al. (2014).

6.1 Implementation details

The implementation of the desparsified Lasso for linear model is inspired by the publicly available code at `https://web.stanford.edu/~montanar/sslasso/code.html`. We also modify the code of Callot et al. (2017) at `https://github.com/lcallot/ttlas` to our desparsified Lasso for threshold model. We use the Generalized Information Criterion of Fan and Tang (2013) (GIC) or ten-fold cross-validation to select the tuning parameters λ. To choose λ_{node}, we use GIC.

The cross-validation does not improve the quality of the results but the processing time is considerably longer.

We focus in turn on the following dimensions: the number of observations, the number of covariates, and the dependence between the threshold variable and the covariates. The covariates and the error terms are assumed to follow a t-distribution with 10 degrees of freedom. The covariance matrix of the covariates is chosen to have a Toeplitz structure with $\Sigma_{j,l} = \rho^{|j-l|}$ for $\rho = 0$. Specifically, each covariate is generated as $X_i^{(j)} \sim t(10)$ for each $j \in 1, \cdots p$, the threshold variable $Q_i \sim$ uniform$(0,1)$, and the error term $U_i \sim t(10)$, which is independent of the covariates. Neither the intercept nor the thresholded intercept is involved in the design to simplify the test. β_0 is $p \times 1$ with $s_0/2$ ones and $p - s_0/2$ zeros. Without loss of generality, we assume the sparsity of β_0 and δ_0 are identical. We choose the threshold parameter $\tau_0 = 0.5$. to avoid either of the split samples being unbalanced.

All tests are carried out at a 5% significance level and all confidence intervals are at the 95% level.

The χ^2-test involves the first non-zero parameter and first zero parameter in β_0 and δ_0. For measuring the size of the χ^2-test using our desparsified Lasso for threshold regression, we test the true hypothesis $H_0 : (\beta_0^{(1)}, \beta_0^{(s_0/2+1)}, \delta_0^{(1)}, \delta_0^{(s_0/2+1)}) = (1, 0, 1, 0)$. Correspondingly, the χ^2-test only involves $\beta_0^{(1)}$ and $\beta_0^{(s_0/2+1)}$ for measuring the size with the desparsified Lasso of van de Geer et al. (2014), as a result of incorrect model selection. In other words, $H_0 : (\beta_0^{(1)}, \beta_0^{(s_0/2+1)}) = (1, 0)$. For measuring the power of the χ^2-test, we test the false hypothesis $H_0 : (\beta_0^{(1)}, \beta_0^{(s_0/2+1)}, \delta_0^{(1)}, \delta_0^{(s_0/2+1)}) = (1, 0, 1, 1)$ in the procedure of our desparsified Lasso for threshold regression, while $H_0 : (\beta_0^{(1)}, \beta_0^{(s_0/2+1)}) = (1, 1)$ in the procedure of the desparsified Lasso of van de Geer et al. (2014). Thus, the hypothesis is only false on $\beta_0^{(s_0/2+1)}$. Similarly, we construct confidence intervals for $(\beta_0^{(1)}, \beta_0^{(s_0/2+1)}, \delta_0^{(1)}, \delta_0^{(s_0/2+1)})$ or $(\beta_0^{(1)}, \beta_0^{(s_0/2+1)})$ depending on which model we select.

6.2 Design 1

In this design, we investigate the effect of using a threshold variable that is part of the set of covariates ($Q \in X$), or that is correlated with the covariate. To be more precise, let $X^{(2)}$ denote the second column of X, and $\rho_{Q,X^{(2)}}$ be the correlation between Q and $X^{(2)}$. We consider the case where Q is independent of X, $Q = X^{(2)}$, as well as $\rho_{Q,X^{(2)}} = 0.9$.

6.3 Design 2

This design is to increase the number of observations or the number of variables to investigate the asymptotic properties of our procedure. We take $s_0 = 10$, which is a very sparse setting to satisfy assumptions on s_0 with relatively large n and p.

The following combinations of n and p are considered:

$$(n, p) \in \{(500, 100), (500, 250), (500, 400), (100, 250), (300, 250)\}$$

6.4 Performance measures

The performance of our desparsified Lasso for threshold regression and the desparsified Lasso of van de Geer et al. (2014) are measured along the following statistics, averaged across iterations.

1. Size: We evaluate the size of the χ^2-test in (5.24) for a hypothesis involving more than one parameter. The null hypothesis is always that the coefficients equal the true value.

2. Power: To measure the power of the test, we test whether $\delta_0^{(s_0/2+1)}$ equals its assigned value plus 1. The difference in alternatives is merely to obtain non-trivial power comparisons (i.e. to avoid either the power of all tests being (very close to) zero or (very close to) one).

3. Coverage rate: The confidence intervals are constructed as in Theorem 5. We calculate the coverage rate of the first nonzero entry and first zero entry of $(\beta_0', \delta_0')'$ $(\beta_0^{(1)}, \beta_0^{(s_0/2+1)}, \delta_0^{(1)}, \delta_0^{(s_0/2+1)}$) or β_0 $(\beta_0^{(1)}, \beta_0^{(s_0/2+1)}$) depending on which method we select.

6.5 Results of simulations

In this section, we report the results of a series of simulation experiments evaluating the finite sample properties of the desparsified Lasso for threshold regression.

Table 1 shows whether the threshold variable is included in the set of covariates or is correlated with one of the covariates, has almost no impact on the performances of our desparsified Lasso Estimator for the high dimensional threshold model.

Table 2 shows that the desparsified Lasso for high dimensional threshold model is always less-size distorted while having slightly more power as n is increased in high-dimensional setting. The size and power approach nominal levels as n is increased. Our desparsified Lasso procedure always has coverages that gradually improve as the sample size is increased. However, it has a tendency to overcover. We also note that our procedure performs better than the misspecified linear desparsified Lasso procedure.

Table 3 shows that our procedure is less size distorted while having much more power. Our procedure is slightly size distorted and over-covers as the number of variables is increased. However, our desparsified Lasso procedure has better performance than the linear desparsified Lasso procedure.

Generally, the desparsified Lasso for threshold regression performs much better for size, power, and coverage rate than the desparsified Lasso of van de Geer et al. (2014) when the threshold effects are present.

Table 1: Summary statistics for Design 1: the dependence between the threshold variable and the covariates

		χ^2		coverage rate			
		size	power	non-zero		zero	
				n th	th	n th	th
		n=500, p=250					
Q is independent of X	DLTH	0.03	0.93	0.99	0.98	1.00	1.00
	DL	0.43	0.57	0.27		1.00	
$\rho_{Q,X^{(2)}} = 0.9$	DLTH	0.05	0.86	0.95	0.98	1.00	1.00
	DL	0.49	0.50	0.23		1.00	
$Q = X^{(2)}$	DLTH	0.04	0.91	0.97	0.98	1.00	1.00
	DL	0.44	0.46	0.21		1.00	

Table 2: Summary statistics for Design 2: the number of observations

		χ^2		coverage rate			
		size	power	non-zero		zero	
				n th	th	n th	th
		p=250, $Q = X^{(2)}$					
n=100	DLTH	0.12	0.19	0.90	0.77	1.00	1.00
	DL	0.03	0.26	0.96		1.00	
n=300	DLTH	0.09	0.72	0.93	0.94	1.00	1.00
	DL	0.21	0.34	0.78		1.00	
n=500	DLTH	0.04	0.91	0.97	0.98	1.00	1.00
	DL	0.44	0.46	0.21		1.00	

Table 3: Summary statistics for Design 2: the number of variables

		χ^2		coverage rate			
		size	power	non-zero		zero	
				n th	th	n th	th
		n=500, $Q = X^{(2)}$					
p=100	DLTH	0.05	0.97	0.95	0.95	1.00	1.00
	DL	0.75	0.65	0.08		1.00	
p=250	DLTH	0.04	0.91	0.97	0.98	1.00	1.00
	DL	0.44	0.46	0.21		1.00	
p=400	DLTH	0.02	0.92	0.99	0.97	1.00	1.00
	DL	0.34	0.44	0.41		1.00	

7 Conclusion

In this paper, we propose a desparsified Lasso estimator for estimating high-dimensional threshold models. The inference based on the estimator is shown to be asymptotically uniformly valid. We do not impose the variables of the model to be sub-gaussian. Future work may include extending the desparsified Lasso to dynamic panels with threshold effects.

8 Appendix

8.1 Proofs for Section 3

In this section of the appendix, firstly we prove the oracle inequality of prediction error. Let the event

$$\mathbb{A}_1 := \left\{ \max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 \le C_2^2 + \mu\lambda \right\}$$

$$\mathbb{A}_2 := \left\{ \min_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 \ge C_3^2 - \mu\lambda \right\},$$

In particular

$$\{\mathbb{A}_2\} \subseteq \left\{ \min_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 \ge C_3^2 - \mu\lambda \right\}$$

The following lemma provides lower bounds on the probabilities of upper bounds of second moment of regressors. B.

Lemma 3 (Probability of $\mathbb{A}_1$ and $\mathbb{A}_2$)**.** *Let Assumption 1 be satisfied and set λ by (3.1). Then*

$$\mathbb{P}\{\mathbb{A}_1\} \ge 1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{\frac{EM_{X^2}^2 \log p}{n}}{(\log p)^2} \right)$$

$$\mathbb{P}\{\mathbb{A}_2\} \ge 1 - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{\frac{EM_{Xt_0}^2 \log p}{n}}{(\log p)^2} \right)$$

Proof of Lemma 3. Consider the term $\|X_i^{(j)}\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2$. Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma):

$$\begin{aligned} &\mathbb{P}\left\{ \max_{1\le j\le p} |\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \ge 2E\left[\max_{1\le j\le p} |\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \right] + \frac{t}{n} \right\} \\ &\le \exp\{-\frac{t^2}{3n\max_{1\le j\le p} Var[(X_i^{(j)})^2]}\} + \tilde{C}\frac{EM_{X^2}^2}{t^2}, \text{ set } t = (n\log p)^{1/2}, \text{ then we have} \\ &\le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{X^2}^2}{n\log p} \end{aligned} \tag{8.1}$$

for a positive constant, $\tilde{C} > 0$. With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017) provides, with $\tilde{C} > 0$ a positive constant,

$$E\left[\max_{1\le j\le p} |\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \right] \le \tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{X^2}^2}\log p}{n}] = O_p(\sqrt{\frac{\log p}{n}}) \tag{8.2}$$

Let $c = \arg\max_{1\le j\le p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2])$

$$
\begin{aligned}
\max_{1\le j\le p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - C_2^2) &\le \max_{1\le j\le p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(c)})^2]) \\
&= \max_{1\le j\le p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]) \\
&\le \max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]|
\end{aligned} \tag{8.3}
$$

Combine (8.1) with (8.2),

(8.4)
$$
\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \ge 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{X^2}^2}\log p}{n}] + \frac{\sqrt{\log p}}{\sqrt{n}}\right\} \le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{X^2}^2}{n\log p}
$$

To get the first part of the lemma, we combine the above display with Assumption 1 (ii), (3.1) and (8.3)

$$
\begin{aligned}
\mathbb{P}\{\mathbb{A}_1^c\} =&\mathbb{P}\left\{\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - C_2^2 > C\frac{\sqrt{\log p}}{\sqrt{n}}\right\} \\
\le&\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \ge C\frac{\sqrt{\log p}}{\sqrt{n}}\right\} \le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{X^2}^2}{n\log p} = o_p(1)
\end{aligned} \tag{8.5}
$$

Therefore we have proved the first part of the lemma.

Next, consider $\mathbb{A}_2$. $\|(X_i^{(j)}(t_0))^2\|_n^2 = \frac{1}{n}\sum_{i=1}^n((X_i^{(j)}(t_0))^2)^2$. Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma):

(8.6)
$$
\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]| \ge 2E\left[\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]|\right] + \frac{t}{n}\right\}
$$

$\le \exp\{-\frac{t^2}{3n\max_{1\le j\le p}Var[(X_i^{(j)}(t_0))^2]}\} + \tilde{C}\frac{EM_{Xt_0}^2}{t^2}$, Set $t = (n\log p)^{1/2}$,then we have

$\le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{Xt_0}^2}{n\log p}$

for a positive constant, $\tilde{C} > 0$. With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017)

provides, with $\tilde{C} > 0$ a positive constant,

$$(8.7)\quad E\left[\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]|\right] \le \tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{Xt_0}^2}\log p}{n}] = O_p(\sqrt{\frac{\log p}{n}})$$

Let $c = \arg\min_{1\le j\le p}(E[(X_i^{(j)}(t_0))^2] - \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2)$

$$(8.8)\quad \begin{aligned} C_3^2 - \min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 &\le \min_{1\le j\le p}(E[(X_i^{(c)}(t_0))^2] - \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2) \\ &= \min_{1\le j\le p}(E[(X_i^{(j)}(t_0))^2] - \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2) \\ &\le \max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]| \end{aligned}$$

Combine (8.6) with (8.7),

(8.9)

$$\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]| \ge 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{Xt_0}^2}\log p}{n}] + \frac{\sqrt{\log p}}{\sqrt{n}}\right\} \le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{Xt_0}^2}{n\log p}$$

To get the probability of event $\mathbb{A}_2$, we combine the above display with Assumption 1 (ii) , (3.1) and (8.8)

$$(8.10)\quad \begin{aligned} \mathbb{P}\{\mathbb{A}_2^c\} =&\mathbb{P}\left\{C_3^2 - \min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 > C\frac{\sqrt{\log p}}{\sqrt{n}}\right\} \\ \le&\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]| \ge C\frac{\sqrt{\log p}}{\sqrt{n}}\right\} \\ \le&\frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{Xt_0}^2}{n\log p} = o_p(1) \end{aligned}$$

Therefore we have proved the lemma. $\square$

Define the events

$$\begin{aligned} \mathbb{A}_3 &:= \left\{\max_{1\le j\le p}\frac{1}{\left\|X^{(j)}\right\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} \le \frac{\mu\lambda}{2}\right\}, \\ \mathbb{A}_4 &:= \left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\left\|X^{(j)}(\tau)\right\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}1\{Q_i<\tau\} \le \frac{\mu\lambda}{2}\right\}, \end{aligned}$$

Next we provide a lower bound on the probabilities of $\mathbb{A}_1 \cap \mathbb{A}_2$ with a suitable choice of λ.

Lemma 4 (Probability of $\mathbb{A}_3 \cap \mathbb{A}_4$)**.** *Conditional on the events*$\mathbb{A}_1 \cap \mathbb{A}_2$*, let Assumption 1 be satisfied and set* λ *by* (3.1)*. Then*

$$\mathbb{P}\{\mathbb{A}_3 \cap \mathbb{A}_4\} \geq 1 - \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM_{UX}^2}{(n\log p)}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM_{UX}^2}{(n \log pn)}\right).$$

Proof of Lemma 4. With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017) yields, with $\tilde{C} > 0$ a positive constant,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\leq j\leq p} |\frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)} - E[(U_i X_i^{(j)}]| \geq 2E\left[\max_{1\leq j\leq p} |\frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)} - E[(U_i X_i^{(j)}]|\right] + \frac{t}{n}\right\} \\
(8.11)\quad &\leq \exp\{-\frac{t^2}{3n\max_{1\leq j\leq p} Var[U_i X_i^{(j)}]}\} + \tilde{C}\frac{EM_{UX}^2}{t^2},\ \text{ set } t = (n\log p)^{1/2}\text{,then we have} \\
&\leq \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{n\log p} = o_p(1)
\end{aligned}$$

Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma) with $\tilde{C} > 0$ a positive constant,

$$(8.12)\qquad E\left[\max_{1\leq j\leq p} |\frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)} - E[(U_i X_i^{(j)}]|\right] \leq \tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{UX}^2}\log p}{n}] = O_p(\sqrt{\frac{\log p}{n}})$$

Using that we are on the set on $\mathbb{A}_1$,

(8.13)

$$\max_{1\leq j\leq p} \frac{1}{\left\|X^{(j)}\right\|_n} \frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)} \leq \frac{1}{\min_{1\leq j\leq p}\left\|X^{(j)}\right\|_n} \max_{1\leq j\leq p} \frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)} \leq \frac{1}{\sqrt{C_3^2 - \mu\lambda}} \max_{1\leq j\leq p} \frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)}$$

Combine (8.11) with (8.12),

$$\mathbb{P}\left\{\max_{1\leq j\leq p} |\frac{1}{n}\sum_{i=1}^{n} U_i X_i^{(j)}| \geq 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{(EM_{UX}^2)^{1/2}\log p}{n}] + \frac{\sqrt{\log p}}{\sqrt{n}}\right\} \leq \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{n\log p} = o_p(1),$$

Then,

$$
\begin{aligned}
\mathbb{P}\{\mathbb{A}_3^c\} =&\mathbb{P}\left\{\max_{1\le j\le p} \frac{1}{\left\|X^{(j)}\right\|_n}\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)} > \frac{\mu C}{2}\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\
\le&\mathbb{P}\left\{\max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)} \ge \frac{\mu C}{2}\sqrt{C_3^2-\mu\lambda}\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\
\le&\mathbb{P}\left\{\max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)} \ge \tilde{C}\frac{\sqrt{\log p}}{\sqrt{n}}\right\} + o_p(1)\\
\le&\frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{n\log p} + o_p(1) = o_p(1)
\end{aligned}
$$

This shows also that

$$
\max_{1\le j\le p}\left|\frac{1}{\left\|X^{(j)}\right\|_n}\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)}\right| = O_p(\frac{\sqrt{\log p}}{\sqrt{n}}) \tag{8.14}
$$

Next consider the event $\mathbb{A}_4$.To show the sup norm over τ, we adapt the proof of equitation (A.1) and (A.2) in Lemma A.1 of Callot et al. (2017) to our purpose. conditional on $\mathbb{A}_4$ and $(Q_1,\cdots Q_n)$, sorted (X_i,U_i,Q_i) $i=\{1\cdots n\}$ by $(Q_1,\cdots Q_n)$ in ascending order, then by the independence of (X_i,U_i) and Q_i, for $j=1,\cdots p$,

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\left\|X^{(j)}(\tau)\right\|_n}|\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)}(\tau)| \ge t|(Q_i,\cdots Q_n)\right\}\\
\le&\mathbb{P}\left\{\frac{1}{\min_{1\le j\le p}\left\|X^{(j)}(t_0)\right\|_n}\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)}(\tau)| \ge t|(Q_i,\cdots Q_n)\right\}\\
=&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^k U_i X_i^{(j)}| \ge t\min_{1\le j\le p}\left\|X^{(j)}(t_0)\right\|_n|(Q_i,\cdots Q_n)\right\}\\
\le&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^k U_i X_i^{(j)}| \ge t\sqrt{C_3^2-\mu\lambda}\right\}
\end{aligned}
$$

