A Functional-Coefficient VAR Model for Dynamic Quantiles and Its Application to Constructing Nonparametric Financial Network*

Zongwu Cai[a], Xiyuan Liu[b], Liangjun Su[b]

[a]Department of Economics, University of Kansas, Lawrence, KS 66045, USA

[b]School of Economics and Management, Tshinghua University, Beijing, Beijing 100084, China

Abstract: In this article, we investigate a functional coefficient vector autoregressive model for conditional quantiles, in which the interdependences among tail risks such as Value-at-Risk are allowed to vary smoothly with a variable of general economy. Methodologically, we develop an easy-to-implement three-stage procedure to estimate functionals in the dynamic network system based on basis function approximation, LASSO-type penalties and the local linear smoothing technique. We establish the consistency and the asymptotic normality of the proposed estimator under strongly mixing time series settings. The simulation studies are conducted to show that our new methods work fairly well. The potential of the proposed estimation procedures is demonstrated by an empirical study of constructing and estimating a new type of nonparametric dynamic financial network.

Keywords: Conditional quantile models; Dynamic financial network; Functional coefficient models; Nonparametric estimation; Tensor-product B-spline; VAR modeling.

1 Introduction

Since the seminal work by Koenker and Bassett (1978), quantile regression, also called conditional quantile or regression quantile or dynamic quantile, has become an increasingly popular tool for risk analysis in many fields in economics such as labor economics, macroeconomics and financial risk management. It is well known that quantile regression is concerned with estimating a collection of conditional quantiles over the entire conditional distribution, instead of investigating the conditional mean function of dependent variable. The reader is referred to the

*Contact information: caiz@ku.edu (Z. Cai),liuxy@sem.tsinghua.edu.cn (X. Liu), sulj@sem.tsinghua.edu.cn (L. Su).

review papers by Koenker (2005) and Koenker, Chernozhukov, He and Peng (2017) for more applications of quantile regression.

In the recent two decades, a great deal of attention has been paid to modeling the tail dependence in the financial statistics/econometrics literature. Among the toolkits of quantile methods, dynamic quantile models are naturally suitable for capturing tail dependence. For example, White, Kim and Manganelli (2015) proposed an innovative method to directly estimate the sensitivity of Value-at-Risk (VaR) of a given financial institution to shocks to the whole financial system by constructing a vector autoregressive (VAR) model for dynamic quantiles, while Härdle, Wang and Yu (2016) developed a model to describe the network relationship among VaRs of financial institutions by a flexible nonparametric quantile model with L_1-penalty. Recently, Zhu, Wang, Wang and Härdle (2019) constructed a quantile autoregressive model that embeds the observed dependency structure in a dynamic network.

A first important question arising in modeling dynamic quantiles is how to capture the nonlinearities of tail dependence. To the best of our knowledge, much of the existing literature assumed constant tail dependence in their models. However, numerous studies have documented time-varying risk interdependence in financial time series and discussed their possible origins and relation to risk spillover; see, for example, Billio, Getmansky, Lo and Pelizzon (2012), Diebold and Yìlmaz (2014), Härdle et al. (2016), Yang and Zhou (2017), Liu, Ji and Fan (2017), Ando and Bai (2020), and references therein. The driving force for the variations of risk interdependence may be the institutional changes or the policy interventions, such as the changes of exchange rate systems and the U.S. quantitative easing policy. A second crucial issue is how to model the dependence among conditional quantiles of financial return distributions. Indeed, since the VaR processes are witnessed to be significantly autocorrelated and interdependent with each other by Engle and Manganelli (2004) and White et al. (2015), some important types of tail dependence (e.g., the co-dependence between VaRs) can be excluded if a model does not allow for interdependence among conditional quantiles.

To deal with the aforementioned issues in a simultaneous way, we propose a nonparametric approach involving multivariate dynamic quantile models with nonlinear structures. Different from previous studies, we allow coefficients of the multivariate dynamic quantile models to vary with a smoothing variable, which is chosen based on an economic theory or some data-driven methods. In addition, the model allows the latent quantiles of several random variables to depend on lagged quantiles and other lagged covariates, which forms a VAR model with unobserved autoregressors. Thus, this model is termed as a functional-coefficient VAR model for dynamic quantiles (FCVAR-DQ) and is presented in (1) later. Since coefficients of dynamic quantile models play an important role in reflecting interdependences among dynamic quantiles, under our model setup, one can easily illustrate the variation of tail dependence and its relation with the variable which is of interest.

Our contributions to the literature can be summarized as follows. First, the model setting in (1) is general enough to nest many well-known dynamic quantile models in the literature; see, for example, the conditional autoregressive value at risk (CaViaR) model proposed by Engle and Manganelli (2004) and further studied by Xiao and Koenker (2009), the threshold CaViaR model in Gerlach, Chen and Chan (2011), and the static VAR for VaR model constructed by White et al. (2015). Second, by allowing coefficients to vary with a smoothing variable, a FCVAR-DQ model provides a new tool to estimate the relationship between the time-varying interdependence of risk and the state variable of economy. Third, a new and simple-to-implement estimation procedure is developed for estimating the proposed quantile model with highly nonlinear structure and latent covariates. In addition, large sample theories for the proposed estimator are established to construct confidence intervals for functional coefficients in the empirical study. Finally, our empirical analysis provides a strong evidence to show that interdependences among VaRs of four world major financial indices' return series vary significantly with the change of U.S. dollar index. This empirical finding is new to the literature and cannot be revealed from using a standard constant-coefficient model

With coefficients being functionals and autoregressors being latent conditional quantiles, the proposed FCVAR-DQ model is highly nonlinear and conventional techniques of nonlinear quantile regression are not directly applicable. Indeed, as pointed out by Xiao and Koenker (2009), estimation of specific class of fixed-coefficient CaViaR models involves fitting a restricted nonlinear quantile regression, which is complicated both computationally and theoretically. The estimation of FCVAR-DQ model makes this problem become more challenging, since the FCVAR-DQ model nests the fixed-coefficient CaViaR model as a special case and all coefficients in our model are allowed to vary with a smoothing variable. To overcome these difficulties, we develop a nonparametric three-stage procedure to estimate functional coefficients in the FCVAR-DQ model. In the first and second steps, latent conditional quantiles are approximated via a set of tensor-product B-spline basis functions, and LASSO (least absolute shrinkage and selection operator)-based approaches are applied for dimension reduction and variable selection. In the third stage, we perform locally weighted estimation for functional coefficients using the estimated conditional quantiles as autoregressors.

The rest of this paper is organized as follows. In Section 2, the model setup and the three-stage estimation procedure are presented for the FCVAR-DQ model. In addition, large sample theories for the proposed estimator are investigated in this section too, together with constructing a consistent estimator of the asymptotic covariance matrix and corresponding confidence interval. A Monte Carlo simulation study is conducted in Section 3 to illustrate the finite sample performance of the proposed estimation procedure. In Section 4, our proposed model and its modeling procedure are applied to constructing a novel class of nonparametric financial networks based on the real example. Finally, a conclusion remark is given in Section 5, and the descriptions of some important notations, assumptions, and mathematical proofs are gathered in Appendix.

2 FCVAR Model for Dynamic Quantiles

2.1 Model Setup

Let Y_{it} $(1 \le i \le \kappa,\ 1 \le t \le n)$, a scalar dependent variable, be the ith observation at time t for fixed κ, $\mathcal{F}_{t-1}$ represent the information set up to time $t-1$ for all individuals $1 \le i \le \kappa$, and $q_{\tau,t,i}$ be the τth conditional quantile of Y_{it} given $\mathcal{F}_{t-1}$. Then, we study the following functional-coefficient VAR model for dynamic quantiles, termed as FCVAR-DQ model, given by,

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_{it}) + \sum_{s=1}^{q} \boldsymbol{\gamma}_{i,s,\tau}^{\top}(Z_{it})\boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}_{i,l,\tau}^{\top}(Z_{it})\mathbb{Y}_{t-l} \tag{1}$$

for some p and q, where $\boldsymbol{q}_{\tau,t} = (q_{\tau,t,1}, \ldots, q_{\tau,t,\kappa})^{\top}$ is a $\kappa \times 1$ vector of conditional quantiles at time t, and $\mathbb{Y}_t$ is a $\kappa_1 \times 1$ vector of covariates with fixed κ_1, including possibly some or all of $\{Y_{it}\}_{i=1}^{\kappa}$ and/or some exogenous information $\{x_{it}\}$. In addition, $\gamma_{i0,\tau}(\cdot)$ is a scalar function and is allowed to depend on τ, both $\boldsymbol{\gamma}_{i,s,\tau}(\cdot) = (\gamma_{si1,\tau}(\cdot), \ldots, \gamma_{si\kappa,\tau}(\cdot))^{\top}$ and $\boldsymbol{\beta}_{i,l,\tau}(\cdot) = (\beta_{li1,\tau}(\cdot), \ldots, \beta_{li\kappa_1,\tau}(\cdot))^{\top}$ are $\kappa \times 1$ and $\kappa_1 \times 1$ vectors of functional coefficients, respectively, and they are allowed to depend on τ too. Here, Z_{it} is an observable scalar smoothing variable, which might be one part of $\mathbb{Y}_{t-l}$ and/or other exogenous variables $\{x_{it}\}$ or their lagged variables. Of course, Z_{it} can be an economic index to characterize economic activities. Also, note that Z_{it} can be set as a multivariate variable. In such a case, the estimation procedures and the related theory for the univariate case still hold for multivariate case, but more complicated notations are involved and models with Z_{it} in very high dimension are often not practically useful due to the "curse of dimensionality". In addition, note that both κ and κ_1 in model (1) are assumed to be fixed, which might be relatively strong settings.[1] It is also noted that similar to the setting of the multi-quantile CaViaR model as in White, Kim and Manganelli (2008), one may further generalize

[1]Indeed, in our empirical study, it is sufficient to consider VAR model with finite dimensions in the sense that we are interested in measuring tail dependence among four major financial market indices. If both κ and κ_1 (or one of them) are allowed to diverge (a function of sample size and going to infinity when sample size goes to infinity), then, model (1) becomes to a high-dimensional VAR model, which involves completely new theories and is out of the scope of this paper.

model (1) by allowing τ in $q_{\tau,t,i}$ to vary across different equations, only with mild changes on asymptotic theory in this paper. Thus, in order to meet our empirical motivation, all of $\tau's$ in model (1) are the same throughout this article.

Importantly, in the case of estimating dynamic financial network in empirical studies, by following White et al. (2015), we consider only the tail dependence between current state and the state of one-period lagged, and take $\mathbb{Y}_t$ to be $\mathbb{Y}_t = (|Y_{1t}|, \ldots, |Y_{\kappa t}|)^\top$ with $|\cdot|$ representing absolute value.[2] Furthermore, the smoothing variable Z_{it} varies only across different time periods but keeps constant over individual units. Therefore, in this paper, for easy exposition, our focus is on the simple case that $q = p = 1$, $\kappa = \kappa_1$, $\mathbb{Y}_t = (|Y_{1t}|, \ldots, |Y_{\kappa t}|)^\top$, and $Z_{it} = Z_t$ for all $1 \le i \le \kappa$. Then, model (1) can be rewritten as

$$q_{\tau,t,i} = \boldsymbol{g}_{i,\tau}^\top(Z_t)\boldsymbol{X}_t, \tag{2}$$

where $\boldsymbol{g}_{i,\tau}(\cdot) = (\gamma_{i0,\tau}(\cdot), \gamma_{i1,\tau}(\cdot), \ldots, \gamma_{i\kappa,\tau}(\cdot), \beta_{i1,\tau}(\cdot), \ldots, \beta_{i\kappa,\tau}(\cdot))^\top$ is a $(2\kappa+1)\times 1$ vector of functional coefficients and $\boldsymbol{X}_t = (1, q_{\tau,t-1,1}, \ldots, q_{\tau,t-1,\kappa}, |Y_{1(t-1)}|, \ldots, |Y_{\kappa(t-1)}|)^\top$. As discussed in Engle and Manganelli (2004), taking absolute value on lagged Y_{it} allows for illustrating the case when both very positive and negative returns increase the VaR in the future.

It is worthwhile to note that if $q_{\tau,t,i}$ in model (2) is defined as VaR of return Y_{it}, then, $\{\gamma_{ij,\tau}(Z_t)\}_{i=1,j=1}^{\kappa}$ in model (2) becomes to the dependence of VaR for one portfolio at time t on that of another at time $t-1$. With these functional coefficients $\{\gamma_{ij,\tau}(Z_t)\}_{i=1,j=1}^{\kappa}$, define the following $\kappa \times \kappa$ matrix

$$\boldsymbol{\Gamma}_{1,\tau}(Z_t) = (\gamma_{ij,\tau}(Z_t))_{\kappa\times\kappa}, \tag{3}$$

which can be used to characterize the cross-sectional dependence among conditional quantiles $\{q_{\tau,t,j}\}_{i=1}^{\kappa}$ at time t.[3] Then, (2) can be expressed as a matrix form, which is a FCVAR model for

[2] Of course, $\mathbb{Y}_t$ can be any covariates.

[3] It is well known that it is not easy to characterize the cross-sectional dependence among individual conditional quantiles $\{q_{\tau,t,j}\}_{j=1}^{\kappa}$ at time t for given τ. Therefore, $\boldsymbol{\Gamma}_{1,\tau}(Z_t)$ defined in (3) can be used as one of measures to capture the dependence.

$\boldsymbol{q}_{\tau,t}$ with exogenous variables,

$$\boldsymbol{q}_{\tau,t} = \boldsymbol{\gamma}_{0,\tau}(Z_t) + \boldsymbol{\Gamma}_{1,\tau}(Z_t)\, \boldsymbol{q}_{\tau,t-1} + \boldsymbol{\Gamma}_{\beta,1,\tau}(Z_t)\, \mathbb{Y}_{t-1}, \tag{4}$$

where $\boldsymbol{\gamma}_{0,\tau}(Z_t)$ and $\boldsymbol{\Gamma}_{\beta,1,\tau}(Z_t)$ are defined obviously. Therefore, $\boldsymbol{\Gamma}_{1,\tau}(Z_t)$ in (4) can serve as a dynamic network system changing with both τ and some information variable Z_t, and it is in a nonparametric nature, so that it is a nonparametric dynamic network. Note that the general setting in the dynamic network system (4) covers some famous network models for characterizing financial risk system, including the one formed by VAR for VaR model in White et al. (2015), which assumes the tail dependence $\{\gamma_{ij,\tau}(Z_t)\}_{i=1,j=1}^{\kappa}$ to be constant and the static financial network in Adrian and Brunnermeier (2016) and Härdle et al. (2016) as special cases.

To investigate the large sample behavior of the proposed estimator (see Theorem 3 later), it is assumed throughout this article that the process $\{(Y_{it}, Z_t)\}$ in model (2) is strictly stationary and α-mixing (strongly mixing).[4] Indeed, in Appendix D, we show that under some regularity conditions, the joint process $\{(Y_{it}, x_{it}, Z_t, q_{\tau,t,i})\}$ generated by model (1) is strictly stationary and α-mixing. By letting $p = q = 1$, these sufficient conditions for the process $\{(Y_{it}, x_{it}, Z_t, q_{\tau,t,i})\}$ in model (1) can be directly applied for verifying the aforementioned probabilistic properties of $\{(Y_{it}, Z_t)\}$ in model (2). The derivation of these two properties in this paper is of independent interest, since our main interests in this article are to derive the asymptotic theory for model (2) and estimate a new class of dynamic financial networks. Therefore, we provide some sufficient conditions that imply these important probabilistic properties and corresponding rigorously theoretical justifications in Appendix D.

Before introducing our estimation procedure in Section 2.2, it should be noted that model (2) can be transformed into a functional-coefficient quantile function, which can serve as an approximation for latent $q_{\tau,t,i}$ in (2) in the preliminary step of our estimation. To see this, we first focus our attention on (4), which is a vector form of (2). Under Assumptions A1 and A7

[4]If some regressors or some of $\{q_{\tau,t,i}\}$ are nonstationary, say, unit root or local to unit root, the model in (1) and its asymptotic theory become very complicated and it needs definitely a further investigation, warranted as a future research; see, for example, the paper by Cheng, Han and Inoue (2022) for details.

given in Appendix A, by recursively substituting in (4), one has

$$\boldsymbol{q}_{\tau,t} = \boldsymbol{\Gamma}_{0,\tau,t} + \sum_{l=1}^{\infty} \boldsymbol{\Gamma}_{l,\tau}(\boldsymbol{Z}_{t,l})\mathbb{Y}_{t-l}, \tag{5}$$

where $\boldsymbol{\Gamma}_{0,\tau,t} = \sum_{l=1}^{\infty} \boldsymbol{\Gamma}_{l0,\tau}(\boldsymbol{Z}_{t,l})$ and $\boldsymbol{Z}_{t,l} = (Z_t, Z_{t-1}, \ldots, Z_{t-l+1})^\top$ is a vector of lagged smoothing variables. Here, $\boldsymbol{\Gamma}_{10,\tau}(\boldsymbol{Z}_{t,1}) = \boldsymbol{\gamma}_{0,\tau}(Z_t)$ and $\boldsymbol{\Gamma}_{l0,\tau}(\boldsymbol{Z}_{t,l}) = \prod_{\ell=0}^{l-2} \boldsymbol{\Gamma}_{1,\tau}(Z_{t-\ell})\boldsymbol{\gamma}_{0,\tau}(Z_{t-l+1})$ for $l \geq 2$; $\boldsymbol{\Gamma}_{1,\tau}(\boldsymbol{Z}_{t,1}) = \boldsymbol{\Gamma}_{\beta,1,\tau}(Z_t)$ and $\boldsymbol{\Gamma}_{l,\tau}(\boldsymbol{Z}_{t,l}) = \prod_{\ell=0}^{l-2} \boldsymbol{\Gamma}_{1,\tau}(Z_{t-\ell})\boldsymbol{\Gamma}_{\beta,1,\tau}(Z_{t-l+1})$ for $l \geq 2$. Now, let $\alpha_{li0,\tau}(\boldsymbol{Z}_{t,l}) = \boldsymbol{\gamma}_{i,1,\tau}^\top(Z_t)\big(\prod_{\ell=1}^{l-2} \boldsymbol{\Gamma}_{1,\tau}(Z_{t-\ell})\boldsymbol{\gamma}_{0,\tau}(Z_{t-l+1})\big)$ for $l \geq 3$, $\alpha_{2i0,\tau}(\boldsymbol{Z}_{t,2}) = \boldsymbol{\gamma}_{i,1,\tau}^\top(Z_t)\boldsymbol{\gamma}_{0,\tau}(Z_{t-1})$, and $\alpha_{1i0,\tau}(\boldsymbol{Z}_{t,1}) = \gamma_{i0,\tau}(Z_t)$; let $\alpha_{lij,\tau}(\boldsymbol{Z}_{t,l}) = \boldsymbol{\gamma}_{i,1,\tau}^\top(Z_t)\big(\prod_{\ell=1}^{l-2} \boldsymbol{\Gamma}_{1,\tau}(Z_{t-\ell})\boldsymbol{\Gamma}_{\beta,1j,\tau}(Z_{t-l+1})\big)$ for $l \geq 3$, $\alpha_{2ij,\tau}(\boldsymbol{Z}_{t,2}) = \boldsymbol{\gamma}_{i,1,\tau}^\top(Z_t)\boldsymbol{\Gamma}_{\beta,1j,\tau}(Z_{t-1})$ and $\alpha_{1ij,\tau}(\boldsymbol{Z}_{t,1}) = \beta_{ij,\tau}(Z_t)$. Here, $\gamma_{i0,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$ are defined in (2), $\boldsymbol{\gamma}_{i,1,\tau}^\top(\cdot)$ is the ith row of $\boldsymbol{\Gamma}_{1,\tau}(\cdot)$, $\boldsymbol{\Gamma}_{\beta,1j,\tau}(\cdot)$ is the jth column of $\boldsymbol{\Gamma}_{\beta,1,\tau}(\cdot)$. Then, $\alpha_{li0,\tau}(\cdot)$ and $\boldsymbol{\alpha}_{li,\tau}^\top(\cdot) = (\alpha_{li1,\tau}(\cdot), \ldots, \alpha_{li\kappa,\tau}(\cdot))$ are the ith row of $\boldsymbol{\Gamma}_{l0,\tau}(\cdot)$ and $\boldsymbol{\Gamma}_{l,\tau}(\cdot)$, respectively. Thus, the ith row of (5) is written as

$$q_{\tau,t,i} = \sum_{l=1}^{\infty} \alpha_{li0,\tau}(\boldsymbol{Z}_{t,l}) + \sum_{l=1}^{\infty} \boldsymbol{\alpha}_{li,\tau}^\top(\boldsymbol{Z}_{t,l})\mathbb{Y}_{t-l}. \tag{6}$$

As discussed in Section 2.2, in the first stage of our estimation procedure, a truncated version of (6) is used to approximate latent $q_{\tau,t,i}$ and is modeled globally by a tensor product of normalized B-spline basis functions. To handle the issue of high-dimensionality, a two-stage LASSO approach is introduced for dimension reduction and variable selection.