Denote a deterministic upper triangular matrix

$$
\Xi = \begin{pmatrix} 1 & 1 & \cdots & 1\\ 0 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}, \tag{8.15}
$$

Let $\xi_i^{(k)}$ is the i-th row, k-th column element of Ξ, then

$$\max_{1\leq j\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{k}U_iX_i^{(j)}| = \max_{1\leq j\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}|$$

$U_iX_i^{(j)}\xi_i^{(k)}$ is independent centered random variable(not identical),

$$\max_{1\leq j\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{k}U_iX_i^{(j)}| = \max_{1\leq j\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}|$$

$U_iX_i^{(j)}\xi_i^{(k)}$ is independent centered random variable(not identical), $\frac{\max_{1\leq j\leq p}\max_{1\leq k\leq n}\sum_{i=1}^{n}Var[U_iX_i^{(j)}\xi_i^{(k)}]}{n} = \frac{\max_{1\leq j\leq p}\max_{1\leq k\leq n}\sum_{i=1}^{k}Var[U_iX_i^{(j)}]}{n} \leq \frac{\max_{1\leq j\leq p}\sum_{i=1}^{n}Var[U_iX_i^{(j)}]}{n} = \max_{1\leq j\leq p}Var[U_iX_i^{(j)}] < \infty$ and $\max_{1\leq j\leq p}\max_{1\leq i\leq k\leq n}|U_iX_i^{(j)}\xi_i^{(k)}| < M_{UX}$. So under assumption 1, conditions for maximal inequalities are stratified automatically. Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma) with $\tilde{C} > 0$ a positive constant,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\leq j\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}| \geq 2E\left[\max_{1\leq j\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}|\right] + t\right\}\\
\leq&\exp\{-\frac{t^2}{3\max_{1\leq j\leq p}\max_{1\leq k\leq n}\sum_{i=1}^{n}Var[U_iX_i^{(j)}\xi_i^{(k)}]}\} + \tilde{C}\frac{EM_{UX}^2}{t^2},\ \text{set } t = (n\log pn)^{1/2}, \text{then we have}\\
\leq&\frac{1}{(pn)^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{(n\log pn)}
\end{aligned} \tag{8.16}$$

With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017) yields, with $\tilde{C} > 0$ a positive constant,

$$\begin{aligned}
&E\left[\max_{1\leq j\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}|\right]\\
\leq&\sqrt{\max_{1\leq j\leq p}\max_{1\leq k\leq n}\sum_{i=1}^{n}Var[U_iX_i^{(j)}\xi_i^{(k)}]}\sqrt{\log pn} + \sqrt{EM_{UX}^2}\log pn\\
\leq&\tilde{C}\sqrt{n\log(pn)} + \sqrt{EM_{UX}^2}\log pn = O_p(\sqrt{n\log(pn)})
\end{aligned} \tag{8.17}$$

Combine (8.16) with (8.17) and we consider $p > n$,

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}|\ge 2\tilde{C}[\frac{\sqrt{\log(pn)}}{\sqrt{n}}+\frac{\sqrt{EM_{UX}^2}\log(pn)}{n}]+\frac{\sqrt{\log(pn)}}{\sqrt{n}}\right\}\\
\le&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}|\ge 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}}+\frac{\sqrt{EM_{UX}^2}\log p}{n}]+\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\
\le&\frac{1}{(pn)^{\tilde{C}}}+\tilde{C}\frac{EM_{UX}^2}{(n\log(pn))},
\end{aligned}
$$

Taking expectations over $(Q_1,\cdots Q_n)$ and set λ by (3.1) yields,

$$
\begin{aligned}
\mathbb{P}\{\mathbb{A}_4^c\}&=\mathbb{P}\left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\left\|X^{(j)}(\tau)\right\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}1\{Q_i<\tau\}>\frac{\mu C}{2}\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\
&\le\mathbb{P}\left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}1\{Q_i<\tau\}\ge\frac{\mu C}{2}\sqrt{C_3^2-\mu\lambda}\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\
&\le\frac{1}{(pn)^{\tilde{C}}}+\tilde{C}\frac{EM_{UX}^2}{(n\log pn)}+o_p(1)=o_p(1)
\end{aligned}
\tag{8.18}
$$

This shows that

$$
\left|\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\left\|X^{(j)}(\tau)\right\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}1\{Q_i<\tau\}\right|=O_p(\frac{\sqrt{\log p}}{\sqrt{n}})
\tag{8.19}
$$

Since $\mathbb{P}\{\mathbb{A}_3\cap\mathbb{A}_4\}\ge 1-\mathbb{P}\{\mathbb{A}_3^c\}-\mathbb{P}\{\mathbb{A}_4^c\}$, we have proved the lemma. □

Define $J_0 := J(\alpha_0)$, $\widehat{\mathbf{D}}=\widehat{\mathbf{D}}(\hat{\tau})$, $\mathbf{D}=\mathbf{D}(\tau_0)$ 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 5. *Conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, for $0<\mu<1$ we have*

$$
\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right\|_1\le 2\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+\left\|f_{(\alpha_0,\widehat{\tau})}-f_0\right\|_n^2.
\tag{8.20}
$$

Proof of Lemma 5. We begin by noting that (2.4)

$$
\widehat{S}_n+\lambda\left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1\le S_n(\alpha,\tau)+\lambda\left\|\mathbf{D}(\tau)\alpha\right\|_1
\tag{8.21}
$$

$$
\widehat{S}_n-S_n(\alpha,\tau)\le\lambda\left\|\mathbf{D}(\tau)\alpha\right\|_1-\lambda\left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1
\tag{8.22}
$$

for all $(\alpha, \tau) \in \mathbb{R}^{2p} \times \mathbb{T}$. Inserting (2.3) to left side of (8.22)

$$\begin{aligned}
&\widehat{S}_n - S_n(\alpha,\tau) \\
&= n^{-1} \left\|\mathbf{y} - \mathbf{X}(\widehat{\tau})\widehat{\alpha}\right\|_2^2 - n^{-1} \left\|\mathbf{y} - \mathbf{X}(\tau)\alpha\right\|_2^2 \\
&= n^{-1} \sum_{i=1}^n \left[U_i - (\mathbf{X}_i(\widehat{\tau})'\widehat{\alpha} - \mathbf{X}_i(\tau_0)'\alpha_0)\right]^2 - n^{-1} \sum_{i=1}^n \left[U_i - (\mathbf{X}_i(\tau)'\alpha - \mathbf{X}_i(\tau_0)'\alpha_0)\right]^2 \\
&= n^{-1} \sum_{i=1}^n \left\{\mathbf{X}_i(\widehat{\tau})'\widehat{\alpha} - \mathbf{X}_i(\tau_0)'\alpha_0\right\}^2 - n^{-1} \sum_{i=1}^n \left\{\mathbf{X}_i(\tau)'\alpha - \mathbf{X}_i(\tau_0)'\alpha_0\right\}^2 \\
&\quad - 2n^{-1} \sum_{i=1}^n U_i \left\{\mathbf{X}_i(\widehat{\tau})'\widehat{\alpha} - \mathbf{X}_i(\tau)'\alpha\right\} \\
&= \left\|\widehat{f} - f_0\right\|_n^2 - \left\|f_{(\alpha,\tau)} - f_0\right\|_n^2 \\
&\quad - 2n^{-1} \sum_{i=1}^n U_i X_i'(\widehat{\beta} - \beta) - 2n^{-1} \sum_{i=1}^n U_i \left\{X_i'\widehat{\delta} 1(Q_i < \widehat{\tau}) - X_i'\delta 1(Q_i < \tau)\right\}.
\end{aligned}$$

Further, the last term on the right side of above can be written as

$$\begin{aligned}
&n^{-1} \sum_{i=1}^n U_i \left\{X_i'\widehat{\delta} 1(Q_i < \widehat{\tau}) - X_i'\delta 1(Q_i < \tau)\right\} \\
&\quad = n^{-1} \sum_{i=1}^n U_i X_i'(\widehat{\delta} - \delta) 1(Q_i < \widehat{\tau}) + n^{-1} \sum_{i=1}^n U_i X_i'\delta \left\{1(Q_i < \widehat{\tau}) - 1(Q_i < \tau)\right\}.
\end{aligned}$$

Then, (8.22) can be bounded as follows:

$$\begin{aligned}
\left\|\widehat{f} - f_0\right\|_n^2 \le{}& \left\|f_{(\alpha,\tau)} - f_0\right\|_n^2 + \lambda \left\|\mathbf{D}(\tau)\alpha\right\|_1 - \lambda \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1 \\
&+ 2n^{-1} \sum_{i=1}^n U_i X_i'(\widehat{\beta} - \beta) + 2n^{-1} \sum_{i=1}^n U_i X_i'(\widehat{\delta} - \delta) 1(Q_i < \widehat{\tau}) \\
&+ 2n^{-1} \sum_{i=1}^n U_i X_i'\delta \left\{1(Q_i < \widehat{\tau}) - 1(Q_i < \tau)\right\}.
\end{aligned}$$

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

$$\begin{aligned}
\left\|\widehat{f} - f_0\right\|_n^2 \le{}& \left\|f_{(\alpha,\tau)} - f_0\right\|_n^2 + \mu\lambda \sum_{j=1}^p \left\|X^{(j)}\right\|_n (\widehat{\beta}_j - \beta_j) + \mu\lambda \sum_{j=1}^p \left\|X^{(j)}(\widehat{\tau})\right\|_n (\widehat{\delta}_j - \delta_j) \\
&+ \lambda \left\|\mathbf{D}(\tau)\alpha\right\|_1 - \lambda \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1 + R_n(\alpha,\tau) \\
\le{}& \left\|f_{(\alpha,\tau)} - f_0\right\|_n^2 + \mu\lambda \left\|\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha)\right\|_1 + \lambda \left\|\mathbf{D}(\tau)\alpha\right\|_1 - \lambda \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1 + R_n(\alpha,\tau)
\end{aligned} \tag{8.23}$$

for all $(\alpha, \tau) \in \mathbb{R}^{2p} \times \mathbb{T}$. Note the fact that

$$\left|\widehat{\alpha}^{(j)} - \alpha_0^{(j)}\right| + \left|\alpha_0^{(j)}\right| - \left|\widehat{\alpha}^{(j)}\right| = 0 \text{ for } j \notin J_0 \tag{8.24}$$

$$\left\|\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right\|_1 = \left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0}\right\|_1 + \left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0^C}\right\|_1 = \left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0}\right\|_1 + \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1, \tag{8.25}$$

$$\begin{aligned}
\|\mathbf{D}\alpha_0\|_1 - \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1 &= \|[\mathbf{D}\alpha_0]_{J_0}\|_1 - \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0}\right\|_1 - \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1 \\
&= \|[\mathbf{D}\alpha_0]_{J_0}\|_1 - \left\|\left[\widehat{\mathbf{D}}\alpha_0\right]_{J_0}\right\|_1 + \left\|\left[\widehat{\mathbf{D}}\alpha_0\right]_{J_0}\right\|_1 - \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0}\right\|_1 - \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1 \\
\text{using triangle inequality } &\leq \left|\|[\mathbf{D}\alpha_0]_{J_0}\|_1 - \left\|\left[\widehat{\mathbf{D}}\alpha_0\right]_{J_0}\right\|_1\right| + \left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0}\right\|_1 - \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1 \\
&= \left|\|\mathbf{D}\alpha_0\|_1 - \left\|\widehat{\mathbf{D}}\alpha_0\right\|_1\right| + \left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0}\right\|_1 - \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1.
\end{aligned} \tag{8.26}$$

Consider (8.20), conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, add $(1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha)\right\|_1$ on both sides of (8.23)(evaluating at $(\alpha, \tau) = (\alpha_0, \widehat{\tau})$, $R_n(\alpha_0, \widehat{\tau}) = 0$),to get

$$\begin{aligned}
&\|\widehat{f} - f_0\|_n^2 + (1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha)\right\|_1 \\
\leq &\lambda\left(\left\|\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right\|_1 + \left\|\widehat{\mathbf{D}}\alpha_0\right\|_1 - \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1\right) + \left\|f_{(\alpha_0,\widehat{\tau})} - f_0\right\|_n^2 \\
&\text{using (8.25) and (8.26)} \\
\leq &\lambda\left(\left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0}\right\|_1 + \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1 + \left|\|\mathbf{D}\alpha_0\|_1 - \left\|\widehat{\mathbf{D}}\alpha_0\right\|_1\right| + \left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0}\right\|_1 - \left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1\right) + \left\|f_{(\alpha_0,\widehat{\tau})} - f_0\right\|_n^2 \\
\leq &2\lambda\left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right]_{J_0}\right\|_1 + \left\|f_{(\alpha_0,\widehat{\tau})} - f_0\right\|_n^2,
\end{aligned}$$

which proves (8.20). □

We are ready to establish the prediction consistency of the Lasso estimator.

Proof of Lemma 1. Conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, add $(1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha)\right\|_1$ on both sides of (8.23) (evaluating at $(\alpha,\tau)=(\alpha_0,\tau_0)$, $\widehat{f}(\alpha_0,\widehat{\tau})-f_0=0$),to get

(8.27)

$$\begin{aligned}
&\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right\|_1\\
\leq&\lambda\left(\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right\|_1+\left\|\widehat{\mathbf{D}}\alpha_0\right\|_1-\left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1\right)+R_n\\
&\text{using (8.25) and (8.26)}\\
\leq&\lambda\left(\left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1+\left|\left\|\mathbf{D}\alpha_0\right\|_1-\left\|\widehat{\mathbf{D}}\alpha_0\right\|_1\right|+\left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1-\left\|\left[\widehat{\mathbf{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1\right)+R_n\\
\leq&2\lambda\left\|\left[\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\lambda\left|\|\widehat{\mathbf{D}}\alpha_0\|_1-\|\mathbf{D}\alpha_0\|_1\right|+R_n,
\end{aligned}$$

The 3 terms on the right side can be bounded as follows using Hölder's inequality :

$$|R_n|\leq 2\mu\lambda\sum_{j=1}^{p}\left\|X^{(j)}\right\|_n|\delta_0^{(j)}|\leq 2\mu\left\|\delta_0\right\|_1\lambda\sqrt{C_2^2+\mu\lambda}., \tag{8.28}$$

$$\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\leq\left\|\widehat{\mathbf{D}}\right\|_\infty\|(\widehat{\alpha}-\alpha_0)_{J_0}\|_1\leq\|(\widehat{\alpha}-\alpha_0)_{J_0}\|_1\sqrt{C_2^2+\mu\lambda}, \tag{8.29}$$

$$\left|\|\widehat{\mathbf{D}}\alpha_0\|_1-\|\mathbf{D}\alpha_0\|_1\right|\leq\left\|(\widehat{\mathbf{D}}-\mathbf{D})\alpha_0\right\|_1\leq\left\|\widehat{\mathbf{D}}-\mathbf{D}\right\|_\infty\|\alpha_0\|_1\leq 2\left\|\alpha_0\right\|_1\sqrt{C_2^2+\mu\lambda} \tag{8.30}$$

Combine (8.28), (8.29) and (8.30) with (8.27) yields

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

which is (3.2). □

8.2 Proofs for Section 3.1

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

Lemma 6. *Let Assumptions 1 hold, for a universal constant $C>0$,*

$$\left\|\frac{1}{n}X_i'X_i-E\left[X_i'X_i\right]\right\|_\infty\leq C\frac{\sqrt{\log p}}{\sqrt{n}}$$

with probability at least $1-\left(\frac{1}{p^{2\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{n\log p^2}\right)$,

$$\sup_{\tau\in\mathbb{T}}\left\|\frac{1}{n}X_i(\tau)'X_i(\tau)-E\left[X_i(\tau)'X_i(\tau)\right]\right\|_\infty\leq C\frac{\sqrt{\log p}}{\sqrt{n}}$$

with probability at least $1-\left(\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log p^2n)}\right)$.

Proof. With Assumption 1, Lemma E.2(ii) of Chernozhukov et al. (2017) with $\tilde{C} > 0$ a positive constant provides (set $\eta = 1$ and $s = 2$ in their Lemma) :

(8.31)
$$
\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j,l\le p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)} - E[(X_i^{(j)}X_i^{(l)}]| \ge 2E\left[\max_{1\le j,l\le p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)} - E[(X_i^{(j)}X_i^{(l)}]|\right] + \frac{t}{n}\right\}\\
&\le \exp\{-\frac{t^2}{3n\max_{1\le j\le p}Var[X_i^{(l)}X_i^{(j)}]}\} + \tilde{C}\frac{EM_{XX}^2}{t^2},\ \text{set } t = (n\log p^2)^{1/2}, \text{then we have}\\
&\le \frac{1}{p^{2\tilde{C}}} + \tilde{C}\frac{EM_{XX}^2}{n\log p^2} = o_p(1)
\end{aligned}
$$

Lemma E.1 of Chernozhukov et al. (2017) with $\tilde{C} > 0$ a positive constant yields:

(8.32)
$$
E\left[\max_{1\le j,l\le p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(j)} - E[(X_i^{(j)}X_i^{(j)}]\right] \le \tilde{C}[\frac{\sqrt{\log p^2}}{\sqrt{n}} + \frac{\sqrt{EM_{XX}^2}\log p^2}{n}] = O_p(\sqrt{\frac{\log p}{n}}),
$$

with $\tilde{C} > 0$ a positive constant. Combine (8.31) with (8.32),

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le l\le p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)} - E[(X_i^{(j)}X_i^{(l)}]| \ge 2\tilde{C}[\frac{\sqrt{\log p^2}}{\sqrt{n}} + \frac{\sqrt{EM_{XX}^2}\log p^2}{n}] + \frac{\sqrt{\log p^2}}{\sqrt{n}}\right\}\\
=&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le l\le p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)} - E[(X_i^{(j)}X_i^{(l)}]| \ge \left[2\tilde{C}\left(1 + \frac{(EM_{XX}^2)^{1/2}\sqrt{\log p^2}}{\sqrt{n}}\right) + 1\right]\frac{\sqrt{\log p^2}}{\sqrt{n}}\right\}\\
\le&\frac{1}{p^{2\tilde{C}}} + \tilde{C}\frac{EM_{XX}^2}{n\log p^2} = o_p(1),
\end{aligned}
$$

This shows that

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

Next, to show the sup norm over τ, we adapt the proof of equitation (A.1) in Lemma A.1 of Callot et al. (2017) to our purpose.