Note that equation (6) corresponds to a functional-coefficient quantile autoregressive (AR)(∞) process, which can be used as an underlying data generating process for Y_{it} in model (1). Indeed, denote U_{it} ($1 \leq i \leq \kappa$, $1 \leq t \leq n$) as independent and identically distributed (iid) standard uniform random variables on the set of $[0, 1]$. Then, following the same argument as in Koenker and Xiao (2006), by assuming that the right side of (6) is monotonically increasing in τ, model (6) corresponds to a functional-coefficient quantile AR(∞) process as follows

$$Y_{it} = \sum_{l=1}^{\infty} \alpha_{li0}(U_{it}, \boldsymbol{Z}_{t,l}) + \sum_{l=1}^{\infty} \boldsymbol{\alpha}_{li}^\top(U_{it}, \boldsymbol{Z}_{t,l})\mathbb{Y}_{t-l} \tag{7}$$

where $\alpha_{li0}(\cdot, \cdot)$ is a scalar and measurable function of U_{it} and $\boldsymbol{Z}_{t,l}$ (from $\mathbb{R} \times \mathbb{R}^l$ to $\mathbb{R}$) and $\boldsymbol{\alpha}_{li}(\cdot, \cdot)$

is a vector of measurable functions from $\mathbb{R} \times \mathbb{R}^l$ to $\mathbb{R}$. Since our objectives in this paper are estimation of model (2) and modeling time-varying tail dependence among dynamic quantiles, we only present identification assumptions for (2) and (4) in Appendix A, instead of focusing on model (7) with more complex coefficients. Therefore, further investigating the properties of (7), omitted in this paper, is left as a future study. Next, we make some remarks on our model in (1).

Remark 1. *(Special Cases) The proposed FCVAR-DQ model* (1) *is related to the papers by Engle and Manganelli (2004) and Xiao and Koenker (2009), which discuss the relation between modeling dynamic structures of conditional quantiles and conditional volatility of returns. Indeed, if* $\kappa = \kappa_1$ *in* (1)*,* Y_{it} *in* (1) *takes a simple form as* $Y_{it} = \sigma_{it}\, e_{it}$*, where* σ_{it}^2 *is the conditional variance of* Y_{it} *and* e_{it} *is an iid sequence of random variables with mean zero and unit variance, then,* $q_{\tau,t,i} = \sigma_{it} F_e^{-1}(\tau)$*, where* $F_e(\cdot)$ *is the distribution function of* e_{it}*. Furthermore, if* $Y_{it} = \sigma_{it}\, e_{it}$ *is generated from a functional coefficient multivariate GARCH* (p, q)*-type process for* κ $(\kappa \geq 1)$ *returns extended from the setting in Taylor (1986) as follows*

$$\sigma_{it} = \gamma_{i0}(Z_t) + \sum_{s=1}^{q} \boldsymbol{\gamma}_{i,s}^{\top}(Z_t)\boldsymbol{\Sigma}_{t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}_{i,l}^{\top}(Z_t)\mathbb{Y}_{t-l},$$

where $\boldsymbol{\Sigma}_t = (\sigma_{it}, \ldots, \sigma_{\kappa t})^{\top}$ *and* $\mathbb{Y}_t = (|Y_{1t}|, \ldots, |Y_{\kappa t}|)^{\top}$*, then, model* (1) *reduces to following dynamic quantile model:*

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_t) + \sum_{s=1}^{q} \boldsymbol{\gamma}_{i,s}^{\top}(Z_t)\boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}_{i,l,\tau}^{\top}(Z_t)\mathbb{Y}_{t-l}, \tag{8}$$

where $\gamma_{i0,\tau}(\cdot) = \gamma_{i0}(\cdot)F_e^{-1}(\tau)$*,* $\boldsymbol{\gamma}_{i,s}(\cdot) = (\gamma_{si1}(\cdot), \ldots, \gamma_{si\kappa}(\cdot))^{\top}$ *and* $\boldsymbol{\beta}_{i,l,\tau}(\cdot) = (\beta_{li1,\tau}(\cdot), \ldots, \beta_{li\kappa,\tau}(\cdot))^{\top}$ *with* $\beta_{lij,\tau}(\cdot) = \beta_{lij}(\cdot)F_e^{-1}(\tau)$*. Note that if* $\gamma' s$ *and* $\beta' s$ *in* (8) *are constant, model* (8) *reduces to those in Engle and Manganelli (2004) and Xiao and Koenker (2009), respectively. For details, the reader is referred to the aforementioned papers. Finally, note that if* $\boldsymbol{q}_{\tau,t}$ *would be observable and all coefficients are threshold functions, model* (1) *covers the model in Tsay (1998).*

Remark 2. *(Monotonicity). The issue of monotonicity is frequently discussed for the quantile autoregression model. A specific case for the monotonicity of* (1) *to hold is that* $\{\boldsymbol{\gamma}_{i,s,\tau}(Z_t)\}_{i=1,s=1}^{\kappa,q}$

and $\{\boldsymbol{\beta}_{i,l,\tau}(Z_t)\}_{i=1,l=1}^{\kappa,p}$ *are all monotone increasing functions with respect to* τ*, and* $\mathbb{Y}_t$ *is a positive random vector. In other cases, the assumption of monotonicity can be satisfied by conducting certain data transformation techniques; see Koenker and Xiao (2006) and Fan and Fan (2006) for detailed discussions.*

Remark 3. *Let* $\kappa = \kappa_1$ *and* $Z_{it} = Z_t$*. Then, the matrix form of model* (1) *is given by* $\boldsymbol{q}_{\tau,t} = \boldsymbol{\gamma}_{0,\tau}(Z_t) + \sum_{s=1}^{q} \boldsymbol{\Gamma}_{s,\tau}(Z_t)\, \boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\Gamma}_{\beta,l,\tau}(Z_t)\, \mathbb{Y}_{t-l}$*, where* $\boldsymbol{\Gamma}_{s,\tau}(Z_t)$ *and* $\boldsymbol{\Gamma}_{\beta,l,\tau}(Z_t)$ *are defined obviously. Note that for* $1 \le t \le n$, $1 \le s \le q$ *and* $1 \le l \le p$*, if all eigenvalues of* $\boldsymbol{\Gamma}_{s,\tau}(Z_t)$ *and* $\boldsymbol{\Gamma}_{\beta,l,\tau}(Z_t)$ *have modulus less than* 1*, then, the matrix form of model* (1) *corresponds to a recursive form similar to equation* (5)*. Thus, under this additional assumption, the proposed estimation procedure and theories for model* (2) *are extendable to model* (1)*, with more complicated notations being involved.*

2.2 Three-stage Estimation Procedure

In this paper, we only focus on estimating functional coefficients $\boldsymbol{g}_{i,\tau}(\cdot)$ in (2) for simplicity, rather than jointly forecasting $q_{\tau,t,i}$ or doing impulse response analysis. So, it is sufficient to estimate $\boldsymbol{g}_{i,\tau}(\cdot)$ in an equation-by-equation way for different i. Thus, by abuse of notation, i is dropped in what follows.

By Assumption A1 in Appendix A, $\alpha_{l0,\tau}(\cdot)$ and each entry of $\boldsymbol{\alpha}_{l,\tau}(\cdot) = (\alpha_{l1,\tau}(\cdot), \ldots, \alpha_{l\kappa,\tau}(\cdot))^\top$ defined in (6) decrease at a geometric rate; that is, there exist positive constants $\rho < 1$ and c, such that $\max_{1\le t\le n} |\alpha_{lj,\tau}(\boldsymbol{Z}_{t,l})| \le c\rho^l$ for $j = 0, \ldots, \kappa$. Since $\alpha_{lj,\tau}(\cdot)$ decreases geometrically, by choosing truncation parameter $m_n = m(n) = m$, (6) becomes to

$$q_{\tau,t} = \sum_{l=1}^{m_n} \alpha_{l0,\tau}(\boldsymbol{Z}_{t,l}) + \sum_{l=1}^{m_n} \boldsymbol{\alpha}_{l,\tau}^\top(\boldsymbol{Z}_{t,l})\mathbb{Y}_{t-l} \equiv \sum_{l=1}^{m_n} \boldsymbol{\alpha}_{1,l,\tau}^\top(\boldsymbol{Z}_{t,l})\boldsymbol{W}_{t-l} \equiv \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t \equiv q_\tau(Z_t, \boldsymbol{W}_t), \tag{9}$$

where $\boldsymbol{\alpha}_{1,l,\tau}(\cdot) = (\alpha_{l0,\tau}(\cdot), \boldsymbol{\alpha}_{l,\tau}^\top(\cdot))^\top = (\alpha_{l0,\tau}(\cdot), \alpha_{l1,\tau}(\cdot), \ldots, \alpha_{l\kappa,\tau}(\cdot))^\top$, $\boldsymbol{W}_t = (1, \mathbb{Y}_t^\top)^\top$, $\boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t) = (\boldsymbol{\alpha}_{1,1,\tau}^\top(\boldsymbol{Z}_{t,1}), \ldots, \boldsymbol{\alpha}_{1,m,\tau}^\top(\boldsymbol{Z}_{t,m}))^\top$ and $\mathbb{W}_t = (\boldsymbol{W}_{t-1}^\top, \ldots, \boldsymbol{W}_{t-m}^\top)^\top$.

Now, denote $\tilde{b}^{K_n}(z) = (\tilde{b}_1(z), \ldots, \tilde{b}_{K_n}(z))^\top$ as a vector of normalized B-spline basis functions on a compact interval $[a, b]$, where $a < b$ are finite numbers. Here, $K_n = K = k_n + \hbar + 1$, with $\hbar$ being the degree of polynomial and k_n being the number of quasi-uniform internal knots; see Schumaker (1981) for details on the construction of normalized B-spline bases. Then, for $1 \leq l \leq m$, define a $K^l \times 1$ vector of tensor-product B-spline bases as follows. Define $\tilde{B}_{lt} = \left(\bigotimes_{\iota=0}^{l-2} \tilde{b}^K(Z_{t-\iota})\right) \otimes \tilde{b}^K(Z_{t-l+1})$ for $l \geq 2$ and $\tilde{B}_{1t} = \tilde{b}^K(Z_t)$, let $B_{lt} = \left(\frac{K^l}{n}\sum_{t=1}^n \tilde{B}_{lt}\tilde{B}_{lt}^\top\right)^{-1/2} K^{l/2}\tilde{B}_{lt}$. Similar transformation to the basis functions is also conducted in Tang, Song, Wang and Zhu (2013), which guarantees that the last part of Assumption A9 in Appendix A is satisfied. Then, there exist vectors $\boldsymbol{c}_{l0,\tau} \in \mathbb{R}^{K^l}$ and $\boldsymbol{c}_{lj,\tau} \in \mathbb{R}^{K^l}$ such that $\alpha_{l0,\tau}(\boldsymbol{Z}_{t,l})$ and $\alpha_{lj,\tau}(\boldsymbol{Z}_{t,l})$ can be approximated by $\alpha_{l0,\tau}(\boldsymbol{Z}_{t,l}) \approx B_{lt}^\top \boldsymbol{c}_{l0,\tau}$ and $\alpha_{lj,\tau}(\boldsymbol{Z}_{t,l}) \approx B_{lt}^\top \boldsymbol{c}_{lj,\tau}$, respectively. Furthermore, denote $\boldsymbol{P}_{lt} = \boldsymbol{W}_{t-l} \otimes B_{lt}$ as a $(1+\kappa)K^l \times 1$ vector and $\boldsymbol{\Pi}_t = (\boldsymbol{P}_{1t}^\top, \ldots, \boldsymbol{P}_{mt}^\top)^\top$. Finally, let $M_n = (1 + \kappa)\sum_{l=1}^m K^l = (1+\kappa)K\left(\frac{1-K^m}{1-K}\right)$ be the dimension of $\boldsymbol{\Pi}_t$ and define $\boldsymbol{c}_\tau = (\boldsymbol{c}_{1,\tau}^\top, \ldots, \boldsymbol{c}_{m,\tau}^\top)^\top \in \mathbb{R}^{M_n}$, where $\boldsymbol{c}_{l,\tau} = (\boldsymbol{c}_{l0,\tau}^\top, \ldots, \boldsymbol{c}_{l\kappa,\tau}^\top)^\top = (c_{l1}, c_{l2}, \ldots, c_{l((1+\kappa)K^l)})^\top$. Thus, $q_{\tau,t}$ can be approximated by $q_{\tau,t} \approx \boldsymbol{\Pi}_t^\top \boldsymbol{c}_\tau$.

Remark 4. *(Tensor-product B-spline bases) The tensor-product B-spline approximation is a standard approach to generate linear sieves of multivariate functions from linear sieves of univariate functions. Different from the paper by He and Shi (1996) using bivariate tensor-product B-splines to approximate unknown functions under M-type regression setting, in this article, the length of the tensor-product B_{lt} is K^l and can diverge to infinity very fast. To overcome this ultra-high-dimensional problem, in the first stage of our procedure, we apply LASSO penalty to the L_1 norm of each coefficient group for dimension reduction.*

First-step: In the first stage, we apply LASSO penalty to the L_1 norms of $\boldsymbol{c}_{l,\tau}$, $l = 1, \ldots, m$. Hence, $\tilde{\boldsymbol{c}} = \underset{\boldsymbol{c}}{\operatorname{argmin}} Q_0(\boldsymbol{c})$, where $Q_0(\boldsymbol{c})$ is given by

$$Q_0(\boldsymbol{c}) = \frac{1}{n}\sum_{t=m+1}^n \rho_\tau\left\{Y_t - \boldsymbol{\Pi}_t^\top \boldsymbol{c}\right\} + \frac{1}{n}\sum_{l=1}^m p_{\lambda_{n,0}}(\|\boldsymbol{c}_l\|_1), \tag{10}$$

$\rho_\tau(y) = y[\tau - I(y < 0)]$ is called the "check" (loss) function, $I(A)$ is the indicator function of any set A and $p_{\lambda_{n,0}}(\cdot)$ is a penalty function with a tuning parameter $\lambda_{n,0}$ for the first step. It is worthwhile to note that different types of penalty function $p_{\lambda_{n,0}}(\cdot)$ are allowed to use in (10), including but not limited to the LASSO, the smoothly clipped absolute deviation (SCAD) of Fan and Li (2001) and the minimax concave penalty (MCP) of Zhang (2010). However, we do not attempt to derive the theory for a general setup of penalty function, since the regularized regression is only applied as a preliminary step of our procedure. Therefore, in this article, we choose to use LASSO by setting $p_{\lambda_{n,0}}(\cdot) = \lambda_{n,0} \cdot (\cdot)$ due to its theoretical and computational simplicity. Note that minimizing (10) with $p_{\lambda_{n,0}}(\cdot) = \lambda_{n,0} \cdot (\cdot)$ can be written as a linear programming problem as follows

$$\min_{(\boldsymbol{\xi}^+,\boldsymbol{\xi}^-,\boldsymbol{c}^+,\boldsymbol{c}^-)\in\mathbb{R}_+^{2(n-m)+2M_n}} n^{-1}\sum_{t=m+1}^{n}\left\{\tau\xi_t^+ + (1-\tau)\xi_t^-\right\} + \frac{\lambda_{n,0}}{n}\sum_{l=1}^{m}\sum_{\mathfrak{u}=1}^{(1+\kappa)K^l}(c_{l\mathfrak{u}}^+ + c_{l\mathfrak{u}}^-)$$

subject to $\xi_t^+ - \xi_t^- = Y_t - \boldsymbol{\Pi}_t^\top(\boldsymbol{c}^+ - \boldsymbol{c}^-)$, $t = m+1, \ldots, n$.

Second-step: Based on the initial estimator $\tilde{\boldsymbol{c}}_\tau$ obtained by minimizing (10), we apply adaptive LASSO penalty to exclude the remained irrelevant covariates survived after the first stage of dimension reduction. Define $\tilde{T}_n = \{l : \|\tilde{\boldsymbol{c}}_{l,\tau}\|_2 > 0, l = 1, \ldots, m\}$, let $\boldsymbol{c}_l = \boldsymbol{0}_{(1+\kappa)K^l}$ for $l \notin \tilde{T}_n$. Then, we propose to minimizing

$$Q_1(\boldsymbol{c}) = \frac{1}{n}\sum_{t=m+1}^{n}\rho_\tau\left\{Y_t - \boldsymbol{\Pi}_t^\top\boldsymbol{c}\right\} + \frac{\lambda_{n,1}}{n}\sum_{l\in\tilde{T}_n}\tilde{\omega}_l\|\boldsymbol{c}_l\|_1 \tag{11}$$

with respect to the unknown components of $\boldsymbol{c}$, where $\lambda_{n,1}$ is the tuning parameter for the second step and $\tilde{\omega}_l = \|\tilde{\boldsymbol{c}}_{l,\tau}\|_2^{-1}$ for $l \in \tilde{T}_n$. After yielding $\hat{\boldsymbol{c}}_\tau$ at τ by minimizing (11), $q_{\tau,t}$ can be estimated by $\hat{q}_{\tau,t} = \boldsymbol{\Pi}_t^\top\hat{\boldsymbol{c}}_\tau$.

Third-step: With $\hat{q}_{\tau,t}$ and $\hat{\boldsymbol{X}}_t = (1, \hat{q}_{\tau,t-1,1}, \ldots, \hat{q}_{\tau,t-1,\kappa}, |Y_{1(t-1)}|, \ldots, |Y_{\kappa(t-1)}|)^\top$ in place, $\boldsymbol{g}_\tau(\cdot)$ in (2) is estimated by a local linear estimation method; see Cai and Xu (2008) for details. In particular, under smoothness condition of coefficient functions $\boldsymbol{g}_\tau(\cdot)$ presented in Assumption A4 in Appendix A, for any given grid point $z_0 \in [a, b]$, when Z_t is in a neighborhood of z_0, $\boldsymbol{g}_\tau(Z_t)$ can be approximated by a polynomial function as $\boldsymbol{g}_\tau(Z_t) \approx \sum_{r=0}^{\varsigma}\boldsymbol{g}_\tau^{(r)}(z_0)(Z_t - z_0)^r/r!$, where $\approx$ denotes the approximation by ignoring the higher orders and $\boldsymbol{g}_\tau^{(r)}(\cdot)$ is the rth derivative of

$\boldsymbol{g}_\tau(\cdot)$. Hence, minimize the following locally weighted loss function $Q_2(\Theta)$ at any given grid point $z_0 \in [a,b]$ to obtain the local linear estimate $\hat{\Theta}$, where

$$Q_2(\Theta) = \sum_{t=1}^{n} \rho_\tau \left\{ Y_t - \sum_{r=0}^{1} \hat{\boldsymbol{X}}_t^\top \Theta_{r,\tau}(Z_t - z_0)^r \right\} K_h(Z_t - z_0). \tag{12}$$

Here, $\Theta_{r,\tau} = \boldsymbol{g}_\tau^{(r)}(\cdot)/r!$, $K(\cdot)$ is a kernel function, $K_h(u) = K(u/h)/h$, and $h = h(n)$ is a sequence of positive numbers tending to zero and controls the amount of smoothing used in estimation.

2.3 Large Sample Theory

We study the asymptotic distribution of the proposed nonparametric estimator in this section. In the first and second stages of our estimation procedure, the coefficient functionals are approximated by a tensor-product B-spline bases. Recently, Tang et al. (2013) considered B-spline approximation and variable selection in ultra-high-dimensional quantile varying coefficient models, with i.i.d. assumption being imposed on observations. To prove Theorems 1 and 2 in this paper, we follow a similar proof strategy as in Tang et al. (2013), but substantially extend their results to allow strictly stationary and α-mixing data setting.