Conditional on $(Q_1,\cdots Q_n)$, sorted (X_i, U_i, Q_i) $i = \{1\cdots n\}$ by $(Q_1,\cdots Q_n)$ in ascending order, then

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j,l\le p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}1(Q_i<\tau) - 1(Q_i<\tau)E[X_i^{(j)}X_i^{(l)}]\right)| \ge t|(Q_1,\cdots Q_n)\right\}\\
\le&\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{k}\left(X_i^{(j)}X_i^{(l)} - E[X_i^{(j)}X_i^{(l)}]\right))| \ge t\right\}
\end{aligned} \tag{8.33}
$$

Recall definition of matrix (8.15), Ξ, and $\xi_i^{(k)}$ which is the i-th row, k-th column element of Ξ, then

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

$\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}$ is independent centered random variable (not identical) across i,

$$\frac{\max_{1\leq j,l\leq p}\max_{1\leq k\leq n}\sum_{i=1}^{n} var[\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}]}{n}$$
$$\leq \frac{\max_{1\leq j,l\leq p}\sum_{i=1}^{n} var[\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)]}{n} < \infty,$$

$$\max_{1\leq j,l\leq p}\max_{1\leq i,k\leq n}|\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|$$
$$\leq \max_{1\leq j,l\leq p}\max_{1\leq i\leq n}|X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]| = M_{XX}.$$

So under assumption 1, conditions for maximal inequalities are stratified automatically. We can apply a 3-layer Maximal Inequalities over j, l, k. Lemma E.2(ii) of Chernozhukov et al. (2017) (set $\eta = 1$ and $s = 2$ in their Lemma) with $\tilde{C} > 0$ a positive constant yields:

(8.34)
$$\mathbb{P}\left\{\max_{1\leq j,l\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}| \geq 2E\left[\max_{1\leq j,l\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\right]+t\right\}$$
$$\leq \exp\{-\frac{t^2}{3\max_{1\leq j,l\leq p}\sum_{i=1}^{n} var[\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}]}\} + \tilde{C}\frac{EM_{XX}^2}{t^2},\ \text{set } t = [n\log(p^2n)]^{1/2}, \text{then we have}$$
$$\leq \frac{1}{(p^2n)^{\tilde{C}}} + \tilde{C}\frac{EM_{XX}^2}{(n\log(p^2n))}$$

Lemma E.1 of Chernozhukov et al. (2017) provides:

(8.35)
$$E\left[\max_{1\leq j,l\leq p}\max_{1\leq k\leq n}|\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\right] \leq \tilde{C}(\sqrt{n}\sqrt{\log(p^2n)}+\sqrt{EM_{XX}^2}\log(p^2n)) = O_p(\sqrt{n\log(p^2n)})$$

Combine (8.34) with (8.35),

$$\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\ge 2\tilde{C}[\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}+\frac{\sqrt{EM_{XX}^2}\log(p^2n)}{n}]+\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log(p^2n))},$$

Then plug-in $t=\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}$ in(8.33) and take expectations over $(Q_1,\cdots,Q_n)\in(0,1)$ yields,

(8.36)

$$\mathbb{P}\{\max_{1\le j,l\le p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)}1\left(Q_i<\tau\right)-E[1\left(Q_i<\tau\right)]E[X_i^{(j)}X_i^{(l)}]|$$
$$\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\}$$
$$=\mathbb{P}\{\max_{1\le j,l\le p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)}1\left(Q_i<\tau\right)-E[1\left(Q_i<\tau\right)]E[X_i^{(j)}X_i^{(l)}]|$$
$$\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\mid(Q_1,\cdots Q_n)\}$$
$$\le\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{k}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right))|\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log(p^2n))}=o_p(1).$$

By Assumption 1,

$$\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}|$$
$$=\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}=O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

since we consider $p>>n$.

This shows that

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

□

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

and recall Assumption 2 (3.3)

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

Define the event

$$\mathbb{A}_5:=\left\{\frac{\kappa(s_0,c_0,\mathbb{T},\hat{\Sigma})^2}{2}<\widehat{\kappa}(c_0,\mathbb{T},\Sigma)^2\right\}$$

The next lemma provides a lower bound on the probability of set $\mathbb{A}_5$.

Lemma 7. *Let Assumptions 1-2 be satisfied,*

$$\mathbb{P}\left\{\mathbb{A}_5\right\}\geq 1-(\frac{1}{(p^2)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log p^2)})-(\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log(p^2n))}).$$

Proof of Lemma 7. Start with

$$\left|\gamma'\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)\gamma\right|=\left|\gamma'\left(\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)-E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]+E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right)\gamma\right| \tag{8.37}$$

$$\geq\left|\gamma'E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\gamma\right|-\left|\gamma'\left(\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)-E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right)\gamma\right| \tag{8.38}$$

by Holders' inequality

$$\left|\gamma'\left(\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)-E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right)\gamma\right|\leq\|\gamma\|_1^2\left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)-E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \tag{8.39}$$

Note that we have the restriction set definition

$$\|\gamma\|_1\leq\|\gamma_{J_0}\|_1+\left\|\gamma_{J_0^c}\right\|_1\leq(1+c_0)\left\|\gamma_{J_0}\right\|_1\leq(1+c_0)\sqrt{s_0}\left\|\gamma_{J_0}\right\|_2 \tag{8.40}$$

So $\frac{\|\gamma\|_1}{\|\gamma_{J_0}\|_2} \leq (1+c_0)\sqrt{s_0}$. Then divide (8.39) by $\|\gamma_{J_0}\|_2^2$ we have

$$\left|\frac{\left|\gamma'\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)\gamma\right|}{\|\gamma_{J_0}\|_2^2} - \frac{\left|\gamma' E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\gamma\right|}{\|\gamma_{J_0}\|_2^2}\right| \leq \frac{\|\gamma\|_1^2}{\|\gamma_{J_0}\|_2^2}\left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \tag{8.41}$$

$$\leq (1+c_0)^2 s_0 \left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \tag{8.42}$$

Since

$$\frac{\left|\gamma' E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\gamma\right|}{\|\gamma_{J_0}\|_2^2} - \frac{\left|\gamma'\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)\gamma\right|}{\|\gamma_{J_0}\|_2^2} \leq \left|\frac{\left|\gamma'\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)\gamma\right|}{\|\gamma_{J_0}\|_2^2} - \frac{\left|\gamma' E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\gamma\right|}{\|\gamma_{J_0}\|_2^2}\right| \tag{8.43}$$

We obtain

$$\frac{\left|\gamma'\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)\gamma\right|}{\|\gamma_{J_0}\|_2^2} \geq \frac{\left|\gamma' E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\gamma\right|}{\|\gamma_{J_0}\|_2^2} - (1+c_0)^2 s_0 \left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \tag{8.44}$$

Minimize over $\tau \in \mathbb{T}$ on right side,

$$\frac{\left|\gamma'\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)\gamma\right|}{\|\gamma_{J_0}\|_2^2} \geq \min_{\tau\in\mathbb{T}} \frac{\left|\gamma' E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\gamma\right|}{\|\gamma_{J_0}\|_2^2} - (1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}} \left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \tag{8.45}$$

Minimize over$\left\{\gamma \in \mathbb{R}^{2p} \setminus 0\right\}$ on right side,

$$\frac{\left|\gamma'\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)\gamma\right|}{\|\gamma_{J_0}\|_2^2} \geq \kappa(c_0,\mathbb{T},\Sigma)^2 - (1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}} \left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \tag{8.46}$$

The above inequality is ture for all $\tau \in \mathbb{T}$ and $\left\{\gamma \in \mathbb{R}^{2p} \setminus 0\right\}$, so minimize over $\tau \in \mathbb{T}$ and $\left\{\gamma \in \mathbb{R}^{2p} \setminus 0\right\}$ on left side we obtain,

$$\hat{\kappa}(c_0,\mathbb{T},\widehat{\Sigma})^2 \geq \kappa(c_0,\mathbb{T},\Sigma)^2 - (1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}} \left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \tag{8.47}$$

So if we can prove that with probability approaching one, $(1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}} \left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \leq \frac{\kappa(c_0,\mathbb{T},\Sigma)^2}{2}$, that will imply of $\frac{\kappa(c_0,\mathbb{T},\Sigma)^2}{2} \leq \widehat{\kappa}(c_0,\mathbb{T},\widehat{\Sigma})^2$ with probability approaching one.

Next, by Lemma 6

$$\sup_{\tau\in\mathbb{T}} \left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau) - E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

$$\mathbb{P}\left\{(1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}}\left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)-E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty \geq (1+c_0)^2 s_0\tilde{C}\frac{\sqrt{\log p}}{\sqrt{n}}\right\}=o(1), \tag{8.48}$$

We get with probability approaching one, $(1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}}\left\|\frac{1}{n}\mathbf{X}(\tau)'\mathbf{X}(\tau)-E\left[\mathbf{X}_i(\tau)'\mathbf{X}_i(\tau)\right]\right\|_\infty <$ $(1+c_0)^2 s_0\tilde{C}\frac{\sqrt{\log p}}{\sqrt{n}} \leq \kappa(c_0,\mathbb{T},\Sigma)^2/2$, since left side of that inequality converges to zero in probability, and the right side is constant. Then by (8.48) and (8.47)

$$\mathbb{P}\left\{\mathbb{A}_5\right\} \geq 1-o(1). \tag{8.49}$$

□

Lemma 8. *Suppose that $\delta_0=0$. Let Assumption1 amd 2 hold with $\kappa=\kappa(\frac{1+\mu}{1-\mu},\mathbb{T},\Sigma)$ for $\mu\in(0,1)$. Let $(\widehat{\alpha},\widehat{\tau})$ be the Lasso estimator defined by (2.4) with 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*

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

Proof of Lemma 8. Note that $\delta_0=0$ implies $\left\|f_{(\alpha_0,\widehat{\tau})}-f_0\right\|^2=0$.Conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ with (8.20), we have

$$\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 \leq 2\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1, \tag{8.50}$$

which implies that

$$\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0^c}\right\|_1 \leq \frac{1+\mu}{1-\mu}\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1. \tag{8.51}$$

As in Lemma 7, conditional on event $\mathbb{A}_5$, apply Assumption 2, specifically UARE $\kappa=\kappa(\frac{1+\mu}{1-\mu},\mathbb{T},\Sigma)$, to yield

$$\begin{aligned}
\kappa^2\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2^2 &\leq 2\widehat{\kappa}(\frac{1+\mu}{1-\mu},\mathbb{T})^2\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2^2\\
&\leq \frac{2}{n}\left\|\mathbf{X}(\widehat{\tau})\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right\|_2^2\\
&= \frac{2}{n}(\widehat{\alpha}-\alpha_0)'\widehat{\mathbf{D}}\mathbf{X}(\widehat{\tau})'\mathbf{X}(\widehat{\tau})\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\\
&\leq \frac{2\max(\widehat{\mathbf{D}})^2}{n}(\widehat{\alpha}-\alpha_0)'\mathbf{X}(\widehat{\tau})'\mathbf{X}(\widehat{\tau})(\widehat{\alpha}-\alpha_0)\\
&= 2\max(\widehat{\mathbf{D}})^2\left\|\widehat{f}-f_0\right\|_n^2,
\end{aligned} \tag{8.52}$$

where the last equality is due to the assumption that $\delta_0 = 0$.

Combining (8.50) with (8.52) yields

$$\begin{aligned}\left\|\widehat{f}-f_0\right\|_n^2 &\le \left\|\widehat{f}-f_0\right\|_n^2 + (1-\mu)\,\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 \le 2\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\\ &\le 2\lambda\sqrt{s_0}\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2 \le \frac{2\sqrt{2}\lambda}{\kappa}\sqrt{s_0}\max(\widehat{\mathbf{D}})\left\|\widehat{f}-f_0\right\|_n.\end{aligned}$$

Cancel $\left\|\widehat{f}-f_0\right\|_n$ on the both sides of the inequality,

$$\left\|\widehat{f}-f_0\right\|_n \le \frac{2\sqrt{2}\lambda}{\kappa}\sqrt{s_0}\max(\widehat{\mathbf{D}})$$

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

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

Next, conditional on $\mathbb{A}_1$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, by (8.51)

$$\begin{aligned}\left\|\widehat{\mathbf{D}}\left(\widehat{\alpha}-\alpha_0\right)\right\|_1 &= \left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1 + \left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0^c}\right\|_1\\ &\le 2\left(1-\mu\right)^{-1}\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\\ &\le 2\left(1-\mu\right)^{-1}\sqrt{s_0}\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2\\ &\le \frac{2}{\kappa\left(1-\mu\right)}\sqrt{s_0}\max(\widehat{\mathbf{D}})\left\|\widehat{f}-f_0\right\|_n\\ &\le \frac{4\sqrt{2}\lambda}{\left(1-\mu\right)\kappa^2}s_0(\max(\widehat{\mathbf{D}})^2)\\ &\le \frac{4\sqrt{2}\lambda}{\left(1-\mu\right)\kappa^2}s_0(C_2^2+\mu\lambda),\end{aligned} \tag{8.53}$$

which proves the second conclusion of the lemma, since conditional on $\mathbb{A}_4$

$$\left\|\widehat{\mathbf{D}}\left(\widehat{\alpha}-\alpha_0\right)\right\|_1 \ge \min(\widehat{\mathbf{D}})\left\|\widehat{\alpha}-\alpha_0\right\|_1 \ge \sqrt{C_3^2-\mu\lambda}\left\|\widehat{\alpha}-\alpha_0\right\|_1. \tag{8.54}$$

$$\left\|\widehat{\alpha}-\alpha_0\right\|_1 \le \frac{4\sqrt{2}}{\left(1-\mu\right)\kappa^2}\frac{C_2^2+\mu\lambda}{\sqrt{C_3^2-\mu\lambda}}s_0\lambda. \tag{8.55}$$

Hence, the second conclusion of lemma follows give the lower bound on the probability of $\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 \cap \mathbb{A}_4 \cap \mathbb{A}_5$. ☐

Proof of Theorem 1. The proof follows immediately from combining Assumption 1 and 2 with Lemma

8. In particular,

$$
\begin{aligned}
\mathbb{P}\{\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 \cap \mathbb{A}_4 \cap \mathbb{A}_5\} \geq 1 &- \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{EM_{X^2}^2}{n \log p}\right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{EM_{Xt_0}^2}{n \log p}\right) \\
&- \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM_{UX}^2}{n \log p}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM_{UX}^2}{n \log(pn)}\right) \\
&- \left(\frac{1}{p^{2\tilde{C}_9}} + \tilde{C}_{10} \frac{EM_{XX}^2}{n \log p^2}\right) - \left(\frac{1}{(p^2 n)^{\tilde{C}_{11}}} + \tilde{C}_{12} \frac{EM_{XX}^2}{n \log(p^2 n)}\right).
\end{aligned}
$$

□

8.3 Proofs for Section 3.2

The following lemma gives an upper bound of $|\widehat{\tau} - \tau_0|$ using only Assumption 3, conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$.

Lemma 9. *Suppose that Assumption 3 holds. Let*

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

where C_4 is the constant defined in Assumption 3. Then conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$

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

Proof of Lemma 9. As in the proof of Lemma 5, we have, on the events $\mathbb{A}_1$,$\mathbb{A}_2$,$\mathbb{A}_3$ and $\mathbb{A}_4$

$$
\begin{aligned}
&\widehat{S}_n - S_n(\alpha_0, \tau_0) \\
&= \left\|\widehat{f} - f_0\right\|_n^2 - 2n^{-1} \sum_{i=1}^n U_i X_i'(\widehat{\beta} - \beta_0) - 2n^{-1} \sum_{i=1}^n U_i X_i'(\widehat{\delta} - \delta_0) 1(Q_i < \widehat{\tau}) - R_n \qquad (8.56)\\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - \mu\lambda \left\|\widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)\right\|_1 - R_n.
\end{aligned}
$$

Then $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$,

$$
\begin{aligned}
&\left[\widehat{S}_n + \lambda \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1\right] - \left[S_n(\alpha_0,\tau_0) + \lambda \left\|\mathbf{D}\alpha_0\right\|_1\right] \\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - \lambda \left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 - \lambda\left[\left\|\mathbf{D}\alpha_0\right\|_1 - \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1\right] - R_n \\
&\quad \text{using (8.25) and (8.26)} \\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - 2\lambda \left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1 - \lambda\left|\left\|\mathbf{D}\alpha_0\right\|_1 - \left\|\widehat{\mathbf{D}}\widehat{\alpha}\right\|_1\right| - R_n \\
&\quad \text{using (8.28), (8.29) and (8.30) to bound the last three terms,} \\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - \left(6\lambda\sqrt{C_2^2+\mu\lambda}C_1 s_0 + 2\mu\lambda\sqrt{C_2^2+\mu\lambda}C_1 s_0\right) \\
&\geq \left\|\widehat{f} - f_0\right\|_n^2 - \left(2C_1(3+\mu)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}} s_0\lambda\right) \geq 0 \\
&\quad \text{by Lemma 1.}
\end{aligned}
\tag{8.57}
$$

Suppose now that $|\hat{\tau}-\tau_0| > \eta^*$, then Assumption 3 and (8.57) together imply that

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

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

We now provide a lemma for bounding the prediction loss as well as the l_1 estimation loss for α_0.To do so, we define a constant G_2, and functions of $(\lambda, c_\alpha, c_\tau, \|\delta_0\|_1)$ G_1 and G_3 accordingly:

$$
\begin{aligned}
G_2 =& \frac{12\left(C_2^2+\mu\lambda\right)}{\kappa^2}, \\
G_1 =& \sqrt{c_\tau} + \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau, \\
G_3 =& \frac{2\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5 C_1}}{\kappa}\left(c_\alpha c_\tau\right)^{1/2}.
\end{aligned}
$$

Lemma 10. *Suppose that $|\hat{\tau}-\tau_0| \leq c_\tau$ and $\|\hat{\alpha}-\alpha_0\|_1 \leq c_\alpha$ for some (c_τ, c_α). Let Assumption 2 and 4 hold with $\mathbb{S} = \{|\tau-\tau_0| \leq c_\tau\}$, $\kappa = \kappa(\frac{2+\mu}{1-\mu}, \mathbb{S}, \Sigma)$ 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*

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

Proof of Lemma 10. Note that

$$|R_n| = \left| 2n^{-1} \sum_{i=1}^{n} U_i X_i' \delta_0 \left\{ 1(Q_i < \widehat{\tau}) - 1(Q_i < \tau_0) \right\} \right| \leq \lambda \sqrt{c_\tau}. \tag{8.58}$$

by Assumption 4 (3.7). Conditioning on $\mathbb{A}_4$, the triangular inequality implies that

(8.59)

$$\begin{aligned}
\left| \left\| \widehat{\mathbf{D}} \alpha_0 \right\|_1 - \|\mathbf{D}\alpha_0\|_1 \right| &\leq \left| \sum_{j=1}^{p} \left(\left\| X^{(j)}(\hat{\tau}) \right\|_n - \left\| X^{(j)}(\tau_0) \right\|_n \right) \left| \delta_0^{(j)} \right| \right| \\
&\quad \text{applying the mean value theorem to } \left\| X^{(j)}(\hat{\tau}) \right\|_n \\
&\leq \sum_{j=1}^{p} \left(2 \left\| X^{(j)}(t_0) \right\|_n \right)^{-1} \left| \left\| X^{(j)}(\hat{\tau}) \right\|_n^2 - \left\| X^{(j)}(\tau_0) \right\|_n^2 \right| \left| \delta_0^{(j)} \right| \\
&\leq \sum_{j=1}^{p} \left(2 \left\| X^{(j)}(t_0) \right\|_n \right)^{-1} \left| \delta_0^{(j)} \right| \frac{1}{n} \sum_{i=1}^{n} \left| X_i^{(j)} \right|^2 \left| 1\{Q_i < \hat{\tau}\} - 1\{Q_i < \tau_0\} \right| \\
&\leq \left(2\sqrt{C_3^2 - \mu\lambda} \right)^{-1} \|\delta_0\|_1 C_5 c_\tau.
\end{aligned}$$

where the last inequality is due to Assumption 4(3.5). We now consider two cases:
(i) $\left\| \widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)_{J_0} \right\|_1 > \sqrt{c_\tau} + \left(2\sqrt{C_3^2 - \mu\lambda} \right)^{-1} C_5 \|\delta_0\|_1 c_\tau$ and
(ii) $\left\| \widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)_{J_0} \right\|_1 \leq \sqrt{c_\tau} + \left(2\sqrt{C_3^2 - \mu\lambda} \right)^{-1} C_5 \|\delta_0\|_1 c_\tau$.