Suppose that each Z_t takes values in a compact interval $[a,b]$. For $l = 1, \ldots, m$, let $\boldsymbol{\mathcal{Z}}_l = [a,b]^l$ be the Cartesian product of $[a,b]$. Given a l-tuple nonnegative integers, $w = (w_0, \ldots, w_{l-1})$, set $|w| = w_0 + \cdots + w_{l-1}$ and let D^w denote the differential operator defined by $D^w = \frac{\partial^{|w|}}{\partial Z_t^{w_0} \ldots \partial Z_{t-l+1}^{w_{l-1}}}$. Let $\upsilon \in (0,1]$ be such that $d \equiv |w| + \upsilon \geq 2$ and $\boldsymbol{\mathcal{H}}_l$ be the class of functions α_τ on $\boldsymbol{\mathcal{Z}}_l$ whose $D^w \alpha_\tau$ exists and satisfies a Lipschitz condition of order υ:

$$|D^w \alpha_\tau(\boldsymbol{z}_l) - D^w \alpha_\tau(\tilde{\boldsymbol{z}}_l)| \leq C \|\boldsymbol{z}_l - \tilde{\boldsymbol{z}}_l\|^\upsilon, \text{ for } \boldsymbol{z}_l, \tilde{\boldsymbol{z}}_l \in \boldsymbol{\mathcal{Z}}_l \text{ and some constant } C > 0.$$

In (9), for $0 \leq j \leq \kappa$, we assume that $\alpha_{lj,\tau}(\cdot) \neq 0$ for $1 \leq l \leq r_n$; $\alpha_{lj,\tau}(\cdot) = 0$ for $r_n + 1 \leq l \leq m$. Let $T_{l,n} = \{\mathfrak{u} : |c_{l\mathfrak{u},\tau}| > 0, \mathfrak{u} = 1, \ldots, (1+\kappa)K^l\}$ be a sparse support of $\boldsymbol{c}_{l,\tau}$, having $\text{card}(T_{l,n}) = R_{l,n}$ nonzero components and $R_n = \sum_{l=1}^{r_n} R_{l,n}$. In addition, recall that $M_n = (1+\kappa)\sum_{l=1}^{m} K^l$ and denote $A_{M_n}^{\mathfrak{r}} = \{\boldsymbol{\delta} \in \mathbb{R}^{M_n} : \|\boldsymbol{\delta}\|_2 = 1, \|\boldsymbol{\delta}\|_0 \leq \mathfrak{r}\}$ as the $\mathfrak{r}$-sparse unit sphere in $\mathbb{R}^{M_n}$. Define the $\mathfrak{r}$-sparse maximal eigenvalue and $\mathfrak{r}$-sparse minimal eigenvalue of

$E[\mathbf{\Pi}_t\mathbf{\Pi}_t^\top]$ as $\tilde{\varphi}(\mathfrak{r}) = \sup_{\boldsymbol{\delta}\in A^{\mathfrak{r}}_{M_n}} E[(\boldsymbol{\delta}^\top\mathbf{\Pi}_t)^2]$ and $\varrho(\mathfrak{r}) = \inf_{\boldsymbol{\delta}\in A^{\mathfrak{r}}_{M_n}} E[(\boldsymbol{\delta}^\top\mathbf{\Pi}_t)^2]$, respectively. Similarly, define $\varphi(\mathfrak{r}) = \sup_{\boldsymbol{\delta}\in A^{\mathfrak{r}}_{M_n}} \boldsymbol{\delta}^\top\left(n^{-1}\sum_{t=m+1}^n \mathbf{\Pi}_t\mathbf{\Pi}_t^\top\right)\boldsymbol{\delta}$, $\vartheta(\mathfrak{r}) = \inf_{\boldsymbol{\delta}\in A^{\mathfrak{r}}_{M_n}} E[|\boldsymbol{\delta}^\top\mathbf{\Pi}_t|^2]/E[|\boldsymbol{\delta}^\top\mathbf{\Pi}_t|^3]$. Let $\mathfrak{q} = \mathfrak{q}(\breve{\mathfrak{m}})$ be a sequence of positive numbers that characterize the strength of identification in the population, which is defined as

$$\mathfrak{q}(\breve{\mathfrak{m}}) = (\varrho(\breve{\mathfrak{m}})\underline{f}/4)\min\{1, (\underline{f}/2\bar{f}')\vartheta(\breve{\mathfrak{m}})\}.$$

Here, $\breve{\mathfrak{m}} = \|\tilde{\boldsymbol{c}}_\tau\|_0$, $\underline{f}$ and $\bar{f}'$ are defined in Assumption A6. Let $\mu = \mu(\breve{\mathfrak{m}}) \geq \mathfrak{q}(\breve{\mathfrak{m}})$ be a sequence of positive constants, defined as $\mu(\breve{\mathfrak{m}}) = \sqrt{\varphi(\breve{\mathfrak{m}})}(\bar{f}\sqrt{\varphi(\breve{\mathfrak{m}})} \vee 1)$.

Finally, let $\mathcal{G}_t = (\ldots, U_{t-1}, U_t) \subset \mathcal{F}_t$ be an information set up to time t with $\{U_t\}$ being sequence of i.i.d. standard uniform random variables and $\mathcal{F}_t$ being defined in Section 2.1. Let $V_{it} \equiv \{(Y_{it}, Z_t)\}$ be strictly stationary process admitting the following representation forms: $V_{it} = \mathcal{L}(\mathcal{G}_t)$, where $\mathcal{L}(\cdot)$ is a measurable function. Now, let U_0 be replaced by an i.i.d. copy of U_0^* and $V_{it}^* = \mathcal{L}(\ldots, U_0^*, \ldots, U_{t-1}, U_t)$. For $\mathfrak{p} \geq 1$, define $\|V_{i\cdot}\|_{\mathfrak{p},0} \equiv \sum_{t=0}^\infty (E|V_{it} - V_{it}^*|^{\mathfrak{p}})^{1/\mathfrak{p}}$, which measures the cumulative effect of U_0 on $\{V_{it}\}_{t\geq 0}$. Based on the dependence adjusted norm $\|\cdot\|_{\mathfrak{p},0}$, we further define $\|V_{i\cdot}\|_{\psi_\nu,0} = \sup_{\mathfrak{p}\geq 2} \mathfrak{p}^{-\nu}\|V_{i\cdot}\|_{\mathfrak{p},0}$, for some $\nu \geq 0$. Let $\psi_\tau(V_{it}, \boldsymbol{c}) \equiv \{I(Y_t \leq \mathbf{\Pi}_t^\top\boldsymbol{c}) - \tau\}\mathbf{\Pi}_t$ be the score function of the tth observation. Similarly, define $\psi^0_{\tau,t} \equiv \psi_\tau(V_{it}, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t)) \equiv \{I(Y_t \leq \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t) - \tau\}\mathbf{\Pi}_t$.

Note that since Z_t is exogenous, the data generating process (7) can be written in the form of $\mathcal{L}(\ldots, U_{i(t-1)}, U_{it})$, with $\mathcal{L}(\cdot)$ being a measurable function and $\{U_{it}\}$ being sequence of i.i.d. standard uniform random variables. Therefore, under a stronger moment condition related to dependence adjusted norm (see Assumption A10 in Appendix A), one can apply the exponential tail bounds proposed in Lemma B.4 of Chernozhukov, Härdle, Huang and Wang (2021) to obtain the error bound of $\tilde{\boldsymbol{c}}_\tau$ and corresponding oracle properties in Theorem 1. In Appendix A, some necessary conditions are provided, though they might not the weakest possible with some discussions on assumptions.

By Assumption A3 in Appendix A, there exists a nonzero spline coefficient vector $\bar{\boldsymbol{c}}_{l,\tau}$ such that $\alpha_{lj,\tau}(\boldsymbol{Z}_{t,l}) = B_{lt}^{\top}\bar{\boldsymbol{c}}_{lj,\tau} + d_{lj}(\boldsymbol{Z}_{t,l})$, $1 \le l \le m$, with $\sup_{l,j,t}|d_{lj}(\boldsymbol{Z}_{t,l})| = O(K^{-d})$. Let $\bar{\boldsymbol{c}}_{\tau} = (\bar{\boldsymbol{c}}_{1,\tau}^{\top}, \ldots, \bar{\boldsymbol{c}}_{r_n,\tau}^{\top}, 0_{M_n-R_n}^{\top})^{\top}$ and denote $\hat{\boldsymbol{c}}_{\tau} = (\hat{\boldsymbol{c}}_{1,\tau}^{\top}, \ldots, \hat{\boldsymbol{c}}_{m,\tau}^{\top})^{\top}$ as the minimizer of (11).

Theorem 1. *Suppose Assumptions A1–A13 given in Appendix A hold, let*

$$m_0 = M_n \wedge \left(\frac{n}{\log(n \vee M_n)}\frac{\mathfrak{q}^2}{\mu^2}\right) \quad \text{and} \quad \lambda_{n,0} = \aleph\sqrt{n\log(n \vee M_n)\varphi(m_0 + R_n)}(m_0\log(n \vee M_n))^{\nu}\frac{\mu}{\mathfrak{q}}$$

for some $0 \le \nu < 1/2$, where $\aleph \to \infty$ is a sequence of positive numbers, possibly data-dependent. Then, (a) $\tilde{\boldsymbol{\zeta}}_n = \|\tilde{\boldsymbol{c}}_{\tau} - \bar{\boldsymbol{c}}_{\tau}\|_2 \lesssim_p \frac{\lambda_{n,0}\sqrt{R_n}}{n\mathfrak{q}}$, and (b) $\|\tilde{\boldsymbol{c}}_{\tau}\|_0 \lesssim_p \frac{\mu^2}{\mathfrak{q}^2}R_n$.

In order to get a consistent estimator in the first stage, Theorem 1 suggests that M_n, the dimension of $\boldsymbol{\Pi}_t$, satisfies $R_n(\log(M_n))^{1+\nu}/n \to 0$, for some $0 \le \nu < 1/2$. Next, Theorem 2 presents the asymptotic properties of the second-stage estimator, including the oracle properties and the convergence rate of $\hat{q}_{\tau,t}$.

Theorem 2. *Let $\hat{\alpha}_{lj,\tau}(\boldsymbol{Z}_{t,l}) = B_{lt}^{\top}\hat{\boldsymbol{c}}_{lj,\tau}$ be the estimate of coefficient function $\alpha_{lj,\tau}(\boldsymbol{Z}_{t,l})$, for $1 \le l \le m$ and $0 \le j \le \kappa$, where $\hat{\boldsymbol{c}}_{lj,\tau}$ is the jth component of $\hat{\boldsymbol{c}}_{l,\tau} = (\hat{\boldsymbol{c}}_{l0,\tau}^{\top}, \ldots, \hat{\boldsymbol{c}}_{l\kappa,\tau}^{\top})^{\top}$. Suppose Assumptions A1–A13 given in Appendix A hold and $R_n^{1/2}\tilde{\boldsymbol{\zeta}}_n \to 0$. Then, (a) $\hat{\alpha}_{lj,\tau}(\boldsymbol{Z}_{t,l}) = 0$, for $l = r_n + 1, \ldots, m$, with probability approaching 1, (b) $\|\hat{\boldsymbol{c}}_{\tau} - \bar{\boldsymbol{c}}_{\tau}\|_2 = O_p((R_n/n)^{1/2})$, and (c) $\max_{m+1 \le t \le n}|\hat{q}_{\tau,t} - q_{\tau,t}| = O_p(R_n/n^{1/2})$.*

Before stating the asymptotic behavior of $\hat{\boldsymbol{g}}_{\tau}(z_0)$ in the following theorem, for notational simplicity, it needs to define some notations. Define $\Omega^*(z_0) \equiv E\left[\boldsymbol{X}_t\boldsymbol{X}_t^{\top}f_{Y|Z,\boldsymbol{X}}(q_{\tau}(z_0, \boldsymbol{X}_t))|Z_t = z_0\right]$ with $q_{\tau}(z_0, \boldsymbol{X}_t) = \boldsymbol{g}_{\tau}^{\top}(z_0)\boldsymbol{X}_t$ and $f_{Y|Z,\boldsymbol{X}}(\cdot) = f_{Y|Z,\boldsymbol{W}}(\cdot)$. In addition, let $\boldsymbol{\Pi}_{a,t} = (\boldsymbol{P}_{1t}^{\top}, \ldots, \boldsymbol{P}_{r_nt}^{\top})^{\top}$ be the submatrix consisting of the first r_n compositions of $\boldsymbol{\Pi}_t = (\boldsymbol{P}_{1t}^{\top}, \ldots, \boldsymbol{P}_{mt}^{\top})^{\top}$ corresponding to the active covariates. Define $\boldsymbol{\Phi}_a = E[f_{Y|Z,\boldsymbol{W}}(q_{\tau}(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t}\boldsymbol{\Pi}_{a,t}^{\top}]$, $\boldsymbol{D}(z_0) \equiv E[\boldsymbol{\Pi}_{a,t}\boldsymbol{\Pi}_{a,t}^{\top}|Z_t = z_0]$ and let $\Gamma(z_0) \equiv E\left\{f_{Y|Z,\boldsymbol{X}}(q_{\tau}(z_0, \boldsymbol{X}_t))\boldsymbol{X}_t\boldsymbol{g}_{\tau}^{\top}(z_0)\,\boldsymbol{\Upsilon}_{a,t}\middle|Z_t = z_0\right\}$ be a $(2\kappa + 1) \times R_n$ matrix, with $\boldsymbol{\Upsilon}_{a,t}^{\top} = (0_{1\times R_n}^{\top}, \boldsymbol{\Pi}_{a,t}, \ldots, \boldsymbol{\Pi}_{a,t}, 0_{\kappa\times R_n}^{\top})$. Finally, let $\boldsymbol{\Xi}(z_0) \equiv \tau(1-\tau)\nu_0[\Omega(z_0) - H_1(z_0) + H_2(z_0)]$, where $H_1(z_0) = E[\boldsymbol{X}_t\boldsymbol{\Pi}_{a,t}^{\top}|Z_t = z_0]\boldsymbol{\Phi}_a^{-1}\Gamma^{\top}(z_0) + \Gamma(z_0)\boldsymbol{\Phi}_a^{-1}E[\boldsymbol{\Pi}_{a,t}\boldsymbol{X}_t^{\top}|Z_t = z_0]$, $H_2(z_0) = \Gamma(z_0)\boldsymbol{\Phi}_a^{-1}\boldsymbol{D}(z_0)\boldsymbol{\Phi}_a^{-1}\Gamma^{\top}(z_0)$ and $\Omega(z_0) \equiv E[\boldsymbol{X}_t\boldsymbol{X}_t^{\top}|Z_t = z_0]$. Now, the asymptotic normality of $\hat{\boldsymbol{g}}_{\tau}(z_0)$ is presented in the following theorem with its detailed proof relegated to Appendix C.

Theorem 3. *Under Assumptions A1–A13 provided in Appendix A, we have*

$$\sqrt{nh}\left[\hat{\boldsymbol{g}}_\tau(z_0)-\boldsymbol{g}_\tau(z_0)-\frac{h^2\mu_2}{2}\boldsymbol{g}_\tau^{(2)}(z_0)+o_p(h^2)\right]\xrightarrow{d}\mathcal{N}(0,\Sigma_\tau(z_0)),$$

where $\Sigma_\tau(z_0)=(\Omega^*(z_0))^{-1}\boldsymbol{\Xi}(z_0)(\Omega^*(z_0))^{-1}/f_z(z_0)$.

Note that the detailed proofs of Theorem 1 and 2 are provided in Appendix B and Appendix C is devoted to the proof for Theorem 3.

Remark 5. *Since* $\hat{\boldsymbol{g}}_\tau(z_0)$ *is based on generated regressors* $\hat{\boldsymbol{X}}_t$*, it is not surprising to see that the asymptotic variance term of* $\hat{\boldsymbol{g}}_\tau(z_0)$ *contains additional two terms* $H_1(z_0)$ *and* $H_2(z_0)$*, which involve* $\boldsymbol{\Pi}_{a,t}$ *in the second step. Similar results of asymptotic variance were also obtained by Xiao and Koenker (2009), which can be seen as a nature of any multi-stage approach; see, for example, Cai, Das, Xiong and Wu (2006) for more discussions.*

2.4 Covariance Estimate

For constructing confidence intervals for the estimated functional coefficients in the empirical study, it turns to discussing how to obtain consistent estimator of the asymptotic covariance matrix $\Sigma_\tau(z_0)$. To this end, one needs to estimate $\boldsymbol{D}(z_0)$, $\boldsymbol{\Phi}_a$, $\Gamma(z_0)$, $H_1(z_0)$, $H_2(z_0)$, $\Omega(z_0)$ and $\Omega^*(z_0)$ consistently. The procedure of estimating covariance matrix and constructing confidence interval is summarized as follows:

Step 1: Given any grid point $z_0\in[a,b]$, use the proposed three-stage procedure to obtain $\boldsymbol{\Pi}_{a,t}$, $\boldsymbol{\Upsilon}_{a,t}$, $\hat{\boldsymbol{X}}_t$ and $\hat{\boldsymbol{g}}_\tau(z_0)$.

Step 2: Obtain estimators $\hat{\boldsymbol{D}}(z_0)=\sum_{t=1}^n\boldsymbol{\Pi}_{a,t}\boldsymbol{\Pi}_{a,t}^\top K_h(Z_t-z_0)/n$, $\hat{\boldsymbol{\Phi}}_a=\sum_{t=m+1}^n w_{1t}\boldsymbol{\Pi}_{a,t}\boldsymbol{\Pi}_{a,t}^\top/n$ and $\boldsymbol{E}_{xw}(z_0)=\sum_{t=1}^n\hat{\boldsymbol{X}}_t\boldsymbol{\Pi}_{a,t}^\top K_h(Z_t-z_0)/n$, where $w_{1t}=I(\hat{q}_{\tau,t}-\delta_{1n}<Y_t\le\hat{q}_{\tau,t}+\delta_{1n})/(2\delta_{1n})$ for any $\delta_{1n}\to 0$ as $n\to\infty$.

Step 3: Construct estimators $\hat{H}_1(z_0)=\boldsymbol{E}_{xw}(z_0)\hat{\boldsymbol{\Phi}}_a^{-1}\hat{\Gamma}^\top(z_0)+\hat{\Gamma}(z_0)\hat{\boldsymbol{\Phi}}_a^{-1}(\boldsymbol{E}_{xw}(z_0))^\top$, $\hat{H}_2(z_0)=\hat{\Gamma}(z_0)\hat{\boldsymbol{\Phi}}_a^{-1}\hat{\boldsymbol{D}}(z_0)\hat{\boldsymbol{\Phi}}_a^{-1}\hat{\Gamma}^\top(z_0)$, $\hat{\Gamma}(z_0)=\sum_{t=1}^n w_{2t}\hat{\boldsymbol{X}}_t\hat{\boldsymbol{g}}_\tau^\top(z_0)\boldsymbol{\Upsilon}_{a,t}K_h(Z_t-z_0)/n$, $\hat{\Omega}(z_0)=\sum_{t=1}^n\hat{\boldsymbol{X}}_t\hat{\boldsymbol{X}}_t^\top K_h(Z_t-z_0)/n$ and $\hat{\Omega}^*(z_0)=\sum_{t=1}^n w_{2t}\hat{\boldsymbol{X}}_t\hat{\boldsymbol{X}}_t^\top K_h(Z_t-z_0)/n$, where $w_{2t}=$

$I(\hat{\boldsymbol{g}}_\tau^\top(z_0)\hat{\boldsymbol{X}}_t - \delta_{2n} < Y_t \le \hat{\boldsymbol{g}}_\tau^\top(z_0)\hat{\boldsymbol{X}}_t + \delta_{2n})/(2\delta_{2n})$ for any $\delta_{2n} \to 0$ as $n \to \infty$.

Step 4: Obtain estimator $\hat{\boldsymbol{\Xi}}(z_0) = \tau(1-\tau)\nu_0[\hat{\Omega}(z_0) - \hat{H}_1(z_0) + \hat{H}_2(z_0)]$. Then, $\hat{\Sigma}_\tau(z_0) = (\hat{\Omega}^*(z_0))^{-1}\hat{\boldsymbol{\Xi}}(z_0)(\hat{\Omega}^*(z_0))^{-1}$ is a consistent estimate of $\Sigma_\tau(z_0)$.

Step 5: With $\hat{\Sigma}_\tau(z_0)$ at hand, let $se(\hat{g}_{j,\tau}(z_0)) = \left[\widehat{Var}(\hat{g}_{j,\tau}(z_0))/nh\right]^{1/2}$, where $\widehat{Var}(\hat{g}_{j,\tau}(z_0))$ is the jth diagonal element of $\hat{\Sigma}_\tau(z_0)$, with $\hat{g}_{j,\tau}(z_0)$ being the jth element of $\hat{\boldsymbol{g}}_\tau(z_0)$.

Then, for given $0 < \alpha < 1$, the $100(1-\alpha)\%$ confidence interval for $g_{j,\tau}(z_0)$ without the asymptotic bias correction can be constructed by $[\hat{g}_{j,\tau}(z_0) - c_{\alpha/2} \times se(\hat{g}_{j,\tau}(z_0)), \hat{g}_{j,\tau}(z_0) + c_{\alpha/2} \times se(\hat{g}_{j,\tau}(z_0))]$, where $c_{\alpha/2}$ is the upper $\alpha/2$-percentile of standard normal random variables.

Finally, note that similar to the proof in Cai and Xu (2008), one can show that $\hat{\boldsymbol{D}}(z_0) = f_z(z_0)\boldsymbol{D}(z_0) + o_p(1)$ and $\hat{\boldsymbol{\Phi}}_a = \boldsymbol{\Phi}_a + o_p(1)$, respectively. Also, in Appendix C, it shows that the above estimators are consistent; that is, $\hat{\Gamma}(z_0) = f_z(z_0)\Gamma(z_0) + o_p(1)$, $\hat{H}_1(z_0) = f_z(z_0)H_1(z_0) + o_p(1)$, $\hat{H}_2(z_0) = f_z(z_0)H_2(z_0) + o_p(1)$, $\hat{\Omega}(z_0) = f_z(z_0)\Omega(z_0) + o_p(1)$, and $\hat{\Omega}^*(z_0) = f_z(z_0)\Omega^*(z_0) + o_p(1)$. The proof of these results relies on the uniform consistency (in probability) of the estimator $\hat{q}_{\tau,t}$, which is guaranteed by Theorem 2(c). Therefore, it shows in Appendix C that indeed, $\hat{\Sigma}_\tau(z_0) = (\hat{\Omega}^*(z_0))^{-1}\hat{\boldsymbol{\Xi}}(z_0)(\hat{\Omega}^*(z_0))^{-1}$ is a consistent estimate of $\Sigma_\tau(z_0)$.

2.5 Practical Implementations

In real application, we need to know how to choose the smoothing variable Z_t, the truncation parameter $m = m_n$, and the bandwidth h. To this end, some suggestions are provided below.

First, it is important to choose an appropriate smoothing variable Z_t in applying functional-coefficient VAR model for dynamic quantiles in (2). Knowledge on physical background or economic theory of the data may be very helpful, as we have witnessed in modeling the real data in Section 4 by choosing Z_t to be the first difference of daily log series of the U.S. dollar index. Without any prior information, it is pertinent to choose Z_t in terms of some data-driven methods such as the Akaike information criterion, cross-validation, and other criteria. Ideally, Z_t can be selected as a linear function of given explanatory variables according to some optimal statistical selection criterion such as LASSO type methods, or an economic index based on some economic theory; see, for instance, Cai, Juhl and Yang (2015). Nevertheless, here we would recommend using a simple and practical approach proposed by Cai, Fan and Yao (2000) or Cai et al. (2015) in practice.

Second, with the help of the first stage, the truncation parameter m can be set such that the dimension of $\mathbf{\Pi}_t$ ($M_n = (1+\kappa)\sum_{l=1}^{m} K^l$) becomes much larger than the sample sizes n. Under Assumption A2 in Appendix A, it suffices to select a truncation m such that $m \geq r_n$. Here, the rate of r_n is given in Assumption A2. In practice, one may apply the forward selection method introduced by Cheng, Honda and Zhang (2016) to select m. Of course, other data-driven methods such as the Akaike information criterion, cross-validation, and other criteria can also be considered for selecting m. To reduce computational burden, we choose m as a sufficiently large constant multiple of $n^{1/8}$, which is used in our simulation study in Section 3 and the empirical analysis in Section 4.

Finally, we would like to address how to select the bandwidth h at the third step. It is well known that the bandwidth plays an essential role in the trade-off between reducing bias and variance. In view of (12), it is about selecting the bandwidth in the context of estimating the coefficient functions in the quantile regression. Therefore, we recommend the method proposed in Cai and Xu (2008) for selecting h in (12), which is used in our simulation study in Section 3 and empirical example in Section 4.

3 A Monte Carlo Simulation Study

In this section, we provide a simulation example to exam the finite sample performance of the proposed three-stage estimation for functional coefficients. In the first stage of estimation, we choose the tuning parameter $\lambda_{n,0}$ by a data-driven pivotal. In particular, define $\Lambda = \left\| \sum_{t=1}^{n} \mathbf{\Pi}_t \{\tau - I(u_t \leq \tau)\} \right\|_\infty$, where $\{u_t\}_{t=1}^{n}$ are i.i.d. Uniform $[0,1]$ random variables, independently distributed from the regressors $\mathbf{\Pi}_t$. Then, $\lambda_{n,0}$ is set as $\lambda_{n,0} = c \cdot \Lambda(1-\alpha|\mathbf{\Pi})$, where $\Lambda(1-\alpha|\mathbf{\Pi})$ is the $(1-\alpha)$-quantile of Λ conditional on $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_n)^\top$, and the constant $c > 1$ depends on the design. In this study, we set $c = 1.1$ and $1 - \alpha = 0.9$, similar to the choice for least squares in Chernozhukov et al. (2021). Using the the exponential tail bounds as in Lemma B.4 in Chernozhukov et al. (2021), one can obtain the asymptotic bound for $\Lambda(1-\alpha|\mathbf{\Pi})$. The proof is omitted, because it is similar to that of Lemmas B.4 and B.6 in Appendix B. In the second stage, we select the parameter $\lambda_{n,1}$ by the generalized information criterion (GIC) as in Fan and Tang (2013) and Zheng, Peng and He (2015), defined as

$$\text{GIC}(\lambda) = \log\left(\sum_{t=1}^{n} \rho_\tau(Y_t - \mathbf{\Pi}_t^\top \hat{\boldsymbol{c}}_\lambda)\right) + df_\lambda \cdot \frac{\log(\text{card}(\tilde{T}_n)) \cdot \log(\log(n))}{n},$$

where $\hat{\boldsymbol{c}}_\lambda$ is the minimizer of (11) given $\lambda_{n,1} = \lambda$, and the degrees of freedom df_λ is defined as the number of nonzero estimated components of $\hat{\boldsymbol{c}}_\lambda$ conditional on λ. At Step 3, we choose optimal bandwidth $\hat{h}$ by the nonparametric Akaiki information criterion (AIC) as in Cai and Xu (2008). Finally, the Epanechnikov kernel $K(x) = 0.75(1-x^2)I(|x| \leq 1)$ is used, $m = \lfloor 0.8n^{1/8} \rfloor$ and $K = \lfloor 1.5n^{1/5} \rfloor$.

In this example, for $1 \leq i \leq 4$, the data generating process (DGP) is given by $Y_{it} = \sigma_{it}\varepsilon_{it}$ with $\sigma_{it} = \gamma_{i0}(Z_t) + \gamma_{i1,\epsilon_{it}}(Z_t)\sigma_{1(t-1)} + \gamma_{i2,\chi_{it}}(Z_t)\sigma_{2(t-1)} + \gamma_{i3,\epsilon_{it}}(Z_t)\sigma_{3(t-1)} + \gamma_{i4,\chi_{it}}(Z_t)\sigma_{4(t-1)} + \beta_{i1}(Z_t)|Y_{1(t-1)}| + \beta_{i2}(Z_t)|Y_{2(t-1)}| + \beta_{i3}(Z_t)|Y_{3(t-1)}| + \beta_{i4}(Z_t)|Y_{4(t-1)}|$, where $\gamma_{10}(z) = \gamma_{30}(z) = 1.5\exp(-3(z+1)^2) + \exp(-8(z-1)^2)$, $\gamma_{20}(z) = \gamma_{40}(z) = 1.5\exp(-3(z-1)^2) + \exp(-8(z+1)^2)$, $\epsilon_{it} = 0.2U_{it}^2 + 0.8$ and $\chi_{it} = 0.2\exp(U_{it}) + 0.8$ with $U_{it} \sim$ iid Uniform $[0,1]$. In addition, let $\gamma_{ij,\epsilon_{it}}(z) = \gamma_{ij}(z) \cdot \epsilon_{it}$, $\gamma_{ij,\chi_{it}}(z) = \gamma_{ij}(z) \cdot \chi_{it}$. Then, $\gamma_{ij,\epsilon_{it}}(z)$, $\gamma_{ij,\chi_{it}}(z)$ and $\beta_{ij}(z)$ for $1 \leq i, j \leq 4$ are defined as follows. For $i = 1$, $\gamma_{i1,\epsilon_{it}}(z) = 0.15\left\{1 + \exp(-4z)\right\}^{-1}\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) = (0.04z^2)\chi_{it}$, $\gamma_{i3,\epsilon_{it}}(z) = (0.15\exp(-4z)\left\{1+\exp(-4z)\right\}^{-1})\epsilon_{it}$, $\gamma_{i4,\chi_{it}}(z) = (0.1\cos^2(z))\chi_{it}$. For $i = 2$, $\gamma_{i1,\epsilon_{it}}(z) = \gamma_{i3,\epsilon_{it}}(z) = (0.1\sin(-0.8\pi z) + 0.1)\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) = 0.15\left\{1 + \exp(-4z)\right\}^{-1}\chi_{it}$, $\gamma_{i4,\chi_{it}}(z) = (0.15\exp(-4z)\left\{1 + \exp(-4z)\right\}^{-1})\chi_{it}$. For $i = 3$, $\gamma_{i1,\epsilon_{it}}(z) = \gamma_{i3,\epsilon_{it}}(z) = (0.1\sin(0.8\pi z) + 0.1)\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) = \gamma_{i4,\chi_{it}}(z) = (0.1\cos(0.8\pi z) + 0.1)\chi_{it}$. For $i = 4$, $\gamma_{i1,\epsilon_{it}}(z) = \gamma_{i3,\epsilon_{it}}(z) =$

$(0.1\cos(0.8\pi z)+0.1)\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) = \gamma_{i4,\chi_{it}}(z) = (0.1\sin(0.8\pi z)+0.1)\chi_{it}$. As for $\beta_{ij}(z)$, we set $\beta_{ij}(z) = \gamma_{ij}(z)$ for $1 \le i, j \le 4$. Finally, ε_{it} are mutually iid from $\mathcal{N}(0,1)$.

Thus, for $1 \le i \le 4$, the quantile function of our DGP is given by the model in (1) with $\kappa = 4$, $\mathbb{Y}_t = (|Y_{1t}|, |Y_{2t}|, |Y_{3t}|, |Y_{4t}|)^T$, $q = p = 1$ and $Z_{it} = Z_t$, where Z_t is generated from Uniform $[-2, 2]$ independently. Also, note that $\gamma_{i0,\tau}(\cdot) = \gamma_{i0}(\cdot)\Phi^{-1}(\tau)$, $\gamma_{i1,\tau}(\cdot) = \gamma_{i1}(\cdot)(0.2\tau^2 + 0.8)$, $\gamma_{i3,\tau}(\cdot) = \gamma_{i3}(\cdot)(0.2\tau^2 + 0.8)$, while $\gamma_{i2,\tau}(\cdot) = \gamma_{i2}(\cdot)(0.2\exp(\tau) + 0.8)$, $\gamma_{i4,\tau}(\cdot) = \gamma_{i4}(\cdot)(0.2\exp(\tau) + 0.8)$ and $\beta_{ij,\tau}(\cdot) = \beta_{ij}(\cdot)\Phi^{-1}(\tau)$ for $1 \le i, j \le 4$, with $\Phi(\cdot)$ being the distribution function of the standard normal. Therefore, $\gamma_{i0,\tau}(\cdot)$, $\gamma_{ij,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$ are functions of τ, suggesting different covariate effects at different levels of τ.

It is worthwhile to note that the DGP used in this Monte Carlo study is a multivariate GARCH model with functional coefficients, which is closely related to our real example in Section 4 for following reasons. First, as discussed in Remark 1, the conditional quantile function of GARCH process has a representation of dynamic quantile model, which is naturally suitable to describe interdependences among conditional VaRs in our new financial network. Second, the types of functional coefficients in this simulation seem to be rich enough to cover different classes of variation of tail dependence in the real application. Third, since GARCH-type models have been proven to be highly successful in financial applications, data that are generated from GARCH-type models may be appropriate to simulate the real example in this paper.

To assess the finite sample performance of the proposed nonparametric estimators, we utilize the mean absolute deviation error (MADE) for $\gamma_{i0,\tau}(\cdot)$, $\gamma_{ij,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$, defined as

$$\text{MADE}(\gamma) = \frac{1}{n_0}\sum_{k}^{n_0} |\hat{\gamma}_\tau(z_k) - \gamma_\tau(z_k)|, \quad \text{and} \quad \text{MADE}(\beta_{ij,\tau}) = \frac{1}{n_0}\sum_{k}^{n_0} |\hat{\beta}_{ij,\tau}(z_k) - \beta_{ij,\tau}(z_k)|,$$

where $\gamma_\tau(\cdot)$ can be either $\gamma_{ij,\tau}(\cdot)$ or $\gamma_{i0,\tau}(\cdot)$, both $\hat{\gamma}_\tau(\cdot)$ and $\hat{\beta}_{ij,\tau}(\cdot)$ are local linear quantile estimates of $\gamma_\tau(\cdot)$ and $\beta_{ij,\tau}(\cdot)$, respectively, and $\{z_k = 0.1(k-1) - 1.75 : 1 \le k \le n_0 = 36\}$ are the grid points. Also note that in this example, $q_{\tau,t,i} = \sigma_{it}F_\varepsilon^{-1}(\tau) = 0$ when $\tau = 0.5$, which leads the quantile regression problem to be ill-posed so that the results for $\tau = 0.5$ are omitted.

Therefore, we only consider τ's level to be 0.05, 0.15, 0.85 and 0.95 and the sample sizes are $n = 500$, 1500 and 4000. For each setting, we replicate simulation 500 times and compute the median and standard deviation (in parentheses) of 500 MADE values. Finally, the results are reported in Tables 1 - 4 only for $\tau = 0.05$, 0.15 and 0.95 but the results for $\tau = 0.85$ are omitted due to the space limitation, available upon request. One can see clearly from Tables 1 - 4 that both median and standard deviation of 500 MADE values steadily decrease as the sample size increases for all four values of τ.

Table 1: Simulation results for $\gamma_{10,\tau}(\cdot)$, $\gamma_{20,\tau}(\cdot)$, $\gamma_{30,\tau}(\cdot)$, $\gamma_{40,\tau}(\cdot)$, and $\gamma_{ij,\tau}(\cdot)$ for $i = 1, 2$ and for $1 \le j \le 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(γ_{10})	MADE(γ_{20})	MADE(γ_{10})	MADE(γ_{20})	MADE(γ_{10})	MADE(γ_{20})
0.05	0.666 (0.120)	0.748 (0.202)	0.456 (0.059)	0.599 (0.087)	0.347 (0.042)	0.501 (0.043)
0.15	0.357 (0.074)	0.415 (0.091)	0.245 (0.047)	0.330 (0.046)	0.202 (0.030)	0.234 (0.043)
0.95	0.694 (0.109)	0.715 (0.165)	0.484 (0.066)	0.603 (0.094)	0.328 (0.045)	0.492 (0.047)
	MADE(γ_{30})	MADE(γ_{40})	MADE(γ_{30})	MADE(γ_{40})	MADE(γ_{30})	MADE(γ_{40})
0.05	0.742 (0.160)	0.690 (0.203)	0.545 (0.102)	0.655 (0.090)	0.455 (0.072)	0.500 (0.066)
0.15	0.455 (0.083)	0.442 (0.093)	0.337 (0.061)	0.325 (0.080)	0.264 (0.042)	0.260 (0.049)
0.95	0.707 (0.204)	0.693 (0.205)	0.557 (0.120)	0.555 (0.121)	0.439 (0.070)	0.469 (0.066)
	MADE(γ_{11})	MADE(γ_{12})	MADE(γ_{11})	MADE(γ_{12})	MADE(γ_{11})	MADE(γ_{12})
0.05	0.182 (0.074)	0.137 (0.058)	0.106 (0.032)	0.094 (0.033)	0.087 (0.029)	0.061 (0.018)
0.15	0.165 (0.074)	0.135 (0.060)	0.108 (0.040)	0.081 (0.046)	0.092 (0.036)	0.075 (0.026)
0.95	0.176 (0.082)	0.140 (0.063)	0.112 (0.034)	0.084 (0.030)	0.107 (0.032)	0.071 (0.020)
	MADE(γ_{13})	MADE(γ_{14})	MADE(γ_{13})	MADE(γ_{14})	MADE(γ_{13})	MADE(γ_{14})
0.05	0.157 (0.067)	0.141 (0.060)	0.086 (0.031)	0.071 (0.027)	0.070 (0.027)	0.068 (0.023)
0.15	0.148 (0.062)	0.127 (0.052)	0.100 (0.041)	0.084 (0.030)	0.069 (0.025)	0.069 (0.023)
0.95	0.152 (0.071)	0.129 (0.063)	0.101 (0.032)	0.083 (0.028)	0.088 (0.031)	0.075 (0.022)
	MADE(γ_{21})	MADE(γ_{22})	MADE(γ_{21})	MADE(γ_{22})	MADE(γ_{21})	MADE(γ_{22})
0.05	0.214 (0.095)	0.175 (0.071)	0.109 (0.044)	0.111 (0.047)	0.088 (0.027)	0.074 (0.032)
0.15	0.206 (0.085)	0.171 (0.062)	0.113 (0.050)	0.101 (0.045)	0.105 (0.039)	0.096 (0.032)
0.95	0.201 (0.088)	0.184 (0.080)	0.120 (0.054)	0.115 (0.042)	0.102 (0.040)	0.101 (0.031)
	MADE(γ_{23})	MADE(γ_{24})	MADE(γ_{23})	MADE(γ_{24})	MADE(γ_{23})	MADE(γ_{24})
0.05	0.169 (0.075)	0.164 (0.064)	0.099 (0.035)	0.098 (0.033)	0.080 (0.029)	0.077 (0.028)
0.15	0.177 (0.073)	0.165 (0.067)	0.108 (0.049)	0.102 (0.038)	0.094 (0.030)	0.086 (0.026)
0.95	0.183 (0.080)	0.176 (0.072)	0.111 (0.042)	0.122 (0.040)	0.099 (0.033)	0.091 (0.033)

Finally, we illustrate the finite sample performance for the consistent covariance estimation

Table 2: Simulation results for $\gamma_{ij,\tau}(\cdot)$ for $i = 3, 4$ and for $1 \leq j \leq 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(γ_{31})	MADE(γ_{32})	MADE(γ_{31})	MADE(γ_{32})	MADE(γ_{31})	MADE(γ_{32})
0.05	0.194 (0.083)	0.165 (0.063)	0.132 (0.036)	0.134 (0.042)	0.090 (0.028)	0.084 (0.026)
0.15	0.187 (0.083)	0.153 (0.069)	0.121 (0.058)	0.107 (0.059)	0.110 (0.038)	0.101 (0.036)
0.95	0.257 (0.087)	0.220 (0.071)	0.142 (0.045)	0.134 (0.040)	0.115 (0.039)	0.109 (0.029)
	MADE(γ_{33})	MADE(γ_{34})	MADE(γ_{33})	MADE(γ_{34})	MADE(γ_{33})	MADE(γ_{34})
0.05	0.161 (0.066)	0.162 (0.065)	0.110 (0.035)	0.116 (0.034)	0.095 (0.030)	0.081 (0.025)
0.15	0.171 (0.073)	0.154 (0.061)	0.109 (0.046)	0.105 (0.045)	0.088 (0.030)	0.092 (0.031)
0.95	0.215 (0.069)	0.196 (0.070)	0.124 (0.045)	0.126 (0.035)	0.115 (0.037)	0.102 (0.026)
	MADE(γ_{41})	MADE(γ_{42})	MADE(γ_{41})	MADE(γ_{42})	MADE(γ_{41})	MADE(γ_{42})
0.05	0.235 (0.099)	0.191 (0.076)	0.108 (0.041)	0.106 (0.036)	0.084 (0.032)	0.083 (0.026)
0.15	0.187 (0.086)	0.176 (0.077)	0.133 (0.066)	0.123 (0.059)	0.110 (0.034)	0.103 (0.038)
0.95	0.234 (0.084)	0.210 (0.078)	0.156 (0.040)	0.152 (0.048)	0.105 (0.037)	0.119 (0.035)
	MADE(γ_{43})	MADE(γ_{44})	MADE(γ_{43})	MADE(γ_{44})	MADE(γ_{43})	MADE(γ_{44})
0.05	0.201 (0.076)	0.176 (0.064)	0.086 (0.043)	0.100 (0.042)	0.083 (0.037)	0.078 (0.030)
0.15	0.160 (0.073)	0.153 (0.072)	0.125 (0.064)	0.123 (0.044)	0.088 (0.028)	0.093 (0.031)
0.95	0.204 (0.072)	0.197 (0.075)	0.139 (0.040)	0.148 (0.049)	0.094 (0.038)	0.090 (0.032)

Table 3: Simulation results for $\beta_{ij,\tau}(\cdot)$ for $i = 1, 2$ and for $1 \leq j \leq 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(β_{11})	MADE(β_{12})	MADE(β_{11})	MADE(β_{12})	MADE(β_{11})	MADE(β_{12})
0.05	0.232 (0.088)	0.165 (0.068)	0.130 (0.050)	0.105 (0.036)	0.092 (0.025)	0.071 (0.020)
0.15	0.140 (0.059)	0.116 (0.047)	0.103 (0.033)	0.077 (0.028)	0.065 (0.024)	0.048 (0.018)
0.95	0.206 (0.086)	0.164 (0.069)	0.124 (0.043)	0.105 (0.033)	0.091 (0.027)	0.071 (0.022)
	MADE(β_{13})	MADE(β_{14})	MADE(β_{13})	MADE(β_{14})	MADE(β_{13})	MADE(β_{14})
0.05	0.155 (0.071)	0.161 (0.065)	0.109 (0.036)	0.101 (0.040)	0.068 (0.028)	0.074 (0.024)
0.15	0.115 (0.047)	0.116 (0.046)	0.086 (0.030)	0.082 (0.028)	0.052 (0.018)	0.052 (0.018)
0.95	0.173 (0.071)	0.155 (0.068)	0.107 (0.038)	0.096 (0.031)	0.080 (0.024)	0.068 (0.022)
	MADE(β_{21})	MADE(β_{22})	MADE(β_{21})	MADE(β_{22})	MADE(β_{21})	MADE(β_{22})
0.05	0.257 (0.103)	0.194 (0.077)	0.159 (0.063)	0.122 (0.053)	0.106 (0.042)	0.078 (0.033)
0.15	0.180 (0.075)	0.139 (0.053)	0.112 (0.045)	0.083 (0.034)	0.082 (0.029)	0.063 (0.021)
0.95	0.260 (0.101)	0.195 (0.092)	0.155 (0.062)	0.114 (0.050)	0.116 (0.042)	0.072 (0.034)
	MADE(β_{23})	MADE(β_{24})	MADE(β_{23})	MADE(β_{24})	MADE(β_{23})	MADE(β_{24})
0.05	0.215 (0.094)	0.196 (0.084)	0.140 (0.055)	0.103 (0.044)	0.087 (0.028)	0.080 (0.029)
0.15	0.145 (0.056)	0.134 (0.053)	0.091 (0.034)	0.082 (0.032)	0.064 (0.024)	0.065 (0.021)
0.95	0.197 (0.089)	0.195 (0.084)	0.136 (0.051)	0.108 (0.044)	0.095 (0.032)	0.076 (0.026)