Case (i): Combine (8.58) and (8.59)

$$\lambda \left| \left\| \widehat{\mathbf{D}} \alpha_0 \right\|_1 - \|\mathbf{D}\alpha_0\|_1 \right| + R_n < \lambda \left(2\sqrt{C_3^2 - \mu\lambda} \right)^{-1} \|\delta_0\|_1 C_5 (c_\tau + \lambda\sqrt{c_\tau}) + \lambda\sqrt{c_\tau} < \lambda \left\| \widehat{\mathbf{D}} \left(\widehat{\alpha} - \alpha_0 \right)_{J_0} \right\|_1.$$

Combine the above results with (8.27), we have

$$\left\| \widehat{f} - f_0 \right\|_n^2 + (1-\mu)\lambda \left\| \widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0) \right\|_1 \leq 3\lambda \left\| \widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)_{J_0} \right\|_1, \tag{8.60}$$

which implies

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

Then subtract $(1-\mu) \left\| \widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)_{J_0} \right\|_1$ on both sides,

$$\left\| \widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)_{J_0^c} \right\|_1 \leq \frac{2+\mu}{1-\mu} \left\| \widehat{\mathbf{D}}(\widehat{\alpha} - \alpha_0)_{J_0} \right\|_1. \tag{8.61}$$

In this case, we are applying Assumption 2 with adaptive restricted eigenvalue condition $\kappa(s_0, \frac{2+\mu}{1-\mu}, \mathbb{S}, \Sigma)$.

Recall that $\mathbf{X}_i(\tau) = (X_i', X_i'1\{Q_i < \tau\})'$ and $\mathbf{X}(\tau) = (\mathbf{X}_1(\tau)', \cdots, \mathbf{X}_n(\tau)')'$.Note the fact that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{2\Big(\mathbf{X}_i(\hat{\tau})'\hat{\alpha} - \mathbf{X}_i(\hat{\tau})'\alpha_0\Big)\Big(X_i'\delta_0\big[1(Q_i < \tau_0) - 1(Q_i < \hat{\tau})\big]\Big)\right\} \\
=&2\Big(\hat{\alpha}'\mathbf{X}(\hat{\tau})' - \alpha_0'\mathbf{X}(\hat{\tau})'\Big)\Big(X'\delta_0\big[1(Q_i < \tau_0) - 1(Q_i < \hat{\tau})\big]\Big) \\
=&2\Big(\hat{\alpha}'\mathbf{X}_i(\hat{\tau})' - \alpha_0'\mathbf{X}_i(\hat{\tau})'\Big)\Big(\mathbf{X}_i(\tau_0)\alpha_0 - \mathbf{X}_i(\hat{\tau})\alpha_0\Big) \\
=&2\hat{\alpha}'\mathbf{X}(\hat{\tau})'\mathbf{X}(\tau_0)\alpha_0 - 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\tau_0)\alpha_0 - 2\hat{\alpha}'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\alpha_0 + 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\alpha_0 \\
=& -2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\tau_0)\alpha_0 + 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\alpha_0 + 2\alpha_0'\mathbf{X}(\tau_0)'\mathbf{X}(\hat{\tau})\hat{\alpha} - 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\hat{\alpha} \\
&\text{since } \alpha_0'\mathbf{X}(\tau_0)'\mathbf{X}(\tau_0)\alpha_0 + \alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\alpha_0 \ge 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\tau_0)\alpha_0 \\
\ge& -\alpha_0'\mathbf{X}(\tau_0)'\mathbf{X}(\tau_0)\alpha_0 - \alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\alpha_0 + 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\alpha_0 + 2\alpha_0'\mathbf{X}(\tau_0)'\mathbf{X}(\hat{\tau})\hat{\alpha} - 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\hat{\alpha} \\
=& -\alpha_0'\mathbf{X}(\tau_0)'\mathbf{X}(\tau_0)\alpha_0 + \alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\alpha_0 + 2\alpha_0'\mathbf{X}(\tau_0)'\mathbf{X}(\hat{\tau})\hat{\alpha} - 2\alpha_0'\mathbf{X}(\hat{\tau})'\mathbf{X}(\hat{\tau})\hat{\alpha}
\end{aligned}
$$

Since it is assumed that $|\hat{\tau} - \tau_0| \le c_\tau$, Assumption 2 only needs to hold with $\mathbb{S}$ in the c_τ neighborhood of τ_0. As $\delta_0 \neq 0$, (8.52) now has an extra term

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

where the last inequality is due to events $\mathbb{A}_1$ and Assumption 4(3.5). Combining this result with

(8.60), we have

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

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

$$\left\|\widehat{f}-f_0\right\|_n^2 \leq \frac{36\left(C_2^2+\mu\lambda\right)}{\kappa^2}\lambda^2 s_0 \vee \frac{6\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{\kappa}\lambda\sqrt{s_0\|\delta_0\|_1}\left(c_\alpha c_\tau\right)^{1/2}. \tag{8.62}$$

To derive the upper bound for $\|\widehat{\alpha}-\alpha_0\|_1$, use (8.61) ,

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

where the last inequality is due to conditional on $\mathbb{A}_3$. Then using the inequality that $a+b\leq 2a\vee 2b$ with (8.55) and (8.62) yields

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

Case (ii): In this case, (8.27) shows

$$\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0^C}\right\|_1 \leq 2\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1-(1-\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+\lambda\left|\left\|\widehat{\mathbf{D}}\alpha_0\right\|_1-\|\mathbf{D}\alpha_0\|_1\right|+R_n \tag{8.63}$$

which implies

$$\left\|\widehat{f}-f_0\right\|_n^2+\leq(1+\mu)\lambda\left\|\widehat{\mathbf{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+\lambda\left|\left\|\widehat{\mathbf{D}}\alpha_0\right\|_1-\|\mathbf{D}\alpha_0\|_1\right|+R_n. \tag{8.64}$$

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

which provides the result. □

The following lemma shows that the bound for $|\hat{\tau}-\tau_0|$ can be further tightened if we combine results obtained in Lemmas 9 and 10.

Lemma 11. *Suppose that $|\hat{\tau}-\tau_0| \leq c_\tau$ and $\|\widehat{\alpha}-\alpha_0\|_1 \leq c_\alpha$ for some (c_τ, c_α). Let Assumption 3 hold, then conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, we have,*

$$|\hat{\tau}-\tau_0| \leq \tilde{\eta}.$$

Proof of Lemma 11. Note that on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ $\mathbb{A}_4$ and Assumption 4(3.7),

$$\begin{aligned}
&\left|\frac{2}{n}\sum_{i=1}^{n}\left[U_iX_i'\left(\hat{\beta}-\beta_0\right)+U_iX_i'1\left(Q_i<\hat{\tau}\right)\left(\hat{\delta}-\delta_0\right)\right]\right| \\
\leq\ & \mu\lambda\left(\sqrt{C_2^2+\mu\lambda}\right)\|\hat{\alpha}-\alpha_0\|_1 \leq \mu\lambda\left(\sqrt{C_2^2+\mu\lambda}\right)c_\alpha
\end{aligned}$$

and

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

Suppose $\tilde{\eta} < |\hat{\tau}-\tau_0| \leq c_\tau$. As in (8.56),

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

Furthermore, we obtain

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

where the last inequality is due to Assumption 3, Hölder's inequality and (8.59).

Since $C_4\tilde{\eta}=\left((1+\mu)\sqrt{C_2^2+\mu\lambda}c_\alpha+G_1\right)\lambda$ by definitation, similarly as in the proof of Lemma 9, proof by contradiction yields the result. □

Lemma 10 provides us with three different bounds for $\|\alpha-\alpha_0\|_1$ and the two terms G_1 amd G_3 are functions of c_τ and c_α. If we can show that the bound for $|\hat{\tau}-\tau_0|$ and $|\hat{\alpha}-\alpha_0|$ in 10 and 11 are further tightened, it is useful to apply Lemmas 10 and 11 iteratively. to tighten up the bounds i

Lemma 9 results in that we can start the iteration with $c_\tau^{(0)}=\frac{2C_1(3+\mu)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}}{C_4}s_0\lambda$. (3.2) in Lemma 1 allow us to choose $c_\alpha^{(0)}=\frac{\left(2C_1(3+\mu)\right)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}}{(1-\mu)\left(C_3^2-\mu\lambda\right)^{\frac{1}{2}}}s_0$.

Lemma 12. *Suppose that Assumption 1 to 4 hold with $\mathbb{S}=\{|\tau-\tau_0|\le\eta^*\}$, $\kappa=\kappa(s_0,\frac{2+\mu}{1-\mu},\mathbb{S},\Sigma)$ for $0<\mu<1$. Let $(\widehat{\alpha},\widehat{\tau})$ be the Lasso estimator defined by (2.4) with λ given by (3.1). In addition, there exists a sequence of constants $\eta_1,...,\eta_{m^*}$ for some finite m^*. With probability at least $1-\left(\frac{1}{p^{\tilde{C}_1}}+\tilde{C}_2\frac{EM_{X^2}^2}{n\log p}\right)-\left(\frac{1}{p^{\tilde{C}_3}}+\tilde{C}_4\frac{EM_{Xt_0}^2}{n\log p}\right)-\left(\frac{1}{p^{\tilde{C}_5}}+\tilde{C}_6\frac{EM_{UX}^2}{n\log p}\right)-\left(\frac{1}{(pn)^{\tilde{C}_7}}+\tilde{C}_8\frac{EM_{UX}^2}{(n\log pn)}\right)-\left(\frac{1}{(p^2)^{\tilde{C}_9}}+\tilde{C}_{10}\frac{EM_{XX}^2}{(n\log p^2)}\right)-\left(\frac{1}{(p^2n)^{\tilde{C}_{11}}}+\tilde{C}_{12}\frac{EM_{XX}^2}{(n\log p^2n)}\right)$ we have*

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

Proof of Lemma 12. The iteration to implement is as follows:

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

Step 2:When $m\ge 1$,

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

Step 3: We stop the iteration if

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

doesn't change.

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

Since $G_1^{(m-1)}$ and $G_2\lambda s_0$ are positive, $\frac{G_1^{(m-1)}}{G_2\lambda s_0} > 0$. Note that $c_\alpha^{(m)} \geq \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} G_2\lambda s_0$, we have

$$
\begin{aligned}
c_\tau^{(m)} =& \frac{\lambda}{C_4}\left((1+\mu)\sqrt{C_2^2+\mu\lambda}c_\alpha^{(m)} + G_1^{(m-1)}\right) \\
\geq& \frac{\lambda}{C_4}\left(\frac{3(1+\mu)\sqrt{C_2^2+\mu\lambda}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} G_2\lambda s_0 + G_1^{(m-1)}\right) \\
\geq& \frac{1}{C_4}\left(\frac{3(1+\mu)\sqrt{C_2^2+\mu\lambda}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} + \frac{G_1^{(m-1)}}{G_2\lambda s_0}\right) G_2\lambda^2 s_0 \\
>& \frac{1}{C_4}\left(\frac{3(1+\mu)\sqrt{C_2^2+\mu\lambda}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}\right) G_2\lambda^2 s_0
\end{aligned}
\tag{8.65}
$$

Note that (8.65) shows that $c_\tau^{(m)} \geq C s_0 \frac{\log p}{2n}$ are valid for all each application of Lemma 9 to Lemma 11. Then $c_\alpha^{(m^*+1)}$ is the bound given in the statement of the lemma for $\|\hat{\alpha}-\alpha_0\|_1$. Next,

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

which is the bound given in the statement of the lemma for $|\hat{\tau}-\tau_0|$.

Next, we turn to Proof-of-Existence for 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 setting of $c_\tau^{(0)}$ and $c_\alpha^{(0)}$ in step 1. By step 2, as long as n,p, s_0 and $\|\delta_0\|_1$ are large enough, we obtain (in the following derivation, $\tilde{C}$ are different constant in each term,but all positive and finite)

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

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

follows from $\|\delta_0\|_1 s_0\lambda = o_p(1)$.

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

Thus we have

$$c_\alpha^{(0)} > c_\alpha^{(1)} \text{ and } c_\tau^{(0)} > c_\tau^{(1)}.$$

We assume

$$c_\alpha^{(m)} > c_\alpha^{(m+1)} \text{ and } c_\tau^{(m)} > c_\tau^{(m+1)},$$

it is easy to show

$$G_1^{(m)} > G_1^{(m+1)} \text{ and } G_3^{(m)} > G_3^{(m+1)}$$

then

$$c_\alpha^{(m+1)} > c_\alpha^{(m+2)} \text{ and } c_\tau^{(m+1)} > c_\tau^{(m+2)}.$$

This means that applying the iteration can tighten up the bounds.

We use proof by contradiction to be shown that there exist m^* such 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 $G_1^{(m-1)}$, $G_3^{(m-1)}$ are decreasing as m increases, and $\left\{G_1^{(m)} \vee G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}\right\}$ is bounded, there are two cases to consider:

Case (1):

$$G_1^{(m)} \leq G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}$$

for m sufficiently large. Let $G_3^{(m)}$ converge to $G_3^{(\infty)}$ and $G_3^{(\infty)} > G_2\lambda s_0$.

$$
\begin{aligned}
c_\alpha^{(\infty)} =& \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} G_3\sqrt{s_0\|\delta_0\|_1} =: H_1\sqrt{s_0\|\delta_0\|_1}\sqrt{c_\alpha^{(\infty)}}\sqrt{c_\tau^{(\infty)}}, \text{ where } H_1 \text{ is defined accordingly as}\\
H_1 =& \frac{6\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}\kappa}.\\
c_\alpha^\infty =& H_1^2 s_0\|\delta_0\|_1 c_\tau^\infty,\\
c_\tau^\infty =& C_4^{-1}\lambda\left((1+\mu)\sqrt{(C_2^2+\mu\lambda)}c_\alpha^\infty+\sqrt{c_\tau^\infty}+\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1}C_5\|\delta_0\|_1 c_\tau^\infty\right)\\
=& C_4^{-1}(1+\mu)\sqrt{(C_2^2+\mu\lambda)}\lambda c_\alpha^\infty + C_4^{-1}\lambda\sqrt{c_\tau^\infty}+C_4^{-1}\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1}C_5\|\delta_0\|_1\lambda c_\tau^\infty\\
=&: H_2\lambda c_\alpha^\infty + H_3\lambda\sqrt{c_\tau^\infty}+H_4\|\delta_0\|_1\lambda c_\tau^\infty,
\end{aligned}
$$

by defining

$$
\begin{aligned}
H_2 =&: C_4^{-1}(1+\mu)\sqrt{(C_2^2+\mu\lambda)},\\
H_3 =&: C_4^{-1},\\
H_4 =&: C_4^{-1}\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1}C_5.
\end{aligned}
$$

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

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

Then,

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

Obviously, the above 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):

$$
G_1^{(m)} > G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}
$$

for m sufficiently large. Let $G_1^{(m)}$ converge to $G_1^{(\infty)}$ and $G_1^{(\infty)} > G_2\lambda s_0$.

Thus, we have that

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

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

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

Then

$$
G_1^{(\infty)} = \left(1 + \left(2\sqrt{C_3^2-\mu\lambda} \right)^{-1} \lambda \|\delta_0\|_1 C_5 \right) \sqrt{c_{\tau}^{(\infty)}} + \left(2\sqrt{C_3^2-\mu\lambda} \right)^{-1} C_5 \|\delta_0\|_1 c_{\tau}^{(\infty)} = O_p(\lambda+\lambda^2),
$$

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

Finally, Lemma 10 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 combining Assumption 1 to 4 with Lemma 12. In particular,

$$
\begin{aligned}
\mathbb{P}\{\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 \cap \mathbb{A}_4 \cap \mathbb{A}_5\} \geq 1 &- \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{EM_{X^2}^2}{n\log p} \right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{EM_{Xt_0}^2}{n\log p} \right) \\
&- \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM_{UX}^2}{n\log p} \right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM_{UX}^2}{n\log(pn)} \right) \\
&- \left(\frac{1}{p^{2\tilde{C}_9}} + \tilde{C}_{10} \frac{EM_{XX}^2}{n\log p^2} \right) - \left(\frac{1}{(p^2n)^{\tilde{C}_{11}}} + \tilde{C}_{12} \frac{EM_{XX}^2}{n\log(p^2n)} \right).
\end{aligned}
$$

□

8.4 Proof of Asymptotic Properties of Nodewise Regression Estimator

The proof is similar to Lemma A.9 in the Appendix of Caner and Kock (2018). We adapt their proof to our purpose.