Table 4: Simulation results for $\beta_{ij,\tau}(\cdot)$ for $i = 3, 4$ and for $1 \le j \le 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(β_{31})	MADE(β_{32})	MADE(β_{31})	MADE(β_{32})	MADE(β_{31})	MADE(β_{32})
0.05	0.249 (0.110)	0.193 (0.082)	0.192 (0.059)	0.141 (0.047)	0.113 (0.038)	0.089 (0.030)
0.15	0.163 (0.068)	0.142 (0.057)	0.118 (0.043)	0.087 (0.039)	0.088 (0.028)	0.067 (0.020)
0.95	0.302 (0.111)	0.238 (0.079)	0.190 (0.056)	0.158 (0.051)	0.119 (0.038)	0.097 (0.030)
	MADE(β_{33})	MADE(β_{34})	MADE(β_{33})	MADE(β_{34})	MADE(β_{33})	MADE(β_{34})
0.05	0.191 (0.076)	0.178 (0.075)	0.150 (0.049)	0.145 (0.052)	0.101 (0.034)	0.096 (0.026)
0.15	0.137 (0.054)	0.133 (0.057)	0.092 (0.031)	0.094 (0.039)	0.071 (0.021)	0.069 (0.024)
0.95	0.246 (0.076)	0.230 (0.078)	0.149 (0.047)	0.135 (0.041)	0.098 (0.033)	0.092 (0.030)
	MADE(β_{41})	MADE(β_{42})	MADE(β_{41})	MADE(β_{42})	MADE(β_{41})	MADE(β_{42})
0.05	0.291 (0.107)	0.242 (0.086)	0.167 (0.053)	0.149 (0.062)	0.119 (0.045)	0.096 (0.034)
0.15	0.172 (0.075)	0.145 (0.057)	0.127 (0.044)	0.104 (0.034)	0.082 (0.028)	0.069 (0.024)
0.95	0.283 (0.107)	0.243 (0.075)	0.206 (0.058)	0.169 (0.059)	0.109 (0.040)	0.095 (0.032)
	MADE(β_{43})	MADE(β_{44})	MADE(β_{43})	MADE(β_{44})	MADE(β_{43})	MADE(β_{44})
0.05	0.234 (0.086)	0.235 (0.073)	0.134 (0.049)	0.149 (0.057)	0.092 (0.035)	0.100 (0.036)
0.15	0.147 (0.059)	0.134 (0.053)	0.107 (0.039)	0.110 (0.038)	0.067 (0.024)	0.071 (0.025)
0.95	0.224 (0.074)	0.221 (0.089)	0.160 (0.056)	0.172 (0.049)	0.091 (0.033)	0.093 (0.029)

given in Section 2.4 via evaluating the pointwise confidence intervals (CI) with the asymptotic bias ignored. To do this, define $\widehat{Var}(\cdot)$ as the asymptotic variance calculated by the estimators presented in Section 2.4. Then, we compute the average of empirical coverage rates (AECR) of 95% pointwise CI of $\gamma_{ij,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$ without the asymptotic bias correction for $1 \le i, j \le 4$, defined as,

$$\text{AECR}(\gamma_{ij,\tau}) = \frac{1}{n_0 B} \sum_{k}^{n_0} \sum_{b=1}^{B} I_b\{\gamma_{ij,\tau}(z_k) \in \hat{\gamma}_{ij,\tau}(z_k) \pm 1.96 \times se(\hat{\gamma}_{ij,\tau}(z_k))\},$$

where $se(\hat{\gamma}_{ij,\tau}(\cdot)) = \left[\widehat{Var}(\hat{\gamma}_{ij,\tau}(\cdot))/nh\right]^{1/2}$, $I_b\{\gamma_{ij,\tau}(\cdot) \in \hat{\gamma}_{ij,\tau}(\cdot) \pm 1.96 \times se(\hat{\gamma}_{ij,\tau}(\cdot))\}$ is an indicator function which equals to 1 if $\gamma_{ij,\tau}(\cdot)$ is covered by the interval $\hat{\gamma}_{ij,\tau}(\cdot) \pm 1.96 \times se(\hat{\gamma}_{ij,\tau}(\cdot))$ in the bth time of replication (equals to 0, otherwise), and the number of replication times is $B = 500$. Similarly, AECR($\beta_{ij,\tau}$), $se(\hat{\beta}_{ij,\tau}(\cdot))$, and $I_b\{\beta_{ij,\tau}(\cdot) \in \hat{\beta}_{ij,\tau}(\cdot) \pm 1.96 \times se(\hat{\beta}_{ij,\tau}(\cdot))\}$ can be defined in the same fashion. The simulation results are presented in Table 5, for $n = 4000$ and $\tau = 0.05$, 0.15 and 0.95. From Table 5, one can see basically that for each setting, AECRs of 95% pointwise CIs are close to the nominal level 95% for all settings. In general, the results of this simulated experiment demonstrate that the proposed procedure is reliable and works fairly well.

Table 5: Average of empirical coverage rates (AECR) of 95% pointwise confidence intervals for $\gamma_{ij,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$ without the asymptotic bias correction, for $1 \leq i, j \leq 4$ and $n = 4000$.

τ	Coverage Rates of $\hat{\gamma}_{ij,\tau}(\cdot)$				Coverage Rates of $\hat{\beta}_{ij,\tau}(\cdot)$			
	$\hat{\gamma}_{11,\tau}$	$\hat{\gamma}_{12,\tau}$	$\hat{\gamma}_{13,\tau}$	$\hat{\gamma}_{14,\tau}$	$\hat{\beta}_{11,\tau}$	$\hat{\beta}_{12,\tau}$	$\hat{\beta}_{13,\tau}$	$\hat{\beta}_{14,\tau}$
0.05	0.945	0.951	0.945	0.967	0.976	0.976	0.972	0.976
0.15	0.940	0.942	0.921	0.931	0.965	0.951	0.954	0.957
0.95	0.944	0.903	0.949	0.935	0.971	0.965	0.967	0.964
	$\hat{\gamma}_{21,\tau}$	$\hat{\gamma}_{22,\tau}$	$\hat{\gamma}_{23,\tau}$	$\hat{\gamma}_{24,\tau}$	$\hat{\beta}_{21,\tau}$	$\hat{\beta}_{22,\tau}$	$\hat{\beta}_{23,\tau}$	$\hat{\beta}_{24,\tau}$
0.05	0.946	0.963	0.955	0.969	0.976	0.980	0.977	0.973
0.15	0.950	0.944	0.905	0.930	0.959	0.960	0.959	0.961
0.95	0.969	0.934	0.963	0.939	0.978	0.985	0.979	0.977
	$\hat{\gamma}_{31,\tau}$	$\hat{\gamma}_{32,\tau}$	$\hat{\gamma}_{33,\tau}$	$\hat{\gamma}_{34,\tau}$	$\hat{\beta}_{31,\tau}$	$\hat{\beta}_{32,\tau}$	$\hat{\beta}_{33,\tau}$	$\hat{\beta}_{34,\tau}$
0.05	0.973	0.954	0.965	0.972	0.981	0.978	0.984	0.977
0.15	0.963	0.957	0.939	0.940	0.965	0.968	0.964	0.949
0.95	0.965	0.902	0.961	0.950	0.974	0.974	0.976	0.975
	$\hat{\gamma}_{41,\tau}$	$\hat{\gamma}_{42,\tau}$	$\hat{\gamma}_{43,\tau}$	$\hat{\gamma}_{44,\tau}$	$\hat{\beta}_{41,\tau}$	$\hat{\beta}_{42,\tau}$	$\hat{\beta}_{43,\tau}$	$\hat{\beta}_{44,\tau}$
0.05	0.972	0.948	0.959	0.973	0.982	0.973	0.978	0.977
0.15	0.959	0.951	0.934	0.929	0.961	0.967	0.966	0.958
0.95	0.979	0.903	0.961	0.956	0.969	0.981	0.968	0.975

4 An Empirical Example

4.1 Empirical Models

In this section, the proposed model and estimation methods are applied to constructing and estimating a new class of dynamic financial networks in international equity markets. Different from the existing literature, the interdependences of this class of networks vary with a smoothing variable of general economy. Motivated by White et al. (2015), we consider the dependence between current and one-day lagged VaR. In particular, we define each linkage in our network as the dependence between VaR of return of one market index at time t and that of another at time $t-1$. Therefore, our network can be written as following equation system:

$$\text{VaR}_{it} = \boldsymbol{\gamma}_{i,\tau}^{\top}(Z_{t-1})\text{VaR}_{t-1}, \quad i = 1, 2, \ldots, \kappa, \tag{13}$$

where $\text{VaR}_{t-1} = (\text{VaR}_{1(t-1)}, \ldots, \text{VaR}_{\kappa(t-1)})^\top$ is a vector of VaRs for all market index returns at time $t-1$ and VaR_{it} is the VaR of the ith market index return at time t, which is described as follows $\text{VaR}_{it} = -\inf\{Y \in \mathbb{R} : P(Y_{it} > Y|\mathcal{F}_{t-1}) \leq 1-\tau\} = -\inf\{Y \in \mathbb{R} : F(Y|\mathcal{F}_{t-1}) > \tau\}$ for $i = 1, 2, \cdots, \kappa$ at a given $\tau \in (0,1)$. Here, $\mathcal{F}_{t-1}$ is the information set to present all information of the ith return available at time $t-1$ and $F(\cdot|\mathcal{F}_{t-1})$ represents the conditional distribution function of Y_{it} given $\mathcal{F}_{t-1}$. In addition, Z_{t-1} is a smoothing variable of general economy and $\boldsymbol{\gamma}_{i,\tau}(\cdot) = (\gamma_{i1,\tau}(\cdot), \ldots, \gamma_{i\kappa,\tau}(\cdot))^T$ is a $\kappa \times 1$ vector of functional coefficients. Then, we extract the quantile estimation of functional coefficients from equation system (13) and construct the matrix $|\hat{\boldsymbol{\Gamma}}_{1,\tau}(Z_{t-1})| = (|\hat{\gamma}_{ij,\tau}(Z_{t-1})|)_{\kappa\times\kappa}$ as our financial network, in which, $|\hat{\gamma}_{ij,\tau}(Z_{t-1})|$ represents the strength of dependence between VaR of return for the market index i at time t and that for the index j at time $t-1$, under τ-th quantile level, and is driven by the smoothing variable Z_{t-1}. Here, taking absolute value on each $\hat{\gamma}_{ij,\tau}(Z_{t-1})$ enables us to calculate and analyze indicators of connectedness, and details are reported in Section 4.3 later. Thus, matrix $|\hat{\boldsymbol{\Gamma}}_{1,\tau}(Z_{t-1})|$ is useful to capture risk interdependence and how it changes with a smoothing variable Z_{t-1}. Note that entries $|\hat{\boldsymbol{\Gamma}}_{1,\tau}(Z_{t-1})|$ correspond to the absolute value of the estimated values of $\{\gamma_{ij,\tau}(\cdot)\}$ in the network model in (3). Therefore, our three-stage procedures can be applied here for direct estimation of the interdependence among VaRs of returns for the market indices.

It is necessary to mention that we restrict our attention to the time-varying interdependence between VaRs of four financial market indices' returns at time t and those at time $t-1$[5], though Xu, Wang, Shin and Zheng (2022) detected significant contemporaneous effect of connected nodes on the conditional quantile function of financial return in the common shareholder network and the headquarter location based network. Different from the predetermined networks studied in Xu et al. (2022), the proposed financial network in our real example is constructed by tail

[5]Note that because of volatility persistence, the VaR of financial return may have a long memory behavior. However, according to the empirical results in Engle and Manganelli (2004) and White et al. (2015), the specification of autoregressive for dynamic quantile with lag one seems to perform well in out-of-sample testing. Thus, to simplify the implementation of our methodology, we follow White et al. (2015) and only consider modeling the dynamic pattern between VaR at time t and its one-period lag in this empirical study.

dependence among unobserved VaRs of financial indices' returns, where observable information of relationship among nodes (e.g., binary network data) can hardly be found. Thus, it still remains unclear to us whether contemporaneous interaction among VaRs of financial indices is significant in our real example or not. A rigorous study of this important issue is left as a future research.

4.2 Data

Our dataset includes the daily series between January 5, 2006 and February 10, 2021 for four major world equity market indices: the U.K. FTSE 100 Index, the Japanese Nikkei 225 Index, the U.S. S&P 500 Composite Index and the Chinese Shanghai Composite Index (SSE). We model the ith index's return series $Y_{it} = 50 \log(\pi_{it}/\pi_{i(t-1)})$, where $i = 1$, 2, 3, and 4 corresponds to the four aforementioned market indices in turn and π_{it} is ith index level at the tth day. The time range of data includes the financial crisis in the U.S. in 2008, the European sovereign debt crisis of 2011-2012, and the COVID-19 pandemic starting from 2019. The daily series of four market indices are downloaded in Yahoo Finance and the estimation sample sizes $n = 3254$. Thus, we take $m = \lfloor 0.8n^{1/8} \rfloor$ and $K = \lfloor 1.5n^{1/5} \rfloor$ in this empirical study. Similar to the example in Monte Carlo simulation study, $\hat{h}$ is selected by the nonparametric AIC criterion as in Cai and Xu (2008) and the Epanechnikov kernel is used. Although it is feasible to introduce more kinds of market index into the equation system (13), due to the computational burdens, we only consider risk co-dependences among four major markets' indices.

As for the smoothing variable Z_t, we choose $Z_t = 50 \log(D_t/D_{t-1})$, where D_t is the U.S. dollar index on the tth day and can be downloaded from the Federal Reserve Bank of St. Louis. The U.S. dollar index measures value of U.S. dollar against the currencies of a broad group of major U.S. trading partners, higher values of the index indicate a stronger U.S. dollar. This choice of smoothing variable is reasonable, because the exchange rate has been regarded as an important factor associated with international transmission of risk in many empirical studies. For instance, Menkhoff, Sarno, Schmelling and Schrimpf (2012) discussed the relation between innovations in global foreign exchange volatility and excess returns arising from strategies of carry trade, through which the risk spillover transmits from one country to others. In addition, Yang and Zhou (2017) showed that volatility spillover intensity increases with U.S. dollar depreciation. We do not claim that the U.S. dollar index is the only choice for smoothing variable, but we choose the U.S. dollar index because it contains more information about risk transmission among international equity markets. It is desirable to consider other variables of economic status as the smoothing variable and this may be left in a future study.

4.3 Empirical Results

The empirical analysis in this section includes two steps: First, we estimate $\gamma_{ij,\tau}(Z_{t-1})$ for each market index in the equation system in (13) under $\tau = 0.05$. Second, we use the estimated value of $\gamma_{ij,\tau}(Z_{t-1})$ to construct the matrix $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$, and do network analysis based on this matrix.

Before exploring the matrix $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$, it is important to exam whether each $\gamma_{ij,\tau}(Z_{t-1})$ in (13) varies significantly with Z_{t-1} or not. To this end, we estimate each $\gamma_{ij,\tau}(Z_{t-1})$ and corresponding 95% pointwise confidence intervals with the asymptotic bias ignored. Figure 1 depicts the corresponding estimation results, in which ij-th panel represents the result for $\gamma_{ij,\tau}(\cdot)$, respectively. The black solid line in each panel of Figure 1 represents the estimates of the $\gamma_{ij,\tau}(\cdot)$ for $1 \leq i, j \leq 4$ in (13) along various values of Z_{t-1} under $\tau = 0.05$, and the red dashed lines are 95% pointwise confidence intervals for each estimate without the asymptotic bias correction. From Figure 1, we clearly see that most of coefficient functions vary significantly over the interval $[-0.75, 0.75]$, which means that we can not use fixed-coefficient dynamic quantiles models to fit the data.

Next, we consider analyzing the matrix $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$, in which each entry is $|\gamma_{ij,\tau}(Z_{t-1})|$. To simplify notation, Z_{t-1} and τ are dropped from $|\hat{\gamma}_{ij,\tau}(Z_{t-1})|$ and $|\hat{\gamma}_{ji,\tau}(Z_{t-1})|$ in the matrix

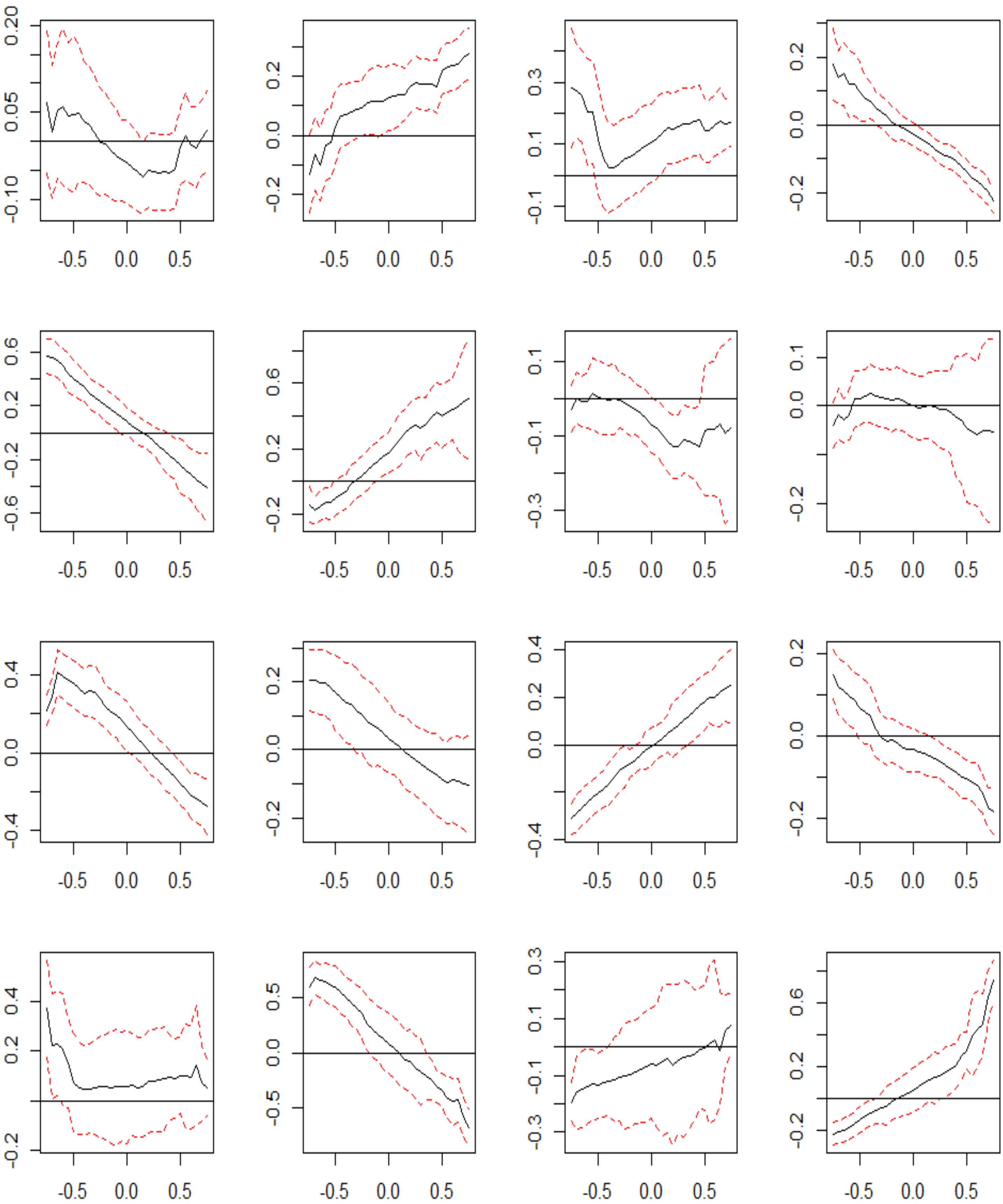

Figure 1: Plots of the estimated coefficient functions $\gamma_{ij,\tau}(\cdot)$ for $1 \leq i \leq 4$ and $1 \leq j \leq 4$ in (13) in the main article under $\tau = 0.05$ (black solid lines), in which ij-th panel represents the result for $\gamma_{ij,\tau}(\cdot)$, respectively. The red dashed lines in each panel indicate the 95% pointwise confidence interval for the estimate with the asymptotic bias ignored.

$|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$, in what follows. Then, $|\hat{\gamma}_{ji}|$ in the matrix $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$ represents the magnitude of dependence between the risk of market index i at time $t-1$ and that of market index j at time t. For the purpose of visualization, by following Härdle et al. (2016), we first define the levels of connectedness. The connectedness with respect to incoming links (CIL) is defined as $\sum_{i=1}^{4}|\hat{\gamma}_{ji}|$, which measures the risk spillover that was emitted from all four market indices one day ago and is received currently by a certain market index. Analogously, the connectedness with respect to outgoing links (COL) is defined as $\sum_{i=1}^{4}|\hat{\gamma}_{ij}|$, which measures the risk spillover emitted from a certain market index one day ago and is received currently by all market indices. Here, $i, j = 1, 2, 3, 4$ correspond to the four aforementioned market indices in turn. Intuitively, the CIL measures exposures of individual indices to systemic shocks from the financial network, while the COL measures contributions of individual indices for risk events in the network. Other than the CIL and COL, we also analyze the total connectedness in the global market, which is equal to $\sum_{j=1}^{4}\sum_{i=1}^{4}|\hat{\gamma}_{ij}|$ and indicates the total risk spillover in the global market, see Härdle et al. (2016) for more applications about these types of connectedness.

Figures 2 and 3 display the corresponding results along the same values of Z_{t-1}, under $\tau = 0.05$, respectively. In Figure 2, each panel displays the CIL and COL subject to the U.S. dollar change. The solid line in each panel represents values of COL and the dashed line indicates values of the CIL. For Figure 3, the vertical axis measures the total connectedness appeared in international equity markets and the horizontal axises in both figures are the same as those in each panel of Figure 1.

Figure 2 shows that the curves of all four major market indices vary greatly over the interval $[-0.75, 0.75]$ and become larger when there is either appreciation ($Z_{t-1} > 0$) or depreciation ($Z_{t-1} < 0$) of U.S. dollar. In particular, when the U.S. dollar experiences appreciation and during the "bad times" of the market (when $Z_{t-1} > 0$ and $\tau = 0.05$), domestic prices of commodity in Europe, Japan and China may increase, which pose risks on domestic companies. Then, all investors who invested corporations in the European, Japanese and Chinese markets suffer from

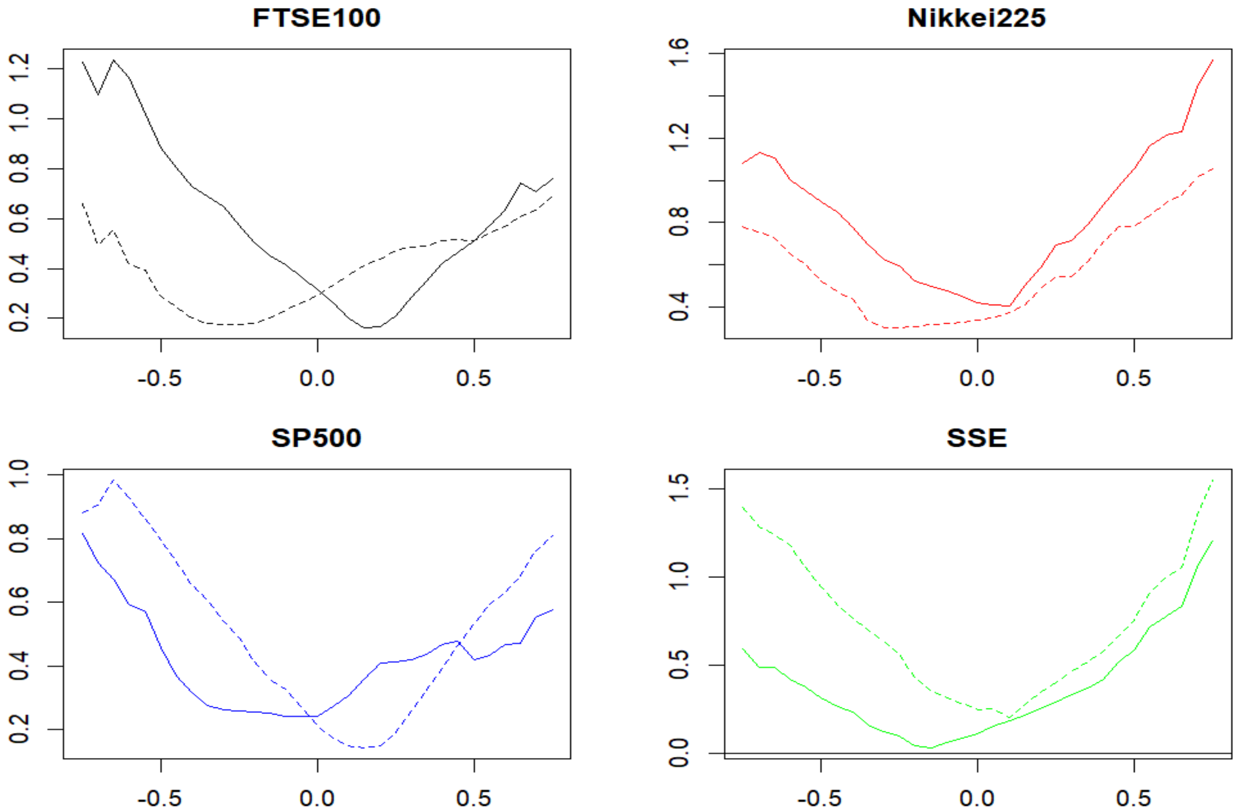

Figure 2: Connectedness with respect to outgoing links and connectedness with respect to incoming links for four market indices with $\tau = 0.05$. The solid line in each panel represents values of connectedness with respect to outgoing links and the dashed line in each panel indicates values of connectedness is for incoming link.

loss of returns, causing both CIL and COL to go up in all three markets. For the U.S. market, U.S. assets may become favorable among global investors during the U.S. dollar appreciation, while investors in the U.S. market who invested corporations in the rest of the world face loss of returns. These two forces lead the U.S. market to be both more important to the global market and to be more vulnerable to risk events in the global market, respectively. Thus, both curves in the panel of S&P 500 index increase.

As for the case when U.S. dollar depreciated ($Z_{t-1} < 0$), profits of investment on domestic corporations in European, Japanese and Chinese markets may increase, which encourage investors to give leverage in investing corporations in these three markets. Therefore, both types of curves in all three markets, as well as the CIL in the U.S. market increase. Nevertheless, global investors who invested assets in the U.S. market subject to adverse situation, which results in an upward movement of COL of S&P 500 index. In addition, due to the increase of leverage, European and Japanese markets can affect the global market more easily, causing COL to dominate CIL in these two markets and CIL to dominate COL in the U.S. market. In the Chinese market, corporations associated with export subject to harmful impact. Under this unfavorable environment, investors in China may be more willing to invest assets from outside of the Chinese market. This trend makes the Chinese market become more vulnerable to global risk events, causing CIL to dominate COL.

It is interesting that in the European and Japanese markets, during the U.S. dollar appreciation ($Z_{t-1} > 0.5$), the COL dominates CIL. These dynamic patterns in the European and Japanese markets may be explained by the so called "carry trade". The carry trade refers to borrowing a low-yielding asset and buying a higher-yielding foreign asset to earn the interest rate differential plus the expected foreign currency appreciation. Due to the relatively lower interest rate in the European and Japanese markets within our time span of study, as $Z_{t-1} > 0.5$, carry traders who borrowed low-yielding assets from the Japanese or European markets and bought assets from the U.S. market enjoy the increase of excess returns to carry trade. As a result, these two markets become less vulnerable to risk events caused by carry traders, which makes the CIL become smaller than COL in these two markets. While in the U.S. market, since the price of risky assets relies heavily on the demand of carry trade during U.S. dollar appreciation, it becomes much easier for the U.S. market to be affected by the global market. Therefore, the CIL dominates the COL in the U.S. market.

Figure 3 sheds light on the variation of risk spillover in the global financial market. Observed that in Figure 3, the total connectedness of all four market indices demonstrates an U-shaped pattern. It means that total risk spillover in the four major markets decreases when Z_{t-1} becomes larger within the interval $[-0.75, 0]$. As Z_{t-1} exceeds 0, the risk spillover intensity is magnified. In general, Figure 3 shows that the relationship between total risk spillover and the U.S. dollar change switches its pattern at a certain threshold of the U.S. dollar change, which is a relatively new result in literature.

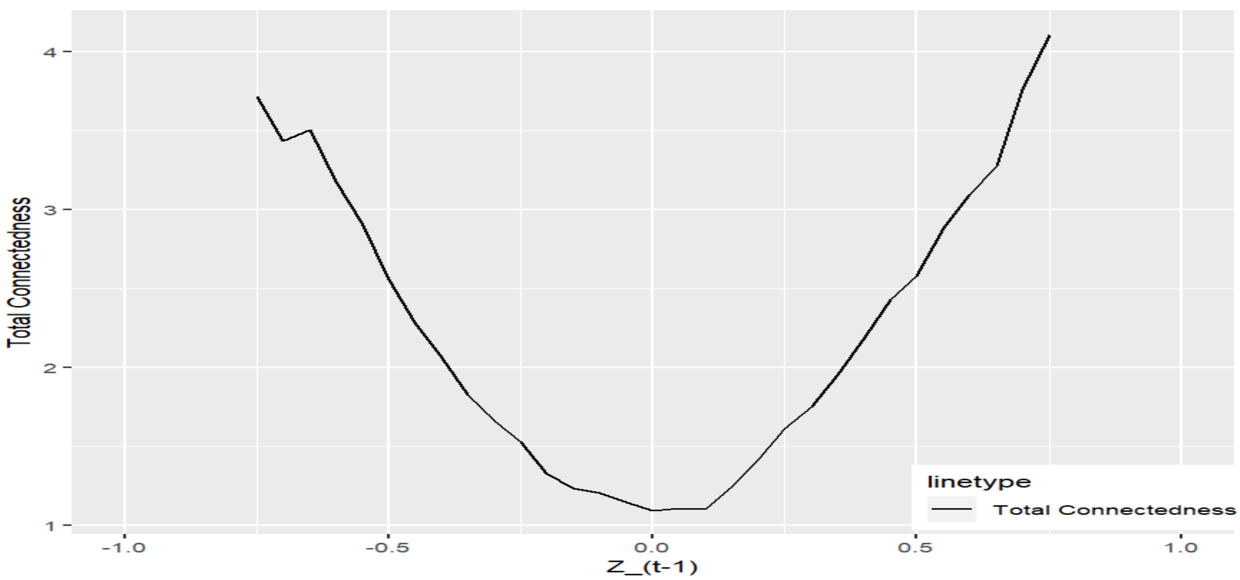

Figure 3: Total connectedness in international equity markets with $\tau = 0.05$.

5 Conclusion

In this paper, we investigate a functional coefficient VAR model for conditional quantiles, which is new to the literature. A three-stage procedure is proposed to estimate coefficients functionals and the properties of asymptotic normality for the proposed estimators are established. The simulation results show that our new estimation methods work fairly well. In addition, there is little literatures regarding the relationship between the variation of financial network and the general state of economy. Based on our approaches, the proposed framework allows to study how the network characteristics of risk spillover in a financial system vary with the state of economy.

There are several issues still worth of further studies. First, it is interesting to visualize the topological change of our financial network and to measure the transition of risk spillover among different market indices when the general economy is shifting. Technically, these studies can be realized by our econometric model. Second, the asymptotic properties of functional coefficients in our model provide solid theory to test the abnormal variation of financial network. Third, it is meaningful to allow for cross-sectional dependence in the current model. Although some methods have been developed to deal with cross-sectional dependence in the literature of conditional mean models, due to the nature of conditional quantile model, it is not obvious to extend these under the quantile setting. Finally, if Z_t in (2) is time, then the model in (2) provides a good start for studying conditional quantile estimation of ARCH- and GARCH-type models with time-varying parameters; see, for example, the papers by Dahlhaus and Subba Rao (2006) and Chen and Hong (2016) for the time-varying GARCH type models. We leave these important issues, together with some possible extensions as mentioned earlier in the paper, as future research topics.

Acknowledgments

We thank the Editor (Professor Atsushi Inoue), the Associate Editor, and two anonymous referees for their constructive and helpful comments and suggestions that improved significantly the quality of the paper. Also, This work was partially supported by the National Natural Science Foundation of China under Grant Numbers 72133002 and 72033008.

Disclosure Statement

The authors claim that there are no relevant financial or non-financial competing interests to report for this article. Also, the authors declare that they do not use any generative AI and AI-assisted technologies in the writing process.

References

Adrian, T., and Brunnermeier, M.K. (2016), CoVaR, *American Economic Review*, 106(7), 1705-1741.

Ando, T., and Bai, J. (2020), Quantile Co-Movement in Financial Markets: A Panel Quantile Model with Unobserved Heterogeneity, *Journal of the American Statistical Association*, 115(529), 266-279.

Belloni, A., and Chernozhukov, V. (2011), l_1-penalized Quantile Regression in High-dimensional Sparse Models, *Annals of Statistics*, 39(1), 82-130.

Billio, M., Getmansky, M., Lo., A.W., and Pelizzon, L. (2012), Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors, *Journal of Financial Economics*, 104(3), 535-559.

Cai, Z., Das, M., Xiong, H., and Wu, X. (2006), Functional Coefficient Instrumental Variables Models, *Journal of Econometrics*, 133(1), 207-241.

Cai, Z., Fan, J., and Yao, Q. (2000), Functional-Coefficient Regression Models for Nonlinear Time Series, *Journal of the American Statistical Association*, 95(451), 941-956.

Cai, Z., Juhl, T., and Yang, B. (2015), Functional Index Coefficient Models with Variable Selection, *Journal of Econometrics*, 189(1), 272-284.

Cai, Z., and Xu, X. (2008), Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models, *Journal of the American Statistical Association*, 103(484), 1595-1608.

Chen, B., and Hong, Y. (2016), Detecting for Smooth Structural Changes in GARCH Models, *Econometric Theory*, 32(3), 740-791.

Cheng, M.Y., Honda, T., and Zhang, J.T. (2016), Forward Variable Selection for Sparse Ultra-High Dimensional Varying Coefficient Models, *Journal of the American Statistical Association*, 111(515), 1209-1221.

Cheng, X., Han, X., and Inoue, A. (2022), Instrumental Estimation of Structural VAR Models Robust to Possible Nonstationarity, *Econometric Theory*, 38(5), 845-874.

Chernozhukov, V., Härdle, W.K., Huang, C., and Wang, W. (2021), LASSO-driven Inference in Time and Space, *Annals of Statistics*, 49(3), 1702-1735.

Dahlhaus, R., and Subba Rao., S. (2006), Statistical Inference for Time-varying ARCH Process, *Annals of Statistics*, 34(3), 1075-1114.

Diebold, F.X., and Yìlmaz, K. (2014), On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms, *Journal of Econometrics*, 182(1), 119-134.

Engle, R.F., and Manganelli, S. (2004), CaViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business & Economic Statistics*, 22(4), 367-381.

Fan, J., and Fan, Y. (2006), Issues on Quantile Autoregression, *Journal of the American Statistical Association*, 101(475), 991-994.

Fan, J., and Li, R. (2001), Variable Selection via Non-concave Penalized Likelihood and its Oracle Properties, *Journal of the American Statistical Association*, 96(456), 1348-1360.

Fan, Y., and Tang, C.Y. (2013), Tuning Parameter Selection in High Dimensional Penalized Likelihood, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 75(3), 531-552.

Gerlach, R., Chen, C.W.S., and Chan, N.Y.C. (2011), Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets, *Journal of Business* & *Economic Statistics*, 29(4), 481-492.

Härdle, W.K., Wang, W., and Yu, L. (2016), TENET: Tail-Event Driven NETwork Risk, *Journal of Econometrics*, 192(2), 499-513.

He, X., and Shi, P. (1996), Bivariate Tensor-Product B-Splines in a Partly Linear Model, *Journal of Multivariate Analysis*, 58, 162-181.

Koenker, R. (2005), *Quantile Regression*. Cambridge University Press, New York, NY.

Koenker, R., and Bassett, G.W. (1978), Regression Quantiles, *Econometrica*, 46(1), 33-50.

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017), *Handbook of Quantile Regression*, Chapman and Hall/CRC: Boca Raton, FL.

Koenker, R., and Xiao, Z. (2006), Quantile Autoregression, *Journal of the American Statistical Association*, 101(475), 980-990.

Liu, B.Y., Ji, Q., and Fan, Y. (2017), A New Time-Varying Optimal Copula Model Identifying the Dependence Across Markets, *Quantitative Finance*, 17(1), 1-17.

Menkhoff, L., Sarno, L., Schmelling, M., and Schrimpf, A. (2012), Carry Trades and Global Foreign Exchange Volatility, *Journal of Finance*, 67(2), 681-718.

Schumaker, L.L. (1981), *Spline Functions: Basic Theory*, Wiley: New York, NY.

Taylor, S. (1986), *Modeling Financial Time Series*, Wiley: New York, NY.

Tang, Y., Song, X., Wang, H.J., and Zhu, Z. (2013), Variable Selection in High-Dimensional Quantile Varying Coefficient Models, *Journal of Multivariate Analysis*, 122(1), 115-132.

Tsay, R.S. (1998), Testing and Modeling Multivariate Threshold Models, *Journal of the American Statistical Association*, 93(433), 1188-1202.

White, H., Kim, T.H., and Manganelli, S. (2008), Modeling Autoregressive Conditional Skewness and Kurtosis With Multi-quantile CAViaR, In: Russell, J., Watson, M. (Eds.), *Volatility and Time Series Econometrics*: A Festschrift in Honor of Robert F. Engle.

White, H., Kim, T.H., and Manganelli, S. (2015), VAR for VaR: Measuring Tail Dependence Using Multivariate Regression Quantiles, *Journal of Econometrics*, 187(1), 169-188.

Xiao, Z., and Koenker, R. (2009), Conditional Quantile Estimation for Generalized Autoregressive Conditional Heteroscedasticity Models, *Journal of the American Statistical Association*, 104(488), 1696-1712.

Xu, X., Wang, W., Shin, Y., and Zheng, C. (2022), Dynamic Network Quantile Regression Model, *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2022.2093882.

Yang, Z., and Zhou, Y. (2017), Quantitative Easing and Volatility Spillovers Across Countries and Asset Classes, *Management Science*, 63(2), 333-354.

Zhang, C.H. (2010), Nearly Unbiased Variable Selection Under Minimax Concave Penalty, *Annals of Statistics*, 38(2), 894-942.

Zheng, Q., Peng, L., and He, X. (2015), Globally Adaptive Quantile Regression with Ultra-high Dimensional Data, *Annals of Statistics*, 43(5), 2225.

Appendix to “A Functional-Coefficient VAR Model for Dynamic Quantiles and Its Application to Constructing Nonparametric Financial Network”

Appendix

Appendix A: Notation and Assumptions

Notation

Throughout this article, $0_{a\times b}$ stands for the $(a\times b)$ matrix of zeros and I_a is the $(a\times a)$ identity matrix. For a vector $v=(v_1,\ldots,v_p)^\top$, let $\|v\|_\infty\equiv\max_{1\le j\le p}|v_j|$ and $\|v\|_s\equiv(\sum_{j=1}^p|v_j|^s)^{1/s}$, $s\ge 1$. Specifically, let $\|v\|_2\equiv\|v\|$ be the Euclidean norm. For a set of vectors $\{\boldsymbol{v}_l\}_{l=1}^m$, let $\bigotimes_{l=1}^m\boldsymbol{v}_l\equiv\boldsymbol{v}_1\otimes\cdots\otimes\boldsymbol{v}_m$. Given two sequences of positive numbers a_n and b_n, write $a_n\lesssim b_n$ if there exists constant $C>0$ (does not depend on n) such that $a_n/b_n\le C$. For a sequence of random variables x_n, we use the notation $x_n\lesssim_p b_n$ to denote $x_n=O_p(b_n)$. For a set A, we denote card(A) as the number of elements contained in A. For any finitely discrete measure $\mathcal{Q}$ on a measurable space, let $\mathcal{L}^q(\mathcal{Q})$ denote the space of all measurable functions $f:\mathcal{X}\to\mathbb{R}$ such that $\|f\|_{\mathcal{Q},q}\equiv(\mathcal{Q}|f|^q)^{1/q}<\infty$, where $\mathcal{Q}f\equiv\int f d\mathcal{Q}$. For a class of measurable functions $\mathcal{F}$, the ϵ-covering number with respect to the $\mathcal{L}^q(\mathcal{Q})$-semi-metric is denoted as $\mathcal{N}(\epsilon,\mathcal{F},\|\cdot\|_{\mathcal{Q},q})$, and let $\text{ent}(\epsilon,\mathcal{F})=\log\sup_{\mathcal{Q}}\mathcal{N}(\epsilon\|\bar{F}\|_{\mathcal{Q},q},\mathcal{F},\|\cdot\|_{\mathcal{Q},q})$ with $\bar{F}=\sup_{f\in\mathcal{F}}|f|$ (the envelope) denote the uniform entropy number. Detailed discussions about the uniform entropy number can be found in the Section 2.6 of Van Der Vaart and Wellner (1996).

Assumptions

First, some necessary conditions are provided for the theoretical proofs of Theorems 1, 2 and 3 as well as the consistency of the estimated covariance matrix, together with some discussions on the assumptions. Note that the same notation is used as in the main text.

Assumption A.