Define

$$\mathbb{A}_{node} = \{\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}} \|X^{(-j)}(\tau)'\xi^{(j)}/n\|_\infty \le \frac{\mu\lambda_{node}}{2}\},$$

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

$$\mathbb{B}_{node} = \{\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}} \|\tilde{X}^{(-j)}(\tau)'\xi^{(j)}/n\|_\infty \le \frac{\mu\lambda_{node}}{2}\},$$

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

The above four series of events are uniformly on $\tau \in \widehat{\mathbb{T}}$.

Lemma 13. *Let Assumptions 1-5 be satisfied and set $\lambda_{node} = \frac{C}{\mu}\sqrt{\frac{\log p}{n}}$. Suppose that $\widehat{\delta}(\widehat{\tau}) \neq 0$ estimated via (2.4). Then*

$$\mathbb{P}\left\{\mathbb{A}_{node} \cap_{j\in H} \mathbb{A}_{EV}^{(j)} \cap \mathbb{B}_{node} \cap_{j\in H} \mathbb{B}_{EV}^{(j)}\right\} \ge 1 - o_p(1).$$

Proof of Lemma 13. To prove probability of event $\mathbb{A}_{node}^C$, we adapt the the proof of Lemma 4 to our purpose,

$$\begin{aligned}\mathbb{P}\left\{\mathbb{A}_{node}^C\right\} &= \mathbb{P}\left\{\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}} \|X^{(-j)}(\tau)'\xi^{(j)}/n\|_\infty \le \frac{\mu\lambda_{node}}{2}\right\}\\ &=\mathbb{P}\left\{\max_{j\in H}\max_{1\le l\le p-1}\sup_{\tau\in\widehat{\mathbb{T}}} \frac{1}{n}\sum_{i=1}^n X_i^{(-j,l)}(\tau)\xi_i^{(j)} \le \frac{\mu\lambda_{node}}{2}\right\}\end{aligned}$$

Conditional on $(Q_1, \cdots Q_n)$, sorted (X_i, U_i, Q_i) $i = \{1\cdots n\}$ by $(Q_1, \cdots Q_n)$ in ascending order, then by the independence of (X_i, U_i) and Q_i,

$$\mathbb{P}\left\{\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\max_{1\le l\le p-1} \frac{1}{n}\sum_{i=1}^n X_i^{(-j,l)}(\tau)\xi_i^{(j)} \le \frac{\mu\lambda_{node}}{2} \mid (Q_1, \cdots Q_n)\right\} \tag{8.66}$$

$$=\mathbb{P}\left\{\max_{j\in H}\max_{1\le k\le n}\max_{1\le l\le p-1} \frac{1}{n}\sum_{i=1}^k X_i^{(-j,l)}\xi_i^{(j)} \le \frac{\mu\lambda_{node}}{2}\right\} \tag{8.67}$$

As there are 3 layers k, j, l across $\max_{1\le k\le n}\max_{j\in H}\max_{1\le l\le p-1}\sum_{i=1}^k X_i^{(-j,l)}\xi_i^{(j)}$, Combine (8.16) with (8.17) with setting $t = \sqrt{n\log((p-1)hn)}$,

$$\begin{aligned}
&\mathbb{P}\{\max_{j\in H}\max_{1\le k\le n}\max_{1\le l\le p-1}\frac{1}{n}\sum_{i=1}^{k}X_i^{(-j,l)}\xi_i^{(j)} \ge \\
&2\tilde{C}[\frac{\sqrt{n\log((p-1)hn)}}{n} + \frac{\sqrt{EM^2_{X\eta}}\log((p-1)hn)}{n}] + \frac{\sqrt{n\log((p-1)hn)}}{n}\} \\
\le&\frac{1}{[(p-1)hn]^{\tilde{C}}} + \tilde{C}\frac{EM^2_{X\eta}}{n\log((p-1)hn)} = o_p(1).
\end{aligned}$$

We see that

$$\begin{aligned}
&2\tilde{C}[\frac{\sqrt{n\log((p-1)hn)}}{n} + \frac{\sqrt{EM^2_{X\eta}}\log((p-1)hn)}{n}] + \frac{\sqrt{n\log((p-1)hn)}}{n} \\
\le&(2\tilde{C}+1)\frac{\sqrt{n\log p^3}}{n} + 2\tilde{C}\frac{\sqrt{EM^2_{X\eta}}\log p^3}{n} \\
\le&\sqrt{\frac{\log p}{n}}((2\tilde{C}+1)\sqrt{3} + 6\tilde{C}\sqrt{\frac{EM^2_{X\eta}\log p}{n}})
\end{aligned}$$

provided that we can find some constant $\tilde{C} > 0$.

Therefore if we choose $\frac{\mu\lambda_{node}}{2} = \sqrt{\frac{\log p}{n}}((2\tilde{C}+1)\sqrt{3} + 6\tilde{C}\sqrt{\frac{EM^2_{X\eta}\log p}{n}})$, the same rate as (3.1),

$$\mathbb{P}\left\{\mathbb{A}^C_{node}\right\} \le \frac{1}{[(p-1)hn]^{\tilde{C}}} + \tilde{C}\frac{EM^2_{X\eta}}{n\log((p-1)hn)} = o_p(1)$$

By the same arguments as above with a lower triangular matrix instead of (8.15), we have

$$\mathbb{P}\left\{\mathbb{B}^C_{node}\right\} \le \frac{1}{[(p-1)hn]^{\tilde{C}}} + \tilde{C}\frac{EM^2_{X\eta}}{n\log((p-1)hn)} = o_p(1)$$

Next, we bound the probability of event $(\cap_{j\in H}\mathbb{A}^{(j)}_{EV})^C$ and $(\cap_{j\in H}\mathbb{B}^{(j)}_{EV})^C$. Note the fact that for each $j \in H$

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

implies that

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

Thus,

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

By Assumption 4(3.5),

$$\sup_{\tau\in\widehat{\mathbb{T}}}\left\|\frac{1}{n}X(\tau)'X(\tau)-E\left[X_i(\tau)'X_i(\tau)\right]\right\|_\infty=O_p(s_0\frac{\log p}{n}).$$

Then by arguments exactly parallel to those in Lemma 7, we can show,

$$\mathbb{P}\left\{(\cap_{j\in H}\mathbb{A}_{EV}^{(j)})^C\right\}\leq o_p(1)$$

provided that $\kappa(s_j,c_0,\widehat{\mathbb{T}},M)>0$. Similarly, we can show

$$\mathbb{P}\left\{(\cap_{j\in H}\mathbb{B}_{EV}^{(j)})^C\right\}\leq o_p(1)$$

Therefore

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

□

Proof of Lemma 2. Given any $\tau\in\widehat{\mathbb{T}}\subset\mathbb{T}$ and each $j\in H$, (5.6) is a loss function for linear model, the pointwise oracle inequalities from Theorem 2.4 in van de Geer et al. (2014) for linear model have been proved.

As the uniform oracle inequalities only involve noise conditions $\mathbb{A}_{node}$ and $\mathbb{B}_{node}$, and adaptive restricted eigenvalue conditions $\cap_{j\in H}\mathbb{A}_{EV}^{(j)}$ and $\cap_{j\in H}\mathbb{B}_{EV}^{(j)}$. Therefore, by Lemma 13, we obtain the following results uniformly in $\widehat{\mathbb{T}}$ and H,

$$\sup_{\tau\in\widehat{\mathbb{T}}}\max_{j\in H}\|X^{(-j)}(\tau)'\gamma_j(\tau)-X^{(-j)}(\tau)'\widehat{\gamma}_j(\tau)\|_n\leq\frac{C}{\widehat{\kappa}(\bar{s},c_0,\widehat{\mathbb{T}},\Sigma)}\sqrt{\bar{s}}\lambda_{node}\tag{8.68}$$

$$\sup_{\tau\in\widehat{\mathbb{T}}}\max_{j\in H}\|\gamma_j(\tau)-\widehat{\gamma}_j(\tau)\|_1\leq\frac{C}{\widehat{\kappa}(\bar{s},c_0,\widehat{\mathbb{T}},\Sigma)^2}\bar{s}\lambda_{node}\tag{8.69}$$

with probability

$$\mathbb{P}\left\{\mathbb{A}_{node}\cap_{j\in H}\mathbb{A}_{EV}^{(j)}\cap\mathbb{B}_{node}\cap_{j\in H}\mathbb{B}_{EV}^{(j)}\right\}\geq 1-o(1)$$

Paralleling to the ones in Lemma A.9 in the Appendix of Caner and Kock (2018), the following inequalities can be established :

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\|\hat{A}_j(\tau)-A_j(\tau)\|_1=O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{8.70}$$

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\|\hat{A}_j(\tau)-A_j(\tau)\|_2=O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{8.71}$$

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\|\hat{A}_j(\tau)\|_1=O_p\left(\sqrt{\bar{s}}\right) \tag{8.72}$$

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\frac{1}{\hat{z}_j(\tau)^2}=O_p\left(1\right) \tag{8.73}$$

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\|\hat{B}_j(\tau)-B_j(\tau)\|_1=O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{8.74}$$

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\|\hat{B}_j(\tau)-B_j(\tau)\|_2=O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{8.75}$$

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\|\hat{B}_j(\tau)\|_1=O_p\left(\sqrt{\bar{s}}\right) \tag{8.76}$$

$$\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\frac{1}{\widehat{\hat{z}}_j(\tau)^2}=O_p\left(1\right) \tag{8.77}$$

We now turn to (3.4) and (5.16), for each $\tau\in\widehat{\mathbb{T}}$ and each $j\in H$ and $j\le p$

$$\|\hat{\Theta}(\tau)_j-\Theta(\tau)_j\|_1\le 2\|\hat{B}_j(\tau)-B_j(\tau)\|_1,$$

$$\|\hat{\Theta}(\tau)_j-\Theta(\tau)_j\|_2\le 2\|\hat{B}_j(\tau)-B_j(\tau)\|_2,$$

$$\|\hat{\Theta}(\tau)_j\|_1\le 2\|\hat{B}_j(\tau)\|_1;$$

for each $\tau\in\widehat{\mathbb{T}}$ and each $j\in H$ and $j>p$

$$\|\hat{\Theta}(\tau)_j-\Theta(\tau)_j\|_1\le\max\{2\|\hat{B}_j(\tau)-B_j(\tau)\|_1,\|\hat{B}_j(\tau)-B_j(\tau)\|_1+\|\hat{A}_j(\tau)-A_j(\tau)\|_1\},$$

$$\|\hat{\Theta}(\tau)_j-\Theta(\tau)_j\|_2\le\max\{2\|\hat{B}_j(\tau)-B_j(\tau)\|_2,\|\hat{B}_j(\tau)-B_j(\tau)\|_2+\|\hat{A}_j(\tau)-A_j(\tau)\|_2\},$$

$$\|\hat{\Theta}(\tau)_j\|_1\le\max\{2\|\hat{B}_j(\tau)\|_1,\|\hat{B}_j(\tau)\|_1+\|\hat{A}_j(\tau)\|_1\}.$$

Combine the two cases, we have proved the first 3 inequalities in Lemma 2.

We now consider $\max_{j\in H}\sup_{\tau\in\widehat{\mathbb{T}}}\|\hat{\Theta}(\tau)_j'\widehat{\mathbf{\Sigma}}(\tau)-e_j'\|_\infty$. For each $\tau\in\widehat{\mathbb{T}}$ and each $j\in H$ and $j\le p$

$$\|\hat{\Theta}(\tau)_j'\widehat{\mathbf{\Sigma}}(\tau)-e_j'\|_\infty=\|\begin{bmatrix}\hat{B}(\tau)_j & -\hat{B}(\tau)_j\end{bmatrix}\begin{bmatrix}\widehat{\mathbf{M}} & \widehat{\mathbf{M}}(\tau)\\ \widehat{\mathbf{M}}(\tau) & \widehat{\mathbf{M}}(\tau)\end{bmatrix}-e_j'\|_\infty$$

$$= \| \begin{bmatrix} \hat{B}(\tau)_j \widehat{\mathbf{N}}(\tau) & 0 \end{bmatrix} - e_j' \|_\infty \le \| \hat{B}(\tau)_j' \widehat{\mathbf{N}}(\tau) - \tilde{e}_j' \|_\infty \le \frac{\lambda_{node}}{\widehat{\tilde{z}}_j(\tau)^2}.$$

For each $\tau \in \widehat{\mathbb{T}}$ and each $j \in H$ and $j > p$

$$\| \hat{\Theta}(\tau)_j' \widehat{\mathbf{\Sigma}}(\tau) - e_j' \|_\infty = \| \begin{bmatrix} -\hat{B}(\tau)_j & \hat{B}(\tau)_j + \hat{A}(\tau)_j \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{M}} & \widehat{\mathbf{M}}(\tau) \\ \widehat{\mathbf{M}}(\tau) & \widehat{\mathbf{M}}(\tau) \end{bmatrix} - e_j' \|_\infty$$

$$= \| \begin{bmatrix} \hat{A}(\tau)_j \widehat{\mathbf{M}}(\tau) - \hat{B}(\tau)_j \widehat{\mathbf{N}}(\tau) & \hat{A}(\tau)_j \widehat{\mathbf{M}}(\tau) \end{bmatrix} - \begin{bmatrix} 0 & \tilde{e}_j' \end{bmatrix} \|_\infty$$

$$\le \max\{ \| \hat{A}(\tau)_j' \widehat{\mathbf{M}}(\tau) - \tilde{e}_j' \|_\infty + \| \hat{B}(\tau)_j' \widehat{\mathbf{N}}(\tau) - \tilde{e}_j' \|_\infty, \| \hat{A}(\tau)_j' \widehat{\mathbf{M}}(\tau) - \tilde{e}_j' \|_\infty \} \le \frac{\lambda_{node}}{\widehat{z}_j(\tau)^2} + \frac{\lambda_{node}}{\widehat{\tilde{z}}_j(\tau)^2}.$$

□

8.5 Proofs for Theorem 4 for Case II: fixed threshold

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

$$t = \frac{\sqrt{n} g'(\hat{a}(\hat{\tau}) - \alpha_0)}{\sqrt{g' \widehat{\Theta}(\hat{\tau}) \hat{\Sigma}(\hat{\tau})_{xu} \widehat{\Theta}(\hat{\tau})' g}} \tag{8.78}$$

is asymptotically standard normal. First, by rewriting (5.5),

$$t = t_1 + t_2,$$

where

$$t_1 = \frac{g' \widehat{\Theta}(\tau_0) \mathbf{X}'(\tau_0) U / n^{1/2}}{\sqrt{g' \widehat{\Theta}(\hat{\tau}) \hat{\Sigma}(\hat{\tau})_{xu} \widehat{\Theta}(\hat{\tau})' g}} \text{ and}$$

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

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

Lemma 14. *Let Assumptions 1, 2, 4, 5 and 6 be satisfied and let g be $2p \times 1$ vector satisfying $\|g\|_2 = 1$. If, furthermore, $\widehat{\delta}(\widehat{\tau}) \neq 0$ estimated via (2.4). Then,conditional on events $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$,*

$\mathbb{A}_4$, $\mathbb{A}_5$, *for any* $\widehat{\tau} \in \widehat{\mathbb{T}}$,

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

Proof of Lemma. As Q_i are continuously distributed and $E\left[|X_i^{(j)} X_i^{(l)}| |Q_i = \tau \right]$ is continuous and bounded in a neighborhood of τ_0, conditions for Lemma A.1 in Hansen (2000) hold. Then

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

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

Consider

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

Then, for any $\widehat{\tau} \in \widehat{\mathbb{T}}$, using Lemma 2

$$\begin{aligned}
&\|g' \left(\Theta(\widehat{\tau}) - \Theta(\tau_0) \right) \|_1 \\
=& \sum_{j \in H} \left(|g_j| \|\Theta_j(\widehat{\tau}) - \Theta_j(\tau_0)\|_1 \right) \\
\leq& \sum_{j \in H} |g_j| \sup_{\tau \in \widehat{\mathbb{T}}} \max_{j \in H} \|\Theta_j(\tau) - \Theta_j(\tau_0)\|_1 \\
\leq& \sqrt{h} \sup_{\tau \in \widehat{\mathbb{T}}} \max_{j \in H} \|\Theta_j(\tau)\|_1 \max_{j \in H} \|\Theta_j(\tau_0)\|_1 \| \left(\Sigma_j(\tau_0) - \Sigma_j(\tau) \right)' \|_\infty \\
=& O_p\left(\sqrt{h} \bar{s} s_0 \frac{\log p}{n} \right)
\end{aligned} \tag{8.79}$$

Finally,

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

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

Lemma 15. *Suppose that Assumption1 to 6 be satisfied, conditional on events $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, for any $\widehat{\tau}\in\widehat{\mathbb{T}}$,*

$$
\|g'\left(\hat{\Theta}(\hat{\tau})\mathbf{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)U\right)/n^{1/2}\|_\infty=O_p\left(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}\right)+O_p\left(\frac{\sqrt{h}\bar{s}s_0\log p}{\sqrt{n}}\right)
$$

Proof of Lemma . To prove this lemma,we need prove the followings

$$
\|g'\left(\hat{\Theta}(\hat{\tau})-\hat{\Theta}(\tau_0)\right)\mathbf{X}'(\tau_0)U\|_\infty/\sqrt{n}=o_p(1),
$$

$$
\|g'\left(\hat{\Theta}(\hat{\tau})\mathbf{X}'(\hat{\tau})-\hat{\Theta}(\hat{\tau})\mathbf{X}'(\tau_0)\right)U\|_\infty/\sqrt{n}=o_p(1).
$$

On the event $\mathbb{A}_1$,$\mathbb{A}_3$ and $\mathbb{A}_4$

$$
\|\frac{\mathbf{X}'(\tau_0)U}{\sqrt{n}}\|_\infty\quad\leq\quad\frac{1}{2}\sqrt{n}\mu\lambda\sqrt{C_2^2+\mu\lambda}
$$

Then, by Hölder's inequality and Lemma 14

$$
\begin{aligned}
&\|g'\left(\hat{\Theta}(\hat{\tau})\mathbf{X}'(\tau_0)-\hat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)\right)U\|_\infty/\sqrt{n}\\
\leq&\|g'\left(\hat{\Theta}(\hat{\tau})-\hat{\Theta}(\tau_0)\right)\|_1\|\frac{\mathbf{X}'(\tau_0)U}{\sqrt{n}}\|_\infty\\
\leq&O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)O_p(\sqrt{\log p})\\
=&O_p\left(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}\right),
\end{aligned}
\tag{8.80}
$$

Considering $\|\frac{\mathbf{X}'(\hat{\tau})U}{\sqrt{n}}-\frac{\mathbf{X}'(\tau_0)U}{\sqrt{n}}\|_\infty$, by Assumption 4 (3.6)

$$
\|\frac{\mathbf{X}'(\hat{\tau})U}{n}-\frac{\mathbf{X}'(\tau_0)U}{n}\|_\infty=O_p(\frac{\sqrt{s_0}\log p}{n}).
$$

$$
\begin{aligned}
&\|g'\hat{\Theta}(\hat{\tau})\left(\mathbf{X}'(\hat{\tau})-\mathbf{X}'(\tau_0)\right)U\|_\infty/\sqrt{n}\\
\leq&\sqrt{n}\|\frac{\mathbf{X}'(\hat{\tau})U}{n}-\frac{\mathbf{X}'(\tau_0)U}{n}\|_\infty\|g'\hat{\Theta}(\hat{\tau})\|_1\\
\leq&\sqrt{n}\|\frac{\mathbf{X}'(\hat{\tau})U}{n}-\frac{\mathbf{X}'(\tau_0)U}{n}\|_\infty\|g\|_1\max_{j\in H}\|\widehat{\Theta}_j(\hat{\tau})\|_1\\
\leq&\sqrt{n}O_p\left(\sqrt{h\bar{s}}\right)O_p\left(\frac{\sqrt{s_0}\log p}{n}\right)=O_p\left(\frac{\sqrt{h\bar{s}s_0}\log p}{\sqrt{n}}\right)
\end{aligned}
\tag{8.81}
$$

Hence, combine (8.80) and (8.81)

$$
\|g'\left(\hat{\Theta}(\hat{\tau})\mathbf{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)U\right)/n^{1/2}\|_\infty=O_p\left(\sqrt{h\bar{s}}\frac{\log p}{\sqrt{n}}\right)+O_p\left(\frac{\sqrt{h\bar{s}s_0}\log p}{\sqrt{n}}\right)
$$

□

Lemma 16. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$ *, then for any* $\widehat{\tau}\in\widehat{\mathbb{T}}$,

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

Proof of Lemma 16. There are only two cases for $\mathbf{X}'(\hat{\tau})\mathbf{X}(\tau_0)$:
$\mathbf{X}'(\hat{\tau})\mathbf{X}(\tau_0)=\mathbf{X}'(\tau_0)\mathbf{X}(\tau_0)$ or $\mathbf{X}'(\hat{\tau})\mathbf{X}(\tau_0)=\mathbf{X}'(\hat{\tau})\mathbf{X}(\hat{\tau})$, then

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

Recall

$$
|\hat{\tau}-\tau_0|\leq C\frac{\log p}{n}s_0
$$

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

where the last equality follows from Assumption 4(3.5).