A1: *For all* $z \in [a,b]$*, let* $\mathcal{A}_\tau(x) = \mathbf{\Gamma}_{\beta,1,\tau}(z)x$ *and* $\mathcal{B}_\tau(x) = I_\kappa - \mathbf{\Gamma}_{1,\tau}(z)x$*, where* $\mathbf{\Gamma}_{\beta,1,\tau}(z)$ *and* $\mathbf{\Gamma}_{1,\tau}(z)$ *are defined in (4) of the main text. Suppose that* $\mathcal{A}_\tau(x)$ *and* $\mathcal{B}_\tau(x)$ *have no common factors so that* $\mathcal{A}_\tau(x) \neq 0$*, for* $|x| \leq 1$ *and* $\mathcal{B}_\tau(x) \neq 0$*, for* $|x| \leq 1$*.*

A2: *For* $d \geq 2$ *defined in Section 2.3 of the main text,* $K = O(n^{c_1})$ *and* $R_n = O(n^{c_2})$ *for some* $c_2 < 1/3$ *and* $\frac{3c_2}{2(1+d)} < c_1 \leq c_2$*. In addition, the number of nonzero components* r_n *satisfies* $r_n = O(n^{c_2 - c_1})$*.*

A3: *There are vectors* $\boldsymbol{c}_{0,i,\tau} \in \mathbb{R}^K$*,* $\boldsymbol{c}_{\gamma,ij,\tau} \in \mathbb{R}^K$ *and* $\boldsymbol{c}_{\beta,ij,\tau} \in \mathbb{R}^K$ *such that* $\sup_{z\in[a,b]} |\gamma_{i0,\tau}(z) - b^{K\top}(z)\boldsymbol{c}_{0,i,\tau}| = O(K^{-d})$*,* $\sup_{z\in[a,b]} |\gamma_{ij,\tau}(z) - b^{K\top}(z)\boldsymbol{c}_{\gamma,ij,\tau}| = O(K^{-d})$ *and* $\sup_{z\in[a,b]} |\beta_{ij,\tau}(z) - b^{K\top}(z)\boldsymbol{c}_{\beta,ij,\tau}| = O(K^{-d})$*, for* $1 \leq i,j \leq \kappa$*.*

A4: *For* $1 \leq l \leq r_n$ *and* $0 \leq j \leq \kappa$*,* $\alpha_{lj,\tau}(\cdot) \in \mathcal{H}_l$*, where* $\mathcal{H}_l$ *is defined in Section 2.3 of the main text. Each entry in the vector* $\boldsymbol{g}_\tau(\cdot)$ *is* $(\varsigma+1)$*th order continuously differentiable in a neighborhood of* z_0 *for any* $z_0 \in [a,b]$*. Here,* ς *is defined in Section 2.2 of the main text.*

A5: $f_z(z)$ *is a continuously marginal density of* Z *and* $f_z(z_0) > 0$*.*

A6: *The distribution of* Y *given* Z *and* $\boldsymbol{W}$ *has an everywhere positive conditional density* $f_{Y|Z,\boldsymbol{W}}(\cdot)$*, which is bounded from below by* $\underline{f}$ *and above by* $\bar{f}$*, and satisfies the Lipschitz continuity condition. Here,* $\boldsymbol{W}_t$ *is defined in (10) of the main text. In addition,* $\left|f^{(1)}_{Y|Z,\boldsymbol{W}}(\cdot)\right| \leq \bar{f}'$*, where* $f^{(1)}_{Y|Z,\boldsymbol{W}}(\cdot)$ *is the first derivative of* $f_{Y|Z,\boldsymbol{W}}(\cdot)$*. Finally, the kernel function* $K(\cdot)$ *is a bounded, symmetric density with a bounded support region. Let* $\mu_2 = \int \nu^2 K(\nu)d\nu$ *and* $\nu_0 = \int K^2(\nu)d\nu$*.*

A7: $\{(Y_{it}, Z_t)\}$ *in model (2) is a strictly stationary sequence with* α*-mixing coefficient* $\alpha(t)$ *which satisfies* $\sum_{t=1}^{\infty} t^{\iota}\alpha^{(\delta-2)/\delta}(t) < \infty$ *for some positive real number* $\delta > 2$ *and* $\iota > (\delta-2)/\delta$*.*

A8: *There exist (small) positive constants* $\varpi_1 > 0$ *and* $\varpi_2 > 0$ *such that* $P\{\max_{1\leq t\leq n} Y_t^2 >$

$n^{\varpi_1}\} \leq \exp(-n^{\varpi_2})$.

A9: *$E[\boldsymbol{W}_t\boldsymbol{W}_t^\top|Z_t = z_0]$ and $E[\boldsymbol{W}_t\boldsymbol{W}_t^\top f_{Y|Z,\boldsymbol{W}}(q_\tau(z_0, \boldsymbol{W}_t))|Z_t = z_0]$ is positive-definite and continuous in a neighborhood of z_0. In addition, $E\|\mathbb{Y}_t\|^{2\delta^*} < \infty$ with $\delta^* > \delta$. Finally, there exists postive constants C_1, C_2, c_1 and c_2, such that $\tilde{\varphi}(\mathfrak{r}) \leq C_1$, $\varphi(\mathfrak{r}) \leq C_2$, $\varrho(\mathfrak{r}) \geq c_1$ and $\vartheta(\mathfrak{r}) \geq c_2$ for any $\mathfrak{r} \leq n$.*

A10: *The class of function $\mathcal{F}_c = \{v \mapsto \psi_\tau(v, \boldsymbol{c}), \|\boldsymbol{c}\|_0 \leq \mathfrak{r}, 1 \leq \mathfrak{r} \leq n\}$ is pointwise measurable and satisfies the entropy condition $\mathrm{ent}(\epsilon, \mathcal{F}_c) \leq C\mathfrak{r}\log((n \vee M_n)/\epsilon)$ for some constant $C > 0$ and for all $0 < \epsilon \leq 1$. In addition, for some $\nu \geq 0$, $\max_{f\in\mathcal{F}_c} \|f(v_\cdot)\|_{\psi_\nu,0} < \infty$ and $\|\psi^0_{\tau,\cdot}\|_{\psi_\nu,0} < \infty$. The map $\boldsymbol{c} \mapsto E\{\psi_\tau(V_t, \boldsymbol{c})|\mathcal{F}_t\}$ is twice continuously differentiable and $\Phi^{\boldsymbol{c}}_{\psi_\nu,0} \equiv \|\max_{\boldsymbol{c}\in\mathcal{C}} \partial_{\boldsymbol{c}} E\{\psi_\tau(V_\cdot, \boldsymbol{c})|\mathcal{F}_\cdot\}\|_{\psi_\nu,0} < \infty$, for some fixed and closed interval $\mathcal{C}$.*

A11: *$\lambda_{n,0}R_n^{1/2}n^{-1} \to 0$, $n^{-1/2}\lambda_{n,1} \to 0$ and $R_n^{-1/2}\lambda_{n,1} \to \infty$. The bandwidth h satisfies $h = O(n^{-1/5})$, $h \to 0$, $nh \to \infty$.*

A12: *$f(\boldsymbol{w}, \boldsymbol{\omega}|\boldsymbol{Y}_0, \boldsymbol{Y}_\ell; \ell) \leq H < \infty$ for $\ell \geq 1$, where $f(\boldsymbol{w}, \boldsymbol{\omega}|\boldsymbol{Y}_0, \boldsymbol{Y}_\ell; \ell)$ is the conditional density of (Z_0, Z_ℓ) given $(\mathbb{Y}_0 = \boldsymbol{Y}_0, \mathbb{Y}_\ell = \boldsymbol{Y}_\ell)$.*

A13: *$n^{1/2-\delta/4}h^{\delta/\delta^*-1/2-\delta/4} = O(1)$.*

Remark A.1. *Assumptions A1 is a condition for the functional coefficients to be well-defined, which is similar to that in Chen and Hong (2016). The assumption on R_n in A2 is for the minimum signal strength of the coefficients in the true active set, which is also used in Sherwood and Wang (2016). Assumptions A3-A6 are common in literature of spline approximation and nonparametric estimation. A7 is a standard assumption for α-mixing. A8 can be implied when the maximum of Y_t^2 follows a generalized extreme value distribution, which is generally satisfied for weakly dependent data; see also Xiao and Koenker (2009). The first and second parts of A9 is commonly required for the model identification, when $\boldsymbol{W}_t$ is α-mixing. By Lemma 7 of Tang et al. (2013), the last part of A9 can be satisfied under the construction of tensor-product B-spline bases proposed in Section 2.2. The first part of A10 is adopted from Chernozhukov et al. (2021), which requires $\psi_\tau(v, \boldsymbol{c})$ not to increase entropy too much. The finite moment conditions* *in the second and the last part of A10 can be implied by some primitive assumptions provided in Chernozhukov et al. (2021). The divergence of* $\lambda_{n,0}$ *and* $\lambda_{n,1}$ *in A11 are necessary to derive Theorems 1 and 2, which are similar to the setting in Tang et al. (2013). A12 is very standard and used for the proof under mixing conditions. A13 allows one to verify standard Lindeberg-Feller conditions for asymptotic normality of the proposed estimators in the proof of Theorem 3; see Cai and Xu (2008) for details on nonparametric quantile regressions models for* α*-mixing time series.*

Appendix B: Mathematical Proofs of Theorems 1 and 2

In this section, we give certain lemmas with their detailed proofs that are useful for proving Theorems 1 and 2 in the paper. Of course, notations and assumptions that are used here are the same as those in the main article. Also note that C and M are denoted as generic constants that may vary across occurrences. Recall that $A^{\mathfrak{r}}_{M_n} = \{\boldsymbol{\delta} \in \mathbb{R}^{M_n} : \|\boldsymbol{\delta}\|_2 = 1, \|\boldsymbol{\delta}\|_0 \leq \mathfrak{r}\}$ is the $\mathfrak{r}$-sparse unit sphere in $\mathbb{R}^{M_n}$ and we use the notation E_t to represent the conditional expectation $E\{\cdot|Z, \boldsymbol{W}\} \equiv E\{\cdot|\mathcal{F}_{t-1}\}$. In proofs of all lemmas and theorems, τ is dropped from $\boldsymbol{c}_\tau$ for simplicity.

B.1 Some Lemmas

Lemma B.1. *Let $\hat{\beta}$ be the minimizer of the function $\sum_{t=1}^n \rho_\tau(Y_t - X_t^\top \beta)$. Then, $\| \sum_{t=1}^n X_t \psi_\tau(Y_t - X_t^\top \hat{\beta})\| \leq \dim(X) \max_{t \leq n} \|X_t\|$.*

Proof. The proof follows from that of Lemma A.2 in Ruppert and Carroll (1980). □

Lemma B.2. *Suppose Assumption A1–A13 hold. Let $Q_\tau(\boldsymbol{c}) = E[\rho_\tau\{Y_t - \boldsymbol{\Pi}_t^\top \boldsymbol{c}\}]$. For each $\boldsymbol{c} \in \mathbb{R}^{M_n}$, satisfying $\|\boldsymbol{c} - \bar{\boldsymbol{c}}\|_2 = \boldsymbol{\zeta}_n \geq C r_n K^{-d}$ and $\|\boldsymbol{c} - \bar{\boldsymbol{c}}\|_0 \leq \breve{\mathfrak{m}}$, we have*

$$Q_\tau(\boldsymbol{c}) - Q_\tau(\bar{\boldsymbol{c}}) \geq C \mathfrak{q}(\breve{\mathfrak{m}}) (\boldsymbol{\zeta}_n^2 \wedge \boldsymbol{\zeta}_n)$$

for some constant $C > 0$.

Proof. Using Knight's identity as in Knight (1998), the law of iterated expectations and mean value theorem, we have, for $\tilde{z} \in [0, \mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})]$,

$$
\begin{aligned}
&Q_\tau(\boldsymbol{c}) - Q_\tau(\bar{\boldsymbol{c}}) - E[-\mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})\psi_\tau(Y_t - \mathbf{\Pi}_t^\top\bar{\boldsymbol{c}})] \\
=&E\left[\int_0^{\mathbf{\Pi}_t^\top(\boldsymbol{c}-\bar{\boldsymbol{c}})} F_{Y|Z,\boldsymbol{W}}(\mathbf{\Pi}_t^\top\bar{\boldsymbol{c}} + z|Z, \boldsymbol{W}) - F_{Y|Z,\boldsymbol{W}}(\mathbf{\Pi}_t^\top\bar{\boldsymbol{c}}|Z, \boldsymbol{W})dz\right] \\
=&E\left[\int_0^{\mathbf{\Pi}_t^\top(\boldsymbol{c}-\bar{\boldsymbol{c}})} z f_{Y|Z,\boldsymbol{W}}(\mathbf{\Pi}_t^\top\bar{\boldsymbol{c}}|Z, \boldsymbol{W}) + \frac{z^2}{2} f_{Y|Z,\boldsymbol{W}}^{(1)}(\mathbf{\Pi}_t^\top\bar{\boldsymbol{c}} + \tilde{z}|Z, \boldsymbol{W})dz\right] \qquad \text{(B.1)} \\
\geq&E\left[\frac{1}{2}\{\mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})\}^2 f_{Y|Z,\boldsymbol{W}}(\mathbf{\Pi}_t^\top\bar{\boldsymbol{c}}|Z, \boldsymbol{W})\right] - \frac{\bar{f}'}{6}E\left[|\mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})|^3\right] \\
\geq&\frac{\underline{f}}{2}E[(\mathbf{\Pi}_t^\top\boldsymbol{\delta})^2]\boldsymbol{\zeta}_n^2 - \frac{\bar{f}'}{6}E[|\mathbf{\Pi}_t^\top\boldsymbol{\delta}|^3]\boldsymbol{\zeta}_n^3,
\end{aligned}
$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{c}-\bar{\boldsymbol{c}}}{\boldsymbol{\zeta}_n} \in A_{M_n}^{\breve{\mathfrak{m}}}$ and $\bar{f}'$ is defined in Assumption A6. Meanwhile, by Assumption A3, there exists a constant $C > 0$ such that $\sup_{\boldsymbol{Z}_t \in [a,b]^{r_n(r_n+1)/2}} |\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t - \mathbf{\Pi}_t^\top\bar{\boldsymbol{c}}| \leq \sum_{l=1}^{r_n} \sup_{\boldsymbol{z}_l \in \boldsymbol{\mathcal{Z}}_l} |\alpha_{lj,\tau}(\boldsymbol{z}_l) - B_{lt}^\top\bar{\boldsymbol{c}}_{lj,\tau}| \leq C r_n K^{-d}$. Then, we have

$$
\begin{aligned}
\left|E[-\mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})\psi_\tau(Y_t - \mathbf{\Pi}_t^\top\bar{\boldsymbol{c}})]\right| =&\left|E\left\{\mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})E[\psi_\tau(Y_t - \mathbf{\Pi}_t^\top\bar{\boldsymbol{c}})|Z, \mathbb{W}]\right\}\right| \\
\leq&\bar{f}Cr_nK^{-d}\boldsymbol{\zeta}_n\left|E[\mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})/\boldsymbol{\zeta}_n]\right| \qquad \text{(B.2)} \\
\leq&C\bar{f}\boldsymbol{\zeta}_n^2\left|E[\mathbf{\Pi}_t^\top(\boldsymbol{c} - \bar{\boldsymbol{c}})/\boldsymbol{\zeta}_n]\right| \leq C\bar{f}\boldsymbol{\zeta}_n^2\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})}.
\end{aligned}
$$

By Assumption A9, $\tilde{\varphi}(\breve{\mathfrak{m}})$ is bounded from above. Thus, combining (B.1) and (B.2) yields

$$
\begin{aligned}
Q_\tau(\boldsymbol{c}) - Q_\tau(\bar{\boldsymbol{c}}) \geq&\frac{\underline{f}}{2}E[(\mathbf{\Pi}_t^\top\boldsymbol{\delta})^2]\boldsymbol{\zeta}_n^2 - \frac{\bar{f}'}{6}E[|\mathbf{\Pi}_t^\top\boldsymbol{\delta}|^3]\boldsymbol{\zeta}_n^3 - C\bar{f}\boldsymbol{\zeta}_n^2\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})} \\
\geq&\frac{\underline{f}}{3}E[(\mathbf{\Pi}_t^\top\boldsymbol{\delta})^2]\boldsymbol{\zeta}_n^2 - \frac{\bar{f}'}{6}E[|\mathbf{\Pi}_t^\top\boldsymbol{\delta}|^3]\boldsymbol{\zeta}_n^3 \qquad \text{(B.3)} \\
=&\frac{\underline{f}}{4}E[(\mathbf{\Pi}_t^\top\boldsymbol{\delta})^2]\boldsymbol{\zeta}_n^2 + \frac{\underline{f}}{12}E[(\mathbf{\Pi}_t^\top\boldsymbol{\delta})^2]\boldsymbol{\zeta}_n^2 - \frac{\bar{f}'}{6}E[|\mathbf{\Pi}_t^\top\boldsymbol{\delta}|^3]\boldsymbol{\zeta}_n^3.
\end{aligned}
$$

Next, define

$$
\boldsymbol{\zeta}_{\breve{\mathfrak{m}}} = \sup\left\{\boldsymbol{\zeta} : Q_\tau(\bar{\boldsymbol{c}} + \boldsymbol{\zeta}\boldsymbol{d}) - Q_\tau(\bar{\boldsymbol{c}}) \geq \frac{\underline{f}}{4}\boldsymbol{\zeta}^2 E[(\mathbf{\Pi}_t^\top\boldsymbol{d})^2], \text{ for all } \boldsymbol{d} \in A_{M_n}^{\breve{\mathfrak{m}}}\right\}.
$$

By construction of $\boldsymbol{\zeta}_{\breve{\mathfrak{m}}}$ and the convexity of $Q_\tau(\cdot)$, for any $\boldsymbol{c}$ such that $\|\boldsymbol{c}-\bar{\boldsymbol{c}}\|_0 \le \breve{\mathfrak{m}}$, we have

$$Q_\tau(\boldsymbol{c}) - Q_\tau(\bar{\boldsymbol{c}}) \ge \frac{\underline{f}}{4} E[(\boldsymbol{\Pi}_t^\top(\boldsymbol{c}-\bar{\boldsymbol{c}}))^2] \wedge \left\{ \boldsymbol{\zeta}_n \frac{\inf_{\boldsymbol{d}\in A_{M_n}^{\breve{\mathfrak{m}}}} Q_\tau(\bar{\boldsymbol{c}} + \boldsymbol{\zeta}_{\breve{\mathfrak{m}}}\boldsymbol{d}) - Q_\tau(\bar{\boldsymbol{c}})}{\boldsymbol{\zeta}_{\breve{\mathfrak{m}}}} \right\}.$$

Since

$$\boldsymbol{\zeta}_n \frac{\inf_{\boldsymbol{d}\in A_{M_n}^{\breve{\mathfrak{m}}}} Q_\tau(\bar{\boldsymbol{c}} + \boldsymbol{\zeta}_{\breve{\mathfrak{m}}}\boldsymbol{d}) - Q_\tau(\bar{\boldsymbol{c}})}{\boldsymbol{\zeta}_{\breve{\mathfrak{m}}}} \ge \boldsymbol{\zeta}_n \frac{\underline{f}\varrho(\breve{\mathfrak{m}})\boldsymbol{\zeta}_{\breve{\mathfrak{m}}}}{4}$$

and

$$\frac{\underline{f}}{4} E[(\boldsymbol{\Pi}_t^\top(\boldsymbol{c}-\bar{\boldsymbol{c}}))^2] \ge \boldsymbol{\zeta}_n^2 \frac{\underline{f}\varrho(\breve{\mathfrak{m}})}{4},$$

we have

$$Q_\tau(\boldsymbol{c}) - Q_\tau(\bar{\boldsymbol{c}}) \ge \boldsymbol{\zeta}_n \frac{\underline{f}\varrho(\breve{\mathfrak{m}})\boldsymbol{\zeta}_{\breve{\mathfrak{m}}}}{4} \wedge \boldsymbol{\zeta}_n^2 \frac{\underline{f}\varrho(\breve{\mathfrak{m}})}{4}. \tag{B.4}$$

Note that for any $\boldsymbol{\zeta}$, if

$$\boldsymbol{\zeta} \ge \frac{\underline{f}}{2\bar{f}'} \inf_{\boldsymbol{\delta}\in A_{M_n}^{\breve{\mathfrak{m}}}} \frac{E[|\boldsymbol{\Pi}_t^\top \boldsymbol{\delta}|^2]}{E[|\boldsymbol{\Pi}_t^\top \boldsymbol{\delta}|^3]} = \frac{\underline{f}}{2\bar{f}'} \vartheta(\breve{\mathfrak{m}}),$$

it follows that

$$\frac{\underline{f}}{12} E[(\boldsymbol{\Pi}_t^\top \boldsymbol{\delta})^2]\boldsymbol{\zeta}_n^2 - \frac{\bar{f}'}{6} E[|\boldsymbol{\Pi}_t^\top \boldsymbol{\delta}|^3]\boldsymbol{\zeta}_n^3 > 0.$$

Then, by (B.3) and the definition of $\boldsymbol{\zeta}_{\breve{\mathfrak{m}}}$, we have $\boldsymbol{\zeta}_{\breve{\mathfrak{m}}} \ge \boldsymbol{\zeta} \ge \frac{\underline{f}}{2\bar{f}'}\vartheta(\breve{\mathfrak{m}})$. This, in conjunction with (B.4), implies that

$$Q_\tau(\boldsymbol{c}) - Q_\tau(\bar{\boldsymbol{c}}) \ge C_1 \frac{\underline{f}\varrho(\breve{\mathfrak{m}})}{4} \left\{ 1 \wedge \frac{\underline{f}}{2\bar{f}'}\vartheta(\breve{\mathfrak{m}}) \right\} (\boldsymbol{\zeta}_n^2 \wedge \boldsymbol{\zeta}_n) = C_1 \mathfrak{q}(\breve{\mathfrak{m}}) (\boldsymbol{\zeta}_n^2 \wedge \boldsymbol{\zeta}_n).$$

These complete the proof of Lemma B.2. □

To obtain an upper bound for $\breve{\mathfrak{m}} = \|\tilde{\boldsymbol{c}}_\tau\|_0$, we focus on the following optimization problem, which is the dual problem of the linear programming problem of (10) in the main text:

$$\begin{aligned}&\max_{a\in\mathbb{R}^n} n^{-1}\sum_{t=1}^{n} Y_t a_t\\ &\text{s.t. } \left|n^{-1}\sum_{t=1}^{n} P_{l\mathfrak{u}t}a_t\right| \le \frac{\lambda_{n,0}}{n},\ \ l=1,\ldots,m,\ \mathfrak{u}=1,\ \ldots,(1+\kappa)K^l,\\ &(\tau-1)\le a_t\le\tau,\ t=1,\ldots,n,\end{aligned} \tag{B.5}$$

where $a=(a_1,\ldots,a_n)^\top$ and $P_{l\mathfrak{u}t}$ is the $\mathfrak{u}$th element of $\boldsymbol{P}_{lt}$.