We know that $\sup_{\tau\in\mathbb{T}}\max_{j\in H}\|\widehat{\Theta}_j(\tau)\|_1 = O_p(\sqrt{\bar{s}})$ by Lemma 2 and $\|\delta_0\|_1 \leq s_0C_1$ by Assumption 1 (vii)

$$
\begin{aligned}
&\|g'\hat{\Theta}(\hat{\tau})(\mathbf{X}'(\hat{\tau})\mathbf{X}(\tau_0)-\mathbf{X}'(\hat{\tau})\mathbf{X}(\hat{\tau}))\alpha_0/n^{1/2}\|_\infty \\
\leq&\sqrt{n}\max_{j\in H}\|\widehat{\Theta}_j(\hat{\tau})\|_1\sum_{j\in H}|g_j|\|\widehat{\mathbf{M}}(\tau_0)-\widehat{\mathbf{M}}(\hat{\tau})\|_\infty\|\delta_0\|_1 \\
\leq&\sqrt{n}O_p(\sqrt{h\bar{s}})O_p\left(\frac{s_0\log p}{n}\right)s_0C_1 \\
=&O_p\left(\frac{s_0^2\sqrt{h\bar{s}}\log p}{\sqrt{n}}\right)
\end{aligned}
$$

□

Lemma 17. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$ *,*$\mathbb{A}_2$*,* $\mathbb{A}_3$*,* $\mathbb{A}_4$*,* $\mathbb{A}_5$*, then for any* $\widehat{\tau}\in\widehat{\mathbb{T}}$*,* $g'\Delta(\hat{\tau}) = O_p(\frac{s_0\sqrt{h}\log p}{\sqrt{n}})$

Proof. Recall that $\Delta(\tau) = \sqrt{n}(\hat{\Theta}(\tau)\hat{\Sigma}(\tau)-I_{2p})(\hat{\alpha}(\tau)-\alpha_0)$
Thus by holder's inequality, Lemma 2 and Theorem 2,

$$
\begin{aligned}
g'\Delta(\hat{\tau}) &\leq \max_{j\in H}|\Delta_j(\hat{\tau})|\sum_{j\in H}|g_j| \\
&= \max_{j\in H}|\left(\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau})-\tilde{e}_j'\right)\sqrt{n}(\hat{\alpha}(\hat{\tau})-\alpha_0)|\sum_{j\in H}|g_j| \\
&\leq \max_{j\in H}|\left(\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau})-\tilde{e}_j'\right)\sqrt{n}(\hat{\alpha}(\hat{\tau})-\alpha_0)|\sum_{j\in H}|g_j| \\
&\leq \max_{1\leq j\leq 2p}\|\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau})-\tilde{e}_j'\|_\infty\sqrt{n}\|\hat{\alpha}(\hat{\tau})-\alpha_0\|_1\sum_{j\in H}|g_j| \\
&\leq C\left(\frac{\lambda_{node}}{\hat{z_1}^2(\hat{\tau})_j}+\frac{\lambda_{node}}{\hat{z_2}^2(\hat{\tau})_j}\right)\cdot\sqrt{n}\cdot\lambda s_0\sqrt{h} \\
&= O_p(\frac{s_0\sqrt{h}\log p}{\sqrt{n}})
\end{aligned}
$$

$\square$

Lemma 18. *Suppose that Assumption1 to 6 be satisfied, then*

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

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

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

Proof. Apply Lemma E.1 and E.2 of Chernozhukov et al. (2017) under Assumption 6 (ii) and (v), by arguments exactly parallel to those in proof of Lemma 6, and its proof therefore omitted. $\square$

Lemma 19. *Suppose that Assumption1 to 6 be satisfied, conditional on events $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, then for any $\widehat{\tau}\in\widehat{\mathbb{T}}$,*

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

Proof of Lemma .

$$\begin{aligned}
\text{Recall } \Sigma(\hat{\tau})_{xu} =& E\left[\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2\right] = E\left[\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\right]E\left[U_i^2\right],\\
\hat{U}_i(\hat{\tau}) =& Y_i - \mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}) = u_i + \mathbf{X}_i'(\tau_0)\alpha_0 - \mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}),\\
\hat{\Sigma}(\hat{\tau})_{xu} =& \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\hat{U}(\hat{\tau})_i^2,\\
\text{and set } \tilde{\Sigma}(\hat{\tau})_{xu} =& \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2
\end{aligned}$$

We first show that

$$\begin{aligned}
&\|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty\\
=\;& \|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g + g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\\
&+g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty\\
\leq\;& \|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g\|_\infty + \|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty\\
&+\|g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty
\end{aligned}$$

To prove this lemma,we need prove the followings

$$\|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g\|_\infty = o_p(1)$$

$$\|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g\|_\infty = o_p(1)$$

$$\|g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty = o_p(1)$$

Step 1.

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

Before we expand and simplify equations, we note that

$$\alpha_0'\mathbf{X}_i(\tau_0)\mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}) = \hat{\alpha}'(\hat{\tau})\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\tau_0)\alpha_0$$

$$
\begin{aligned}
&\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu} \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\hat{U}_i^2(\hat{\tau}) - \mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})(U_i + \mathbf{X}_i'(\tau_0)\alpha_0 - \mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}))^2 - \mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2 \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\alpha_0'\mathbf{X}_i(\tau_0)\mathbf{X}_i'(\tau_0)\alpha_0\right) \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\hat{\alpha}'(\hat{\tau})\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right) \\
-&\frac{2}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\alpha_0'\mathbf{X}_i(\tau_0)\mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right) \\
+&\frac{2}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\alpha_0'\mathbf{X}_i(\tau_0)U_i\right) \\
-&\frac{2}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})'\mathbf{X}_i(\hat{\tau})U_i\right) \\
-&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\alpha_0\mathbf{X}_i'(\tau_0)\left(\mathbf{X}_i'(\tau_0)\alpha_0 - \mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)\right) \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\mathbf{X}_i'(\hat{\tau})\left(\mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \mathbf{X}_i'(\tau_0)\alpha_0\right)\right) \\
+&\frac{2}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i\left(\alpha_0'\mathbf{X}_i(\tau_0) - \hat{\alpha}(\hat{\tau})'\mathbf{X}_i(\hat{\tau})\right)\right)
\end{aligned}
$$

Recall Lemma 18,

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

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

Applying Theorem 2,

$$
\|\hat{\alpha}(\hat{\tau})\|_1 \le \|\alpha_0\|_1 + O_p\left(s_0\sqrt{\frac{\log p}{n}}\right)
$$

$$\|\mathbf{X}(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \mathbf{X}(\tau_0)\alpha_0\|_n = O_p\left(\sqrt{s_0}\sqrt{\frac{\log p}{n}}\right)$$

$$\|\alpha_0\|_1 = O_p(s_0)$$

By Cauchy-Schwarz inequality and holder's inequality

$$\begin{aligned}
& \max_{1\le k,l\le 2p} |\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\mathbf{X}_i^{(k)}(\hat{\tau})\mathbf{X}_i^{(l)}(\hat{\tau})\alpha_0'\mathbf{X}_i(\tau_0)\left(\mathbf{X}_i'(\tau_0)\alpha_0 - \mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)\right)| \\
\le & \sqrt{\max_{1\le k,l\le 2p}\max_{1\le i\le n}\tfrac{1}{n}\textstyle\sum_{i=1}^n(\mathbf{X}_i^{(k)}(\hat{\tau})\mathbf{X}_i^{(l)}(\hat{\tau}))^2\left(\alpha_0\mathbf{X}_i'(\tau_0)\right)^2}\|\mathbf{X}(\tau_0)\alpha_0 - \mathbf{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\le & \sqrt{\max_{1\le k,l\le 2p}\max_{1\le i\le n}\tfrac{1}{n}\textstyle\sum_{i=1}^n(\mathbf{X}_i^{(k)}(\hat{\tau})\mathbf{X}_i^{(l)}(\hat{\tau}))^2\left(\max_{1\le k\le 2p}\mathbf{X}_i^{(k)}(\tau_0)\right)^2\|\alpha_0\|_1^2}\|\mathbf{X}(\tau_0)\alpha_0 - \mathbf{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\le & \sqrt{\max_{1\le k,l,j\le p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\|\alpha_0\|_1^2\cdot 1\left(Q_i<\tau_0\right)\cdot 1\left(Q_i<\hat{\tau}\right)}\|\mathbf{X}(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \mathbf{X}(\tau_0)\alpha_0\|_n \\
\le & O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

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

$$\begin{aligned}
& \max_{1\le k,l\le 2p} |\tfrac{2}{n}\textstyle\sum_{i=1}^n\left(\mathbf{X}_i'(\tau_0)\alpha_0 - \mathbf{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)(\mathbf{X}_i^{(k)}(\hat{\tau})\mathbf{X}_i^{(l)}(\hat{\tau}))U_i| \\
\le & 2\sqrt{\max_{1\le k,l\le p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}U_i)^2\cdot 1\left(Q_i<\hat{\tau}\right)}\|\mathbf{X}(\tau_0)\alpha_0 - \mathbf{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\le & O_p\left(\sqrt{s_0}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

Hence,

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

$$\begin{aligned}
& \|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g\|_\infty \\
\le & \|g'\widehat{\Theta}(\hat{\tau})\|_1^2\|\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty \\
\le & \left(\textstyle\sum_{j\in H}|g_j|\max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\widehat{\Theta}(\hat{\tau})\|_1\right)^2\|\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty \\
= & O_p\left(h\bar{s}\right)O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

Step 2.Next, we show that

$$\|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g\|_\infty = o_p(1)$$

Note that

$$\begin{aligned} & \tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu} \\ = \ & \tfrac{1}{n}\textstyle\sum_{i=1}^n \mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2 - E\left[\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})U_i^2\right] \end{aligned}$$

Recall Lemma 18,

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

Therefore

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

Step 3.Next, we show that

$$\|g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty = o_p(1)$$

By Lemma 6.1 in van de Geer et al. (2014)

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

As $\|\Sigma(\hat{\tau})_{xu}\|_\infty = \max_{1\le l,k\le 2p} E\left[\mathbf{X}_i^{(k)}(\hat{\tau})\mathbf{X}_i^{(l)}(\hat{\tau})u_i^2\right]$, $\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})$ and $\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Theta)$ are assumed bounded from Assumption 6,

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

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

Furthermore,

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

Finally, by Assumption 6 (i),

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

$\square$

Lemma 20. *Suppose that Assumption1 to 6 be satisfied, conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$,* *then for any* $\widehat{\tau} \in \widehat{\mathbb{T}}$,

$$\|g'\widehat{\Theta}(\hat{\tau})\widehat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\|_\infty = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right) + O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right) = o_p(1)$$

Proof of Lemma 20.

$$\begin{aligned}
& \|g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\|_\infty \\
= \quad & \|\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})' - \Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})' + g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g \\
& + g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\|_\infty \\
\leq \quad & \|g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty + \|g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g\|_\infty \\
& + \|g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\|_\infty
\end{aligned}$$

To prove this lemma,we need prove the followings

$$\|g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty = o_p(1)$$

$$\|g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g\|_\infty = o_p(1)$$

$$\|g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\|_\infty = o_p(1)$$

Firstly, since $\Theta(\hat{\tau})$ is sysmetric,$\|\Theta(\hat{\tau})'g\|_1 = \|g'\Theta(\hat{\tau})\|_1$, also $\|\Sigma(\hat{\tau})_{xu}\|_\infty$ is bounded by Assumption 1, combine with (8.79)

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

Secondly, as $\|g'\left(\Theta(\hat{\tau}) - \Theta(\tau_0)\right)\|_1 = \|\left(\Theta(\hat{\tau}) - \Theta(\tau_0)\right)'g\|_1$

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

Note,

$$
\begin{aligned}
& \Sigma(\hat{\tau})_{xu} - \Sigma(\tau_0)_{xu} \\
= & \quad E\left[\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau})u_i^2\right] - E\left[\mathbf{X}_i(\tau_0)\mathbf{X}_i'(\tau_0)u_i^2\right] \\
= & \quad E\left[\mathbf{X}_i(\hat{\tau})\mathbf{X}_i'(\hat{\tau}) - \mathbf{X}_i(\tau_0)\mathbf{X}_i'(\tau_0)\right] E\left[u_i^2\right] \\
= & \quad \left(\Sigma(\hat{\tau}) - \Sigma(\tau_0)\right) E\left[u_i^2\right] \\
= & \quad E\left[X_i X_i'\right] E\left[u_i^2\right] |\tau_0 - \widehat{\tau}|
\end{aligned}
$$

and

$$
\|\Sigma(\hat{\tau})_{xu} - \Sigma(\tau_0)_{xu}\|_\infty = \|\Sigma(\hat{\tau}) - \Sigma(\tau_0)\|_\infty E\left[u_i^2\right] = O_p\left(s_0 \frac{\log p}{n}\right)
$$

$$
\begin{aligned}
& \|g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\|_\infty \\
\leq & \quad \|g'\Theta(\tau_0)\left(\Sigma(\hat{\tau})_{xu} - \Sigma(\tau_0)_{xu}\right)\Theta(\tau_0)'g\|_\infty \\
\leq & \quad \|g'\Theta(\tau_0)\|_1^2 \|\Sigma(\hat{\tau})_{xu} - \Sigma(\tau_0)_{xu}\|_\infty \\
= & \quad O_p\left(h\bar{s}\right) O_p\left(s_0 \tfrac{\log p}{n}\right) = O_p\left(h\bar{s}s_0 \tfrac{\log p}{n}\right)
\end{aligned}
$$

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

Combine with Lemma 19

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

□

Proof of Theorem 4 Case II: fixed threshold. We show that the ratio

$$
t = \frac{\sqrt{n}g'\left(\hat{a}(\hat{\tau}) - \alpha_0\right)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \tag{8.82}
$$

is asymptotically standard normal. First, note that by (5.5) one can write

$$
t = t_1 + t_2,
$$

where

$$t_1 = \frac{g'\widehat{\Theta}(\tau_0)\mathbf{X}'(\tau_0)U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \text{ and}$$

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

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

Step 1. We first show that t_1 is asymptotically standard normal.

Step 1.1) To show that t_1 is asymptotically standard normal we first show that

$$t_1' = \frac{g'\Theta(\tau_0)X'(\tau_0)U/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}$$

converges in distribution to a standard normal where $\Sigma_{xu}(\tau_0) = \frac{1}{n}\sum_{i=1}^n E(\mathbf{X}_i((\tau_0))\mathbf{X}_i(\tau_0)'U_i^2)$. Then we show that t_1' and t_1 are asymptotically equivalent. Note that, using $E(U_i|X_i) = 0$ for all $i = 1, ..., n$, we obtain

$$E\left[\frac{g'\Theta(\tau_0)\mathbf{X}'(\tau_0)U/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\right] = E\left[\frac{g'\Theta(\tau_0)\sum_{i=1}^n \mathbf{X}_i(\tau_0)U_i/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\right] = 0, \tag{8.83}$$

and

$$E\left[\frac{g'\Theta(\tau_0)\mathbf{X}'(\tau_0)U/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\right]^2 = E\left[\frac{g'\Theta(\tau_0)\sum_{i=1}^n \mathbf{X}_i(\tau_0)U_i/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\right]^2 = 1.$$

Hence, to use Lyapounov's central limit theorem, we check the conditions for a sequence of independent random variables, it suffices to show that for some $\varepsilon > 0$

$$\frac{\sum_{i=1}^n E|g'\Theta(\tau_0)\mathbf{X}_i'(\tau_0)U_i/n^{1/2}|^{2+\varepsilon}}{(g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g)^{1+\varepsilon/2}} \to 0.$$

Let $\bar{S} = \cup_{j \in H} S_j$, then the cardinality $|\bar{S}| = p \wedge h\bar{s}$

$$
\begin{aligned}
E\left|g'\Theta(\tau_0)\mathbf{X}_i'(\tau_0)U_i/n^{1/2}\right|^{2+\varepsilon} &\leq E\left|\left[\|g'\Theta(\tau_0)/n^{1/2}\|_1 \max_{j\in\bar{S}}\left(X_i^{(j)}U_i\right)\right]\right|^{2+\varepsilon} \\
&\leq E\left[\|g'\Theta(\tau_0)/n^{1/2}\|_1^{2+\varepsilon} \max_{j\in\bar{S}}\left|X_i^{(j)}U_i\right|^{2+\varepsilon}\right] \\
&\leq \|g'\Theta(\tau_0)/n^{1/2}\|_1^{2+\varepsilon} E\left[\max_{j\in\bar{S}}\left|X_i^{(j)}U_i\right|^{2+\varepsilon}\right] \\
&\leq \|g'\Theta(\tau_0)/n^{1/2}\|_1^{2+\varepsilon} E\left[\sum_{j\in\bar{S}}\left|X_i^{(j)}U_i\right|^{2+\varepsilon}\right] \\
&\leq \|g'\Theta(\tau_0)/n^{1/2}\|_1^{2+\varepsilon}(p\wedge h\bar{s}) \max_{j\in\bar{S}} E\left[\left|X_i^{(j)}U_i\right|^{2+\varepsilon}\right] \\
&\leq \|g'\Theta(\tau_0)/n^{1/2}\|_1^{2+\varepsilon}(p\wedge h\bar{s}) \max_{1\leq j\leq p} E\left[\left|X_i^{(j)}U_i\right|^{2+\varepsilon}\right] \\
&= O_p\left(\frac{(h\bar{s})^{2+\varepsilon/2}}{n^{1+\varepsilon/2}}\right) \max_{1\leq j\leq p} E\left[\left(X_i^{(j)}U_i\right)^{2+\varepsilon}\right] \wedge O_p\left(\frac{(h\bar{s})^{1+\varepsilon/2}p}{n^{1+\varepsilon/2}}\right) \max_{1\leq j\leq p} E\left[\left(X_i^{(j)}U_i\right)^{2+\varepsilon}\right]
\end{aligned}
$$

where the 1st inequality follows from the Holder's inequality.

By Cauchy–Schwarz inequality and take $\varepsilon = 2$

$$
E\left[\left(X_i^{(j)}U_i\right)^4\right] \leq E\left[\left(X_i^{(j)}\right)^4\right] E\left[(U_i)^4\right]
$$

is bounded by assumption 1.

Thus $\sum_{i=1}^n E\left|g'\Theta(\tau_0)\mathbf{X}_i(\tau_0)U_i/n^{1/2}\right|^4 = O_p\left(\frac{(h\bar{s})^3}{n^2}\right) \wedge O_p\left(\frac{(h\bar{s})^2 p}{n^2}\right) = o_p(1)$ by Assumption 6 (iv)

Next, we show that $g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g$ is asymptotically bounded away from zero. Clearly,

$$
\begin{aligned}
g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g &\geq \kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|g'\Theta(\tau_0)\|_2^2 \\
&\geq \kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|g'\|_2^2 \kappa(\bar{s}, c_0, \mathbb{T}, \Theta)^2 \\
&= \kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\kappa(\bar{s}, c_0, \mathbb{T}, \Theta)^2,
\end{aligned}
\tag{8.84}
$$

which is bounded away from zero since $\kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})$ and $\kappa(\bar{s}, c_0, \mathbb{T}, \Theta)$ are bounded away from zero by Assumption 6 (iv). Hence, the Lyapunov condition is satisfied and t_1' converges in distribution to a standard normal.

Step 1.2). Let

$$
t_1'' = \frac{g'\Theta(\tau_0)X'(\tau_0)u/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})g}}
$$

$$
\begin{aligned}
&\|g'\widehat{\Theta}(\tau_0)X'(\tau_0)U/n^{1/2} - g'\Theta(\tau_0)X'(\tau_0)U/n^{1/2}\|_\infty \\
\le&\|g'\left(\hat{\Theta}(\tau_0) - \Theta(\tau_0)\right)\|_1\|\mathbf{X}(\tau_0)U/n^{1/2}\|_\infty
\end{aligned}
$$

Conditional on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$ and by Lemma 2

$$
=O_p(\sqrt{h}\bar{s}\frac{\sqrt{\log p}}{\sqrt{n}})O_p(\sqrt{\log p}) = O_p(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}) = o_p(1)
$$

$$
\begin{aligned}
|t_1'' - t_1| =&\frac{g'\left(\widehat{\Theta}(\tau_0)X'(\tau_0)U/n^{1/2} - \Theta(\tau_0)X'(\tau_0)U/n^{1/2}\right)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \\
\le&o_p(1)\frac{1}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}
\end{aligned}
$$

$$
\begin{aligned}
|t_1' - t_1''| =&\frac{\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} - \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)g'\Theta(\tau_0)X'(\tau_0)u/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}} \\
=&\frac{\left(g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\right)g'\Theta(\tau_0)X'(\tau_0)u/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)} \\
\le&\frac{\|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\|_\infty g'\Theta(\tau_0)X'(\tau_0)u/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)} \\
=&o_p(1)t_1'\frac{1}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)}
\end{aligned}
$$

by Lemma 20.

Then combine the above two,

$$
\begin{aligned}
t_1 = t_1'\pm&o_p(1)t_1'\frac{1}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)} \\
\pm&o_p(1)\frac{1}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}
\end{aligned}
$$

Step 2. By Lemma 15, 16 and 17,

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

Step 3. Finally, by Slutsky's theorem

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

therefore Lemma 20 implies that

$$\sup_{\alpha_0 \in \mathcal{A}_{\ell_0}(s_0)} |\hat{\Theta}(\hat{\tau})\hat{\Sigma}_{xu}(\hat{\tau})\hat{\Theta}'(\hat{\tau}) - \Theta(\tau_0)\Sigma_{xu}(\tau_0)\Theta'(\tau_0)| = o_p(1) \tag{8.85}$$

where

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

□

8.6 Asymptotic Distribution of Threshold Parameter

To develop the asymptotic properties of the threshold parameter estimator, we rely on the empirical process results introduced by Hansen (2000) and adopt the shrinking-threshold-effect framework. In this framework, the threshold effect diminishes as the sample size tends to infinity. By constructing a likelihood ratio (LR) statistic, we can derive inferences regarding the threshold parameter.

Assumption 7. *(i)For some fixed $\delta_0^* < \infty$ and $0 < \varphi < \frac{1}{2}$, $\delta_0 = n^{-\varphi}\delta_0^*$ and $n^{-\varphi}\|\delta_0^*\|_1 > 0$.*
(ii) $E\left[X_iX_i'U_i^2|Q_i = \tau\right]$ is continuous and bounded when τ is in a neighborhood of τ_0.
(iii)For any $\eta > 0$, and $\tau_1, \tau_2 \in \mathbb{T}$ such that wpa1

$$\sup_{|\tau-\tau_0|<\eta} \left|\frac{1}{n}\sum_{i=1}^n U_iX_i'\delta_0\left[1\left(Q_i<\tau_0\right) - 1\left(Q_i<\tau\right)\right]\right| \le \frac{\lambda(\sqrt{\eta})^\varpi}{2} \tag{8.86}$$

$$\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\eta} \frac{1}{n}\sum_{i=1}^n \left|X_i^{(j)}X_i^{(l)}\right| \left|1\left(Q_i<\tau_0\right) - 1\left(Q_i<\tau\right)\right| \le C_5(\eta)^\varpi, \tag{8.87}$$

$$\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\eta} \|\delta_0\|_1 \left|\frac{1}{n}\sum_{i=1}^n U_iX_i^{(j)}\left[1\left(Q_i<\tau_0\right) - 1\left(Q_i<\tau\right)\right]\right| \le \frac{\lambda(\sqrt{\eta})^\varpi}{2}, \tag{8.88}$$

where $0 < \varphi < \frac{1}{2}$, $\varpi > \frac{1}{1-2\varphi}$ and $\frac{s_0^2 \log p}{n^{\varpi(1-2\varphi)-1}}\|\delta_0\|_1^2 = o_p(1)$, $\sqrt{\frac{logp}{n^{1-2\varphi}}} = o_p(1)$, $\|\delta_0\|_1^4\frac{\log p}{n} = o_p(1)$.

Assumption 7 is an extension of the fixed dimension case in the literature when working with a fixed regressor design(e.g., Hansen (2000)). Assumption 7 (i) has been widely used in the threshold model to obtain a tractable asymptotic distribution for the least squares estimator of τ(e.g., Hansen (2000)) The re-normalization is to force δ_0 to be small, reducing the information in the sample concerning the threshold and hence slowing down the rate of convergence of the threshold estimate. This assumption need not be viewed as very restrictive since the rate at which δ_0 decreases to zero can be set quite low. It does suggest, however, that the asymptotic approximation is more likely to provide good approximations when δ_0 is small relative to the case where δ_0 is large. The unknown parameter $0 < \varphi < \frac{1}{2}$ reflects the difficulty of estimating and affects the identification and estimation of the change point. Both the rate of convergence and the asymptotic distribution depend on. In Assumption 7 (iii), (8.88) implies (8.86).

The following arguments are parallel to those in Lemma 11, Lemma 13, Theorem 1, and Theorem 2 of Hansen (2000). To describe the asymptotic distribution, we introduce additional notations. For any $v \in \Psi$, an arbitrary compact set, let

$$\Delta_n(v) = n\left[S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0) + \lambda \left\|\mathbf{D}(\tau_0)\widehat{\alpha}(\widehat{\tau})\right\|_1\right] - n\left[S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0 + \frac{v}{n^{1-2\varphi}}) + \lambda \left\|\mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\widehat{\alpha}(\widehat{\tau})\right\|_1\right]$$

We can then derive the process using (2.4)

(8.89)

$$\begin{aligned}
&\operatorname{argmax}_v \Delta_n(v)\\
=&\operatorname{argmax}_v S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0) + \lambda \left\|\mathbf{D}(\tau_0)\widehat{\alpha}(\widehat{\tau})\right\|_1 - S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0 + \frac{v}{n^{1-2\varphi}}) - \lambda \left\|\mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\widehat{\alpha}(\widehat{\tau})\right\|_1\\
=&n^{1-2\varphi}(\widehat{\tau} - \tau_0).
\end{aligned}$$

Let $\widehat{v} = n^{1-2\varphi}(\widehat{\tau} - \tau_0)$.

$$\begin{aligned}
&\Delta_n(v)\\
=&\widehat{\delta}(\widehat{\tau})'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'(X(\tau_0 + \frac{v}{n^{1-2\varphi}})\widehat{\delta}(\widehat{\tau}) - 2\widehat{\delta}(\widehat{\tau})(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U\\
&+2\widehat{\delta}(\widehat{\tau})(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'(X(\tau_0 + \frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau}) - \beta_0))\\
&+n\lambda \left\|\mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau})\right\|_1 - n\lambda \left\|\mathbf{D}(\tau_0)\alpha(\widehat{\tau})\right\|_1\\
=&\delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'X(\tau_0 + \frac{v}{n^{1-2\varphi}})\delta_0 - 2\delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'U\\
&-2(\widehat{\delta}(\widehat{\tau}) - \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'U\\
&+2\widehat{\delta}(\widehat{\tau})'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'X(\tau_0 + \frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau}) - \beta_0)\\
&+(\widehat{\delta}(\widehat{\tau}) + \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'X(\tau_0 + \frac{v}{n^{1-2\varphi}})(\widehat{\delta}(\widehat{\tau}) - \delta_0)\\
&+n\lambda \left\|\mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau})\right\|_1 - n\lambda \left\|\mathbf{D}(\tau_0)\alpha(\widehat{\tau})\right\|_1
\end{aligned} \tag{8.90}$$

Regarding the second term in (8.90), we introduce additional notations. Let $R_n(v) = \frac{\sqrt{n^{1-2\varphi}}}{\sqrt{n}} \sum_{i=1}^n \delta_0^{*\prime}(X_i(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X_i(\tau_0))U_i$ and $V_n(v) = \frac{n^{1-2\varphi}}{n} \sum_{i=1}^n \delta_0^{*\prime}(X_i(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X_i(\tau_0))(X_i(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X_i(\tau_0))'\delta_0^* U_i^2$. First, we show for any given v the convergence of the finite-dimensional distributions of $R_n(v)$ to those of $B(v)$. It suffices to show the first and second terms in (8.90) converge somewhere correspondingly, and then show the convergence of $\Delta_n(v)$.

Lemma 21. *Under Assumption 1,2 and 7, for any $v \in \Psi$, a arbitrary compact set,*

$$\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0 \xrightarrow{p} v\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*$$

and

$$R_n(v) \rightsquigarrow B(v),$$

where $B(v)$ can be written as $\sqrt{\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*}W(v)$, and $W(v)$ is a standard Brownian motion.

The notation $R_n(v) \rightsquigarrow B(v)$ defines a general concept of convergence in distribution introduced by Dudley (1985).

Proof. To show the first part of the lemma,

$$\begin{aligned}
&E\left[\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0\right]\\
=&\delta_0'E\left[(X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0)'(X(\tau_0+\frac{v}{n^{1-2\varphi}})\right]\delta_0\\
=&\frac{n^{1-2\varphi}}{n}\delta_0^{*\prime}E\left[X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0)'X(\tau_0)\right]\delta_0^*\\
=&n^{1-2\varphi}\delta_0^{*\prime}E\left[\frac{1}{n}X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0)'X(\tau_0)\right]\delta_0^*\\
=&n^{1-2\varphi}\delta_0^{*\prime}E\left[\frac{1}{n}\sum_{i=1}^n X_iX_i'\left[\mathbf{1}\left(Q_i<\tau_0+\frac{v}{n^{1-2\varphi}}\right)-\mathbf{1}\left(Q_i<\tau_0\right)\right]U_i^2\right]\delta_0^*\\
=&n^{1-2\varphi}\delta_0^{*\prime}E\left[X_iX_i'|\tau_0\le Q_i\le\tau_0+\frac{v}{n^{1-2\varphi}}\right]\delta_0^*\\
\xrightarrow{p}&v\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*
\end{aligned} \tag{8.91}$$

as $n\to\infty$.

(8.92)

$$\begin{aligned}
&E\left[\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'(X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0-E\big[\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0\big]\right]^2\\
=&E\left[\delta_0'\Big[X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-E\big[X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})\big]-X(\tau_0)'X(\tau_0)+E\big[X(\tau_0)'X(\tau_0)\big]\Big]\delta_0\right]^2\\
\le&E\left[\|\delta_0\|_1^2\Big[\|\frac{1}{n}X_i(\tau_0+v)'X_i(\tau_0+v)-E\left[X_i(\tau_0+v)'X_i(\tau_0+v)\right]\|_\infty+\|\frac{1}{n}X_i(\tau_0)'X_i(\tau_0)-E\left[X_i(\tau_0)'X_i(\tau_0)\right]\|_\infty\Big]\right]^2\\
\xrightarrow{p}&\|\delta_0\|_1^4O_p(\frac{\log p}{n})
\end{aligned}$$

where we used Lemma 6 in the last step. Combine the above with Markov's inequality,

$$\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0 \xrightarrow{p} v\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*.$$

Our proof proceeds by establishing the convergence of the finite-dimensional distributions of $R_n(v)$ to those of $B(v)$ for any given v, then extending that by showing the tightness of $R_n(v)$.

$$
\begin{aligned}
&E\left[V_n(v)\right]\\
&=E\left[\frac{n^{1-2\varphi}}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[\mathbf{1}\left(Q_i<\tau_0+\frac{v}{n^{1-2\varphi}}\right)-\mathbf{1}\left(Q_i<\tau_0\right)\right]U_i^2\right]\\
&=\sum_{i=1}^{n}E\left[\frac{n^{1-2\varphi}}{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[\mathbf{1}\left(Q_i<\tau_0+\frac{v}{n^{1-2\varphi}}\right)-\mathbf{1}\left(Q_i<\tau_0\right)\right]U_i^2\right]\\
&\xrightarrow{p}v\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*
\end{aligned}
\tag{8.93}
$$

(8.94)

$$
\begin{aligned}
&E\left[V_n(v)-E[V_n(v)]\right]^2\\
&=E\left[\frac{n^{1-2\varphi}}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]U_i^2-E\left[\frac{n^{1-2\varphi}}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]U_i^2\right]\right]\\
&=n^{1-2\varphi}E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]U_i^2-E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]U_i^2\right]\right]\\
&=n^{1-2\varphi}E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]U_i^2-E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]U_i^2\right]\right]\\
&\leq n^{1-2\varphi}E\left[\|\delta_0^*\|_1^2\left[\|\frac{1}{n}X_i(\tau_0+v)'X_i(\tau_0+v)-E\left[X_i(\tau_0+v)'X_i(\tau_0+v)\right]\|_\infty+\|\frac{1}{n}X_i(\tau_0)'X_i(\tau_0)-E\left[X_i(\tau_0)'X_i(\tau_0)\right]\|_\infty\right]\right]\\
&\xrightarrow{p}0
\end{aligned}
$$

which establishes that $V_n(v)\xrightarrow{p}|v|\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*$ by Markov's inequality.

Since $\{X_i,U_i,Q_i\}_{i=1}^n$ is an independent and identically distributed sequence, $E\left[R_n(v)\right]=0$. We conclude that $R_n(v)\xrightarrow{d}N(0,|v|\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*)$ for any fix v. This argument can be extended to include any finite collection $[v_1,\cdots v_k]$, to yield the convergence of the finite-dimensional distributions of $R_n(v)$ to those of $B(v)$.

Then we show the tightness of $R_n(v)$. Fix $\eta>0$ and set $\tau_1=\tau_0+\frac{v_1}{n^{1-2\varphi}}$, then by Assumption 7(8.86),

$$
\sup_{v_1\leq v\leq v_1+\eta}R_n(v)-R_n(v_1)\leq\sup_{v_1\leq v\leq v_1+\eta}R_n(v)-R_n(0)+\sup_{v_1\leq v\leq v_1+\eta}R_n(0)-R_n(v_1)\leq\frac{1}{2}\lambda n^{1-\varpi(\frac{1}{2}-\varphi)}[(\sqrt{v})^{\varpi}+(\sqrt{v_1})^{\varpi}].
$$

Thus, $\mathbb{P}\{\sup_{v_1\leq v\leq v_1+\eta}|R_n(v)-R_n(v_1)|>\lambda n^{1-\frac{1}{2}\varpi(1-2\varphi)}(\sqrt{v})^{\varpi}\}\to0$, as $n\to\infty$. So $R_n(v)$ is tight.

As $R_n(v)$ is tight, we conclude that $R_n(v)\rightsquigarrow B(v)$. □

In next part, we present an auxiliary technical lemma and its proof. We start with some matrix norm inequalities. Let A be a generic $q \times p$ matrix and x a $p \times 1$ vector and z a $q \times 1$ vector.

Lemma 22.

$$x'Az \leq \|x\|_1 \|A\|_\infty \|z\|_1$$

Proof. Observe that

$$x'Az \leq \|x\|_1 \|Az\|_\infty \leq \|x\|_1 \|A\|_\infty \|z\|_1$$

by Hölder's inequality. $\square$

Lemma 23. *Under Assumption 1,2 and 7, for any $v \in \Psi$,on any compact set,*

$$\Delta_n(v) \rightsquigarrow \Delta(v),$$

where $\Delta(v) = -|v| \delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right] \delta_0^* + 2\sqrt{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right] \delta_0^*} W(v)$, and $W(v)$ is a standard Brownian motion.

Proof. Rearranging (8.90), yields

$$\begin{aligned}
\Delta_n(v) =& \delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))X(\tau_0 + \frac{v}{n^{1-2\varphi}})'\delta_0 - 2\delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U \\
&-2(\widehat{\delta}(\widehat{\tau}) - \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U \\
&+2\widehat{\delta}(\widehat{\tau})'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))X(\tau_0 + \frac{v}{n^{1-2\varphi}})'(\widehat{\beta}(\widehat{\tau}) - \beta_0) \\
&+(\widehat{\delta}(\widehat{\tau}) + \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))X(\tau_0 + \frac{v}{n^{1-2\varphi}})'(\widehat{\delta}(\widehat{\tau}) - \delta_0) \\
&+n\lambda \left\| \mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau}) \right\|_1 - n\lambda \left\| \mathbf{D}(\tau_0)\alpha(\widehat{\tau}) \right\|_1 \\
=& \delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))X(\tau_0 + \frac{v}{n^{1-2\varphi}})'\delta_0 + 2R_n(v) + \Upsilon(v),
\end{aligned} \tag{8.95}$$

where

$$\begin{aligned}
\Upsilon(v) = &-2(\widehat{\delta}(\widehat{\tau}) - \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U \\
&+2\widehat{\delta}(\widehat{\tau})'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))X(\tau_0 + \frac{v}{n^{1-2\varphi}})'(\widehat{\beta}(\widehat{\tau}) - \beta_0) \\
&+(\widehat{\delta}(\widehat{\tau}) + \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))X(\tau_0 + \frac{v}{n^{1-2\varphi}})'(\widehat{\delta}(\widehat{\tau}) - \delta_0) \\
&+n\lambda \left\| \mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau}) \right\|_1 - n\lambda \left\| \mathbf{D}(\tau_0)\alpha(\widehat{\tau}) \right\|_1
\end{aligned} \tag{8.96}$$

It suffices to show $\Upsilon(v) \Rightarrow 0$. Note that by triangle inequality and Hölder's inequality

$$
\begin{aligned}
& n\lambda \left\| \mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau}) \right\|_1 - n\lambda \left\| \mathbf{D}(\tau_0)\alpha(\widehat{\tau}) \right\|_1 \\
\leq & n\lambda \left\| \left(\mathbf{D}(\tau_0 + \frac{v}{n^{1-2\varphi}}) - \mathbf{D}(\tau_0) \right) \alpha(\widehat{\tau}) \right\|_1 \\
= & n\lambda \left| \sum_{j=1}^{p} \left(\|X^{(j)}(\tau_0 + \frac{v}{n^{1-2\varphi}})\|_n - \|X^{(j)}(\tau_0)\|_n \right) \widehat{\delta}^{(j)}(\widehat{\tau}) \right| \\
\leq & n\lambda \left| \sum_{j=1}^{p} \|X^{(j)}(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X^{(j)}(\tau_0)\|_n \widehat{\delta}^{(j)}(\widehat{\tau}) \right| \\
\leq & n\lambda \left(\max_{j=1\cdots p} \left\| X^{(j)}(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X^{(j)}(\tau_0) \right\|_n \right) \left\| \widehat{\delta}(\widehat{\tau}) \right\|_1 \\
\leq & n\lambda \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \left\| \widehat{\delta}(\widehat{\tau}) \right\|_1 \\
\leq & n\lambda \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \left(\|\delta_0\|_1 + Cs_0\lambda \right) \\
\leq & n\lambda \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \left(n^{-\varphi} \|\delta_0^*\|_1 + Cs_0\lambda \right) \\
\leq & Cv^{\frac{\varpi}{2}} \left(\frac{\sqrt{\log p}\, \|\delta_0^*\|_1}{\sqrt{n^{(1-2\varphi)(\varpi-1)}}} + \frac{s_0 \log p}{\sqrt{n^{(1-2\varphi)\varpi}}} \right) \\
= & O_p \left(\frac{\sqrt{\log p}\, \|\delta_0^*\|_1}{\sqrt{n^{(1-2\varphi)(\varpi-1)}}} + \frac{s_0 \log p}{\sqrt{n^{(1-2\varphi)\varpi}}} \right)
\end{aligned}
\tag{8.97}
$$

where the last inequality is by Assumption 7 (8.87), Theorem 1 or 2.

Note that by Hölder's inequality and Assumption 7 (8.88)

$$
\begin{aligned}
& 2(\widehat{\delta}(\widehat{\tau}) - \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U \\
\leq & 2\|\widehat{\delta}(\widehat{\tau}) - \delta_0\|_1 \sup_{1 \leq j \leq p} \sup_{|\tau - \tau_0| < \frac{v}{n^{1-2\varphi}}} \left| \sum_{i=1}^{n} U_i X_i^{(j)} \left[1\left(Q_i < \tau_0\right) - 1\left(Q_i < \tau\right) \right] \right| \\
\leq & \|\widehat{\delta}(\widehat{\tau}) - \delta_0\|_1 n \frac{\lambda (\sqrt{\frac{v}{n^{1-2\varphi}}})^{\varpi}}{\|\delta_0\|_1} \\
\leq & C \frac{s_0 \log p}{n^{-\varphi} \|\delta_0^*\|_1} \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \\
= & Cv^{\frac{\varpi}{2}} \frac{s_0 \log p}{\|\delta_0^*\|_1 \sqrt{n^{(1-2\varphi)\varpi - 2\varpi}}} \\
= & O_p \left(\frac{s_0 \log p}{\|\delta_0^*\|_1 \sqrt{n^{(1-2\varphi)\varpi - 2\varpi}}} \right).
\end{aligned}
\tag{8.98}
$$

By Lemma 22 and Assumption 7 (8.87)

$$
\begin{aligned}
&2\widehat{\delta}(\widehat{\tau})'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau})-\beta_0)\\
=&2(\widehat{\delta}(\widehat{\tau})'-\delta_0'+\delta_0')(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau})-\beta_0)\\
\leq&2\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\sup_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|\|\widehat{\beta}(\widehat{\tau})-\beta_0\|_1\\
&+2\|\delta_0\|_1\sup_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|\|\widehat{\beta}(\widehat{\tau})-\beta_0\|_1\\
\leq&C(\frac{v}{n^{1-2\varphi}})^{\varpi}n(\frac{s_0\sqrt{\log p}}{\sqrt{n}})^2+C(\frac{v}{n^{1-2\varphi}})^{\varpi}n\frac{s_0\sqrt{\log p}}{\sqrt{n}}\|\delta_0\|_1
\end{aligned}
\tag{8.99}
$$

By Lemma 22 and Assumption 7 (8.87)

$$
\begin{aligned}
&(\widehat{\delta}(\widehat{\tau})+\delta_0)'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))X(\tau_0+\frac{v}{n^{1-2\varphi}})'(\widehat{\delta}(\widehat{\tau})-\delta_0)\\
=&(\widehat{\delta}(\widehat{\tau})-\delta_0+2\delta_0)'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))X(\tau_0+\frac{v}{n^{1-2\varphi}})'(\widehat{\delta}(\widehat{\tau})-\delta_0)\\
\leq&\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\sup_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\\
&+2\|\delta_0\|_1\sup_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}\left[1\left(Q_i<\tau_0\right)-1\left(Q_i<\tau\right)\right]\right|\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\\
\leq&C(\frac{v}{n^{1-2\varphi}})^{\varpi}n(\frac{s_0\sqrt{\log p}}{\sqrt{n}})^2+C(\frac{v}{n^{1-2\varphi}})^{\varpi}n\frac{s_0\sqrt{\log p}}{\sqrt{n}}\|\delta_0\|_1.
\end{aligned}
\tag{8.100}
$$

Thus, $\Upsilon(v)\rightsquigarrow 0$. Combine with Lemma 21,

$$\Delta_n(v)\rightsquigarrow\Delta(v).$$

□

Lemma 24. *Under Assumption 1,2 and 7,*

$$n^{1-2\varphi}(\widehat{\tau}-\tau_0) \xrightarrow{d} \omega T,$$

where $\omega = \frac{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right]\delta_0^*}{(\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^*)^2}$ *and* $T = \mathrm{argmax}_r \left[-\frac{|r|}{2} + W(r)\right]$.
$W(r)$ is defined as a two-sided Brownian motion on the real line,

$$W(r) = \begin{cases} W_1(r), & \text{if } r \geq 0, \\ W_2(r), & \text{if } r < 0. \end{cases}$$

where $W_1(r)$ and $W_2(r)$ are independent standard Brownian motions on $[0, \infty)$.

Proof. By Theorem 2,

$$n^{1-2\varphi}(\widehat{\tau}-\tau_0) \leq C\frac{s_0 \log p}{n^{2\varphi}} = O_p(1)$$

and by Lemma 23

$$\Delta_n(v) \rightsquigarrow \Delta(v).$$

Next, as $\lim_{v\to\infty} \frac{W(v)}{v} = 0$, $\lim_{|v|\to\infty} \Delta(v) = -\infty$. Then the limit functional $\Delta(v)$ is continuous, so $Q(v)$ has a unique maximum. Therefore, all conditions of Theorem 2.7 of Kim and Pollard (1990) are satisfied, which implies

$$n^{1-2\varphi}(\widehat{\tau}-\tau_0) \xrightarrow{d} \mathrm{argmax}_v Q(v).$$

Making the change-of-variables $v = \frac{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right]\delta_0^*}{(\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^*)^2} r$, we can re-write the asymptotic distribution as

$$\mathrm{argmax}_v Q(v) = \frac{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right]\delta_0^*}{(\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^*)^2} \mathrm{argmax}_r \left[-\frac{|r|}{2} + W(r)\right]$$

□

To test hypothesis $H_0 : \tau = \tau_0$, a standard approach is to use the likelihood ratio statistic. Let $LR_n(\tau) = n\frac{S_n(\widehat{\alpha}(\tau),\tau) + \lambda\|\mathbf{D}(\tau)\widehat{\alpha}(\tau)\|_1 - S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau}) - \lambda\|\mathbf{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1}{S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau}) + \lambda\|\mathbf{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1}$

Lemma 25. *Under Assumption 1,2 and 7,*

$$LR_n(\tau) \xrightarrow{d} \varrho^2 \Lambda,$$

where $\varrho^2 = \frac{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right]\delta_0^*}{\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^* \sigma^2}$ *and* $\Lambda = \max_r \left[-\frac{|r|}{2} + W(r)\right]$.
The distribution function of Λ is $\mathbb{P}\{\Lambda < x\} = (1 - e^{-\frac{x^2}{2}})^2$.

Proof. We note that

$$S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau}) = \frac{1}{n}\sum_{i=1}^{n} \widehat{U}_i(\widehat{\tau})^2$$

and

$$\lambda \left\|\mathbf{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\right\|_1 \xrightarrow{p} 0.$$

$$(S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau})+\lambda\left\|\mathbf{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\right\|_1)LR_n(\tau_0)-\Delta_n(\widehat{v})$$

$$(8.101)\qquad =S_n(\widehat{\alpha}(\tau_0),\tau_0)+\lambda\left\|\mathbf{D}(\tau_0)\widehat{\alpha}(\tau_0)\right\|_1-S_n(\widehat{\alpha}(\widehat{\tau}),\tau_0)-\lambda\left\|\mathbf{D}(\widehat{\tau})\widehat{\alpha}(\tau_0)\right\|_1$$

$$=(\widehat{\alpha}(\tau_0)-\widehat{\alpha}(\widehat{\tau}))'\mathbf{X}(\tau_0)'\mathbf{X}(\tau_0)(\widehat{\alpha}(\tau_0)-\widehat{\alpha}(\widehat{\tau}))\xrightarrow{p}0.$$

$$(8.102)\qquad LR_n(\tau_0)=\frac{\Delta_n(n^{1-2\varphi}(\tau_0-\widehat{\tau}))}{S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau})+\lambda\left\|\mathbf{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\right\|_1}=\frac{Sup_v\Delta_n(v)}{S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau})+\lambda\left\|\mathbf{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\right\|_1}\xrightarrow{d}\frac{Sup_v\Delta(v)}{\sigma^2}$$

by continuous mapping theorem. This limiting distribution equals

$$(8.103)\qquad \begin{aligned}&\frac{1}{\sigma^2}Sup_v\left[-v\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*+2\sqrt{|v|\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*}W(v)\right]\\ =&\frac{\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*}{\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*\sigma^2}\sup_r\left[-\frac{|r|}{2}+W(r)\right]=\varrho^2\Lambda\end{aligned}$$

To find the distribution function of Λ, note that

$$\sup_r\left[-\frac{|r|}{2}+W(r)\right]=2\max\left[\sup_{r>0}[-\frac{|r|}{2}+W(r)],\sup_{r<0}[-\frac{|r|}{2}+W(r)]\right]=2\max\left[\Lambda_+,\Lambda_-\right].$$

which becomes the two-sided Brownian motion, as in Hansen (2000). $[\Lambda_+,\Lambda_-]$ are iid exponential random variables with distribution $\mathbb{P}\{\Lambda_+<x\}=1-e^{-x}$. see Bhattacharya and Brockwell 1976.

Thus

$$\mathbb{P}\{\Lambda<x\}=\mathbb{P}\{2\max\left[\Lambda_+,\Lambda_-\right]<x\}=\mathbb{P}\{2\Lambda_+<x\}\mathbb{P}\{2\Lambda_-<x\}=(1-e^{-\frac{x^2}{2}})^2.$$

□

Our Likelihood ratio test corresponds to a modified version of the LR Test used in Hansen (2000). The asymptotic distribution of Lemma 25 depends on the nuisance parameter ϱ^2, which can be constructed by following Section 3.4 in Hansen (2000). We can then obtain a confidence interval for τ.

8.7 Time Series Model

The above results are allowed for cross-section observations. Here we

Assumption 8. *(i)$\{X_i,Q_i,U_i\}_{i=1}^n$ are sequences of (strictly) stationary and ergodic random variables. Furthermore, $\{X_i\}_{i=1}^n$,$\{U_i\}_{i=1}^n$ are independent. The error terms $E(U_i|X_i,Q_i)=0$ and $E(U_i^2)=\sigma^2$. (ii)For the strong mixing variables X_i, U_i: $\alpha(i)\geq exp(-\sigma^2 ir_0)$, for a positive constant $r_0>0$. (iii)There exists positive constants r_1, r_2,and another set of positive constants b_1,b_2,s_1, $s_2>0$, and for $i=1,\cdots,n$, and $j=1,\cdots,p$ $\mathbb{P}\left\{|U_i|>s_1\right\}\leq exp(-(s_1/b_1)^{r_1})$ and $\mathbb{P}\left\{|X_i^{(j)}|>s_2\right\}\leq exp(-(s_2/b_2)^{r_2})$*

Assumption 8 is relevant for time series applications and is trivially satisfied for independent observations. The assumption of stationary excludes time trends and integrated processes. It is sufficiently flexible to embrace many nonlinear time series processes including threshold auto-regressions. Assumption 8(i) imposes that is a correct specification of the conditional mean. Assumptions 8 (ii) and (iii) are standard assumptions, and are used in Fan et al. (2011) as well.

By utilizing Lemma A.3 and Lemma B.1 of Fan et al. (2011) under Assumption 8 in place of the maximal inequality attributed to Lemma E.1 and E.2 of Chernozhukov et al. (2017) used in all of the preceding proofs, we can show that $\mathbb{P}\{\mathbb{A}_1\}$,$\mathbb{P}\{\mathbb{A}_2\}$, $\mathbb{P}\{\mathbb{A}_3\}$,$\mathbb{P}\{\mathbb{A}_4\}$ and $\mathbb{P}\{\mathbb{A}_5\}$ approach 0 for all sufficiently large n. This implies that our framework encompasses the time series data threshold regression model.

References

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

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls†. *The Review of Economic Studies 81*(2), 608–650.

Belloni, A., V. Chernozhukov, and Y. Wei (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics 34*(4), 606–619.

Belloni, A., C. Hansen, and W. Newey (2019). Simultaneous confidence intervals for high-dimensional linear models with many endogenous variables.

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.

Cai, T., W. Liu, and X. Luo (2011). A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association 106*(494), 594–607.

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. (2021). Generalized linear models with structured sparsity estimators.

Caner, M. and B. E. Hansen (2004). Instrumental variable estimation of a threshold model. *Econometric Theory 20*(5), 813–843.

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.

Caner, M. and A. B. Kock (2019). High dimensional linear gmm.

Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability 45*(4), 2309–2352.

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.

Dudley, R. M. (1985). An extended wichura theorem, definitions of donsker class, and weighted empirical distributions. In A. Beck, R. Dudley, M. Hahn, J. Kuelbs, and M. Marcus (Eds.), *Probability in Banach Spaces V*, Berlin, Heidelberg, pp. 141–178. Springer Berlin Heidelberg.

Fan, J., Y. Liao, and M. Mincheva (2011). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics 39*(6), 3320 – 3356.

Fan, Y. and C. Y. Tang (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(3), 531–552.

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

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

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

Kim, J. and D. Pollard (1990). Cube Root Asymptotics. *The Annals of Statistics 18*(1), 191 – 219.

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.

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

Miao, K., K. Li, and L. Su (2020). Panel threshold models with interactive fixed effects. *Journal of Econometrics 219*(1), 137–170.

Miao, K., L. Su, and W. Wang (2020). Panel threshold regressions with latent group structures. *Journal of Econometrics 214*(2), 451–481.

Semenova, V., M. Goldman, V. Chernozhukov, and M. Taddy (2021). Estimation and inference on heterogeneous treatment effects in high-dimensional dynamic panels.

Seo, M. H. and Y. Shin (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics 195*(2), 169–186.

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.

Yuan, M. (2010, August). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res. 11*, 2261–2286.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(1), 217–242.