Lemma B.3. *Suppose Assumption A1–A13 hold. The number* $\breve{\mathfrak{m}}=\|\tilde{\boldsymbol{c}}_\tau\|_0$ *of nonzero components in* $\tilde{\boldsymbol{c}}_\tau$ *satisfies*

$$\breve{\mathfrak{m}}=\|\tilde{\boldsymbol{c}}_\tau\|_0\le n\wedge M_n\wedge\frac{n^2\tilde{\varphi}(\breve{\mathfrak{m}})}{\lambda_{n,0}^2}.$$

Suppose that $Y_1,\ldots,Y_n$ *are absolutely continuous conditional on* $\boldsymbol{W}_1,\ldots,\boldsymbol{W}_n$*,* $Z_1,\ldots,Z_n$*, then the number of interpolated points,* $\mathrm{card}(\{t:Y_t=\boldsymbol{\Pi}_t^\top\tilde{\boldsymbol{c}}\})$ *is equal to* $\breve{\mathfrak{m}}$ *with probability approaching to* 1.

Proof. The proof follows directly from that of Lemma 6 in Belloni and Chernozhukov (2011). □

Based on the rough upper bound of $\breve{\mathfrak{m}}=\|\tilde{\boldsymbol{c}}_\tau\|_0$ in Lemma B.3, one can further refine the upper bound of $\breve{\mathfrak{m}}$. Indeed, by the complementary slackness condition of linear programming in Theorem 4.5 of Bertsimas and Tsitsiklis (1997), we have

$$\tilde{c}_{l\mathfrak{u}}>0 \text{ iff } n^{-1}\sum_{t=1}^{n}P_{l\mathfrak{u}t}\tilde{a}_{t,\tau}=\frac{\lambda_{n,0}}{n},$$

and

$$\tilde{c}_{l\mathfrak{u}}<0 \text{ iff } n^{-1}\sum_{t=1}^{n}P_{l\mathfrak{u}t}\tilde{a}_{t,\tau}=-\frac{\lambda_{n,0}}{n}, \tag{B.6}$$

for $1\le l\le m$ and $1\le\mathfrak{u}\le(1+\kappa)K^l$, where $P_{l\mathfrak{u}t}$ is the $\mathfrak{u}$th component of $\boldsymbol{P}_{lt}$ and $\tilde{a}_\tau=(\tilde{a}_{1,\tau},\ldots,\tilde{a}_{n,\tau})^\top$ solves the dual problem (B.5).

Let $\psi_\tau(V_t,\boldsymbol{c})\equiv\{I(Y_t\le\boldsymbol{\Pi}_t^\top\boldsymbol{c})-\tau\}\boldsymbol{\Pi}_t$ be the score function of the tth observation, where

$V_t \equiv \{Y_t, \boldsymbol{\Pi}_t\}$. Similarly, let $\psi_\tau(V_t, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t)) \equiv \{I(Y_t \leq \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t) - \tau\}\boldsymbol{\Pi}_t$, where $\boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t) = (\boldsymbol{\alpha}_{1,1,\tau}^\top(\boldsymbol{Z}_{t,1}), \ldots, \boldsymbol{\alpha}_{1,m,\tau}^\top(\boldsymbol{Z}_{t,m}))^\top$ and $\mathbb{W}_t = (\boldsymbol{W}_{t-1}^\top, \ldots, \boldsymbol{W}_{t-m}^\top)^\top$. Then, define

$$S_n(\boldsymbol{c}) \equiv n^{-1}\sum_{t=1}^{n}\psi_\tau(V_t, \boldsymbol{c}) = n^{-1}\sum_{t=1}^{n}\{I(Y_t \leq \boldsymbol{\Pi}_t^\top \boldsymbol{c}) - \tau\}\boldsymbol{\Pi}_t$$

and

$$S^0 \equiv n^{-1}\sum_{t=1}^{n}\psi_\tau(V_t, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t)) = n^{-1}\sum_{t=1}^{n}\{I(Y_t \leq \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t) - \tau\}\boldsymbol{\Pi}_t.$$

In addition, define the set of $\breve{\mathfrak{m}}$-sparse vectors near $\bar{\boldsymbol{c}}$ as $R(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}) = \{\boldsymbol{c} \in \mathbb{R}^{M_n} : \|\boldsymbol{c}\|_0 \leq \breve{\mathfrak{m}}, Cr_nK^{-d} \leq \|\boldsymbol{c}-\bar{\boldsymbol{c}}\|_2 \leq \tilde{\boldsymbol{\zeta}}_n\}$ and the sparse sphere associated with a given vector $\boldsymbol{c} = (c_{\mathfrak{s}})_{1\leq\mathfrak{s}\leq M_n}$ as $\boldsymbol{S}(\boldsymbol{c}) = \{\boldsymbol{\delta} \in \mathbb{R}^{M_n} : \|\boldsymbol{\delta}\|_2 \leq 1, \text{support}(\boldsymbol{\delta}) \subseteq \text{support}(\boldsymbol{c})\}$, where $\text{support}(\boldsymbol{\delta}) = \{\jmath : \delta_\jmath \neq 0, 1 \leq \jmath \leq M_n, \boldsymbol{\delta} = (\delta_\jmath)_{1\leq\jmath\leq M_n}\}$ and $\text{support}(\boldsymbol{c}) = \{\mathfrak{s} : c_{\mathfrak{s}} \neq 0, 1 \leq \mathfrak{s} \leq M_n, \boldsymbol{c} = (c_{\mathfrak{s}})_{1\leq\mathfrak{s}\leq M_n}\}$. Also, define

$$\varepsilon_0(\breve{\mathfrak{m}}, n, M_n) \equiv \sup_{\boldsymbol{\delta}\in A_{M_n}^{\breve{\mathfrak{m}}}} n^{1/2}\big|\boldsymbol{\delta}^\top[S^0 - E\{S^0\}]\big|,$$

$$\varepsilon_1(\breve{\mathfrak{m}}, n, M_n) \equiv \sup_{\boldsymbol{\delta}\in A_{M_n}^{\breve{\mathfrak{m}}}} n^{1/2}\big|\boldsymbol{\delta}^\top[S_n(\bar{\boldsymbol{c}}) - E\{S_n(\bar{\boldsymbol{c}})\} - (S^0 - E\{S^0\})]\big|,$$

$$\varepsilon_2(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}, n, M_n) \equiv \sup_{\boldsymbol{c}\in R(\tilde{\boldsymbol{\zeta}}_n,\breve{\mathfrak{m}}),\boldsymbol{\delta}\in\boldsymbol{S}(\boldsymbol{c})} n^{1/2}\big|\boldsymbol{\delta}^\top[S_n(\boldsymbol{c}) - E\{S_n(\boldsymbol{c})\} - S_n(\bar{\boldsymbol{c}}) + E\{S_n(\bar{\boldsymbol{c}})\}]\big|,$$

and

$$\varepsilon_3(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}, n, M_n) \equiv \sup_{\boldsymbol{c}\in R(\tilde{\boldsymbol{\zeta}}_n,\breve{\mathfrak{m}}),\boldsymbol{\delta}\in\boldsymbol{S}(\boldsymbol{c})} n^{1/2}\big|\boldsymbol{\delta}^\top[E\{S_n(\boldsymbol{c})\} - E\{S_n(\bar{\boldsymbol{c}})\}]\big|,$$

where $A_{M_n}^{\breve{\mathfrak{m}}} \equiv \{\boldsymbol{\delta} \in \mathbb{R}^{M_n} : \|\boldsymbol{\delta}\|_2 = 1, \|\boldsymbol{\delta}\|_0 \leq \breve{\mathfrak{m}}\}$.

Lemma B.4. *Suppose that Assumption A1–A13 hold. Then, for any $0 < \epsilon \leq 1$ and for some $0 \leq \nu < 1/2$, we have*

$$\varepsilon_0(\breve{\mathfrak{m}}, n, M_n) \lesssim_p (\log((n \vee M_n)/\epsilon))^{\nu+1/2}\|\psi_{\tau,\cdot}^0\|_{\psi_\nu,0}.$$

where $\psi_{\tau,\cdot}^0$ is presented in Section 2.3 of the main text.

Proof. Similar to the proof of Lemma B.10 in Chernozhukov et al. (2021), since

$$\varepsilon_0(\breve{\mathfrak{m}}, n, M_n) = \sup_{\boldsymbol{\delta} \in A^{\breve{\mathfrak{m}}}_{M_n}} n^{-1/2} \sum_{t=1}^{n} \boldsymbol{\delta}^\top [\psi_\tau(V_t, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t)) - E\{\psi_\tau(V_t, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t))\}],$$

consider the class of function $\mathcal{F}_\alpha = \{v \mapsto \boldsymbol{\delta}^\top \psi_\tau(v, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t)), \boldsymbol{\delta} \in A^{\breve{\mathfrak{m}}}_{M_n}\}$, the cardinality of the set is $\text{card}(\mathcal{F}_\alpha) = n \vee M_n$. Then, for any $0 < \epsilon \leq 1$, the corresponding covering number is given by $\sup_{\mathcal{Q}} \mathcal{N}(\epsilon \|\bar{F}_\alpha\|_{\mathcal{Q},2}, \mathcal{F}_\alpha, \|\cdot\|_{\mathcal{Q},2}) = (n \vee M_n)/\epsilon$, with $\bar{F}_\alpha = \sup_{f \in \mathcal{F}_\alpha} |f|$. Recall that $\psi^0_{\tau,t} \equiv \psi_\tau(V_t, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t))$ and applying the tail probability bounds in Lemma B.4 in Chernozhukov et al. (2021), we have

$$\varepsilon_0(\breve{\mathfrak{m}}, n, M_n) \lesssim (\log((n \vee M_n)/\epsilon))^{\nu+1/2} \max_{f \in \mathcal{F}_\alpha} \|f(v_\cdot)\|_{\psi_\nu,0} \leq C(\log((n \vee M_n)/\epsilon))^{\nu+1/2} \|\psi^0_{\tau,\cdot}\|_{\psi_\nu,0}.$$

Then, we complete the proof of Lemma B.4. □

Lemma B.5. *Under Assumption A1–A13, we have*

$$\varepsilon_1(\breve{\mathfrak{m}}, n, M_n) \lesssim_p \sqrt{\breve{\mathfrak{m}}(\tilde{\varphi}(\breve{\mathfrak{m}}) \vee \varphi(\breve{\mathfrak{m}}))}.$$

Proof. First, write $\varepsilon_1(\breve{\mathfrak{m}}, n, M_n)$ as

$$\begin{aligned}
\varepsilon_1(\breve{\mathfrak{m}}, n, M_n) &= \sup_{\boldsymbol{\delta} \in A^{\breve{\mathfrak{m}}}_{M_n}} n^{1/2} \boldsymbol{\delta}^\top [S_n(\bar{\boldsymbol{c}}) - S^0 - E\{S_n(\bar{\boldsymbol{c}}) - S^0\}]\Big| \\
&\equiv \sup_{\boldsymbol{\delta} \in A^{\breve{\mathfrak{m}}}_{M_n}} n^{-1/2} \left| \sum_{t=1}^{n} V_{nt}(\boldsymbol{\delta}) - E(V_{nt}(\boldsymbol{\delta})) \right| \\
&\leq \sup_{\boldsymbol{\delta} \in A^{\breve{\mathfrak{m}}}_{M_n}} n^{-1/2} \left| \sum_{t=1}^{n} V_{nt}(\boldsymbol{\delta}) - E_t(V_{nt}(\boldsymbol{\delta})) \right| + \sup_{\boldsymbol{\delta} \in A^{\breve{\mathfrak{m}}}_{M_n}} n^{-1/2} \left| \sum_{t=1}^{n} E_t(V_{nt}(\boldsymbol{\delta})) - E(V_{nt}(\boldsymbol{\delta})) \right| \\
&\equiv B_{n1} + B_{n2}
\end{aligned}$$

where $V_{nt}(\boldsymbol{\delta}) \equiv [\psi_\tau(Y_t - \boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}}) - \psi_\tau(Y_t - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t)] \boldsymbol{\delta}^\top \boldsymbol{\Pi}_t$.

Now, we derive the upper bound of B_{n1}. For any $0 < \epsilon \leq 1$, covering the ball $A^{\breve{\mathfrak{m}}}_{M_n}$ with cubes

$\mathcal{C} = \{\mathcal{C}_\imath\}$, where $\mathcal{C}_\imath$ is a cube with center $\boldsymbol{\delta}_\imath$ and side length $1/n$, we have that $\text{card}(\mathcal{C}) = n^{\breve{\mathfrak{m}}} = N(n)$ and for $\boldsymbol{\delta} \in \mathcal{C}_\imath$, $\|\boldsymbol{\delta} - \boldsymbol{\delta}_\imath\| \leq \breve{\mathfrak{m}}^{1/2}/n$. Thus,

$$\begin{aligned}
B_{n1} \equiv & \sup_{\boldsymbol{\delta} \in A_{M_n}^{\breve{\mathfrak{m}}}} n^{-1/2} \left| \sum_{t=1}^{n} V_{nt}(\boldsymbol{\delta}) - E_t(V_{nt}(\boldsymbol{\delta})) \right| \\
\leq & \max_{1 \leq \imath \leq N(n)} n^{-1/2} \left| \sum_{t=1}^{n} V_{nt}(\boldsymbol{\delta}_\imath) - E_t(V_{nt}(\boldsymbol{\delta}_\imath)) \right| \\
& + \breve{\mathfrak{m}}^{1/2} n^{-3/2} \sum_{t=1}^{n} \Big\{ [\psi_\tau(Y_t - \boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}}) - \psi_\tau(Y_t - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t)] \\
& - E_t\{\psi_\tau(Y_t - \boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}}) - \psi_\tau(Y_t - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t)\} \Big\} \\
\equiv & B_{n11} + B_{n12}.
\end{aligned}$$

We only focus on B_{n11}, since B_{n12} can be bounded in the same way. Notice that for any $\flat > 0$, $|\psi_\tau(Y_t - \boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}}) - \psi_\tau(Y_t - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t)|^\flat = I(u_{3t} < Y_t \leq u_{4t})$, where $u_{3t} = \min(q_{2t}, q_{3t})$ and $u_{4t} = \max(q_{2t}, q_{3t})$ with $q_{2t} = \boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}}$ and $q_{3t} = \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t$. Therefore, by Assumption A6, there exists a constant $C > 0$ such that

$$\begin{aligned}
E\{|\psi_\tau(Y_t - \boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}}) - \psi_\tau(Y_t - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t)|^\flat | Z_t, \boldsymbol{W}_t\} &= F_{Y|Z,\boldsymbol{W}}(u_{4t}) - F_{Y|Z,\boldsymbol{W}}(u_{3t}) \\
&\leq C|\boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}} - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t|,
\end{aligned} \tag{B.7}$$

which implies by Assumption A3 and A9 that

$$E_t|V_{nt}(\boldsymbol{\delta})|^2 \leq C|\boldsymbol{\Pi}_t^\top \bar{\boldsymbol{c}} - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t| \left(\frac{1}{n} \sum_{t=m+1}^{n} \boldsymbol{\delta}^\top \boldsymbol{\Pi}_t \boldsymbol{\Pi}_t^\top \boldsymbol{\delta} \right) \leq C r_n K^{-d} \varphi(\breve{\mathfrak{m}}).$$

Thus, we have

$$W_n^2 = \sum_{t=m+1}^{n} E_t[V_{nt}(\boldsymbol{\delta}_\imath) - E_t(V_{nt}(\boldsymbol{\delta}_\imath))]^2 \leq \sum_{t=m+1}^{n} E_t[V_{nt}(\boldsymbol{\delta}_\imath)]^2 = O(r_n n K^{-d} \varphi(\breve{\mathfrak{m}}))$$

and

$$U_n^2 = \sum_{t=m+1}^{n} [V_{nt}(\boldsymbol{\delta}_\imath) - E_t(V_{nt}(\boldsymbol{\delta}_\imath))]^2 = O_p(r_n n K^{-d} \varphi(\breve{\mathfrak{m}})).$$

Also, notice that $\eta_{nt}(\boldsymbol{\delta}_\imath) \equiv \{V_{nt}(\boldsymbol{\delta}_\imath) - E_t(V_{nt}(\boldsymbol{\delta}_\imath))\}$ is a martingale difference sequence. Therefore, let $L = r_n n K^{-d}\varphi(\breve{\mathfrak{m}})$ and $M = \sqrt{\breve{\mathfrak{m}}\varphi(\breve{\mathfrak{m}})}$. Thus, we have

$$\begin{aligned}
& P\left[\max_{1\leq \imath \leq N(n)} \left| \frac{1}{\sqrt{n}} \sum_{t=m+1}^{n} \{V_{nt}(\boldsymbol{\delta}_\imath) - E_t(V_{nt}(\boldsymbol{\delta}_\imath))\} \right| > M \right] \\
\leq\ & N(n) \max_\imath P\left[\left| \frac{1}{\sqrt{n}} \sum_{t=m+1}^{n} \{V_{nt}(\boldsymbol{\delta}_\imath) - E_t(V_{nt}(\boldsymbol{\delta}_\imath))\} \right| > M \right] \\
\leq\ & N(n) \max_\imath P\left[\left| \sum_{t=m+1}^{n} \eta_{nt}(\boldsymbol{\delta}_\imath) \right| > \sqrt{n}M, W_n^2 + U_n^2 \leq L \right] \\
& + N(n) \max_\imath P\left[\left| \sum_{t=m+1}^{n} \eta_{nt}(\boldsymbol{\delta}_\imath) \right| > \sqrt{n}M, W_n^2 + U_n^2 > L \right] \equiv D_{n,1} + D_{n,2}. \qquad \text{(B.8)}
\end{aligned}$$

For $D_{n,1}$, by exponential inequality for martingale difference sequences (see, e.g., Bercu and Touati, 2008), we have

$$\begin{aligned}
& N(n) \max_\imath P\left[\left| \sum_{t=m+1}^{n} \eta_{nt}(\boldsymbol{\delta}_\imath) \right| > \sqrt{n}M, W_n^2 + U_n^2 \leq L \right] \\
\leq & 2N(n) \exp\left(-\frac{nM^2}{2L} \right).
\end{aligned}$$

For $D_{n,2}$, because $P\left[W_n^2 + U_n^2 > L\right] \leq P\left[W_n^2 > L/2\right] + P\left[U_n^2 > L/2\right]$ and each term can be bounded exponentially under Assumptions A1, A7 and A8. Thus, $B_{n11} = O_p(\sqrt{\breve{\mathfrak{m}}\varphi(\breve{\mathfrak{m}})})$. Similarly, one can show that $B_{n12} = o_p(\sqrt{\breve{\mathfrak{m}}\varphi(\breve{\mathfrak{m}})})$. Therefore, $B_{n1} = O_p(\sqrt{\breve{\mathfrak{m}}\varphi(\breve{\mathfrak{m}})})$.

Next, we consider B_{n2}. Similar to the proof of Lemma B.4, define the class of function $\mathcal{F}_{\alpha\bar{c}} = \{v \mapsto \boldsymbol{\delta}^\top[\psi_\tau(v, \boldsymbol{\alpha}_\tau(\boldsymbol{Z}_t)) - \psi_\tau(v, \bar{\boldsymbol{c}})], \boldsymbol{\delta} \in A_{M_n}^{\breve{\mathfrak{m}}}\}$, the cardinality of the set is $\text{card}(\mathcal{F}_{\alpha\bar{c}}) = n \vee M_n$. For any $0 < \epsilon \leq 1$, the corresponding covering number of $\mathcal{F}_{\alpha\bar{c}}$ is given by $\sup_{\mathcal{Q}} \mathcal{N}(\epsilon\|\bar{F}_{\alpha\bar{c}}\|_{\mathcal{Q},2}, \mathcal{F}_{\alpha\bar{c}}, \|\cdot\|_{\mathcal{Q},2}) = (n \vee M_n)/\epsilon$, with $\bar{F}_{\alpha\bar{c}} = \sup_{f\in\mathcal{F}_{\alpha\bar{c}}} |f|$. Therefore, for any $f \in \mathcal{F}_{\alpha\bar{c}}$, there exists a set

$F_{\alpha\bar{c},n}$ such that $\min_{f'\in F_{\alpha\bar{c},n}} \|f-f'\|_{\mathcal{Q},2} \le \tilde{\epsilon}$, where $\tilde{\epsilon} \equiv \epsilon\|2\bar{F}_{\alpha\bar{c}}\|_{\mathcal{Q},2}$, and the cardinality of $F_{\alpha\bar{c},n}$ is $\text{card}(F_{\alpha\bar{c},n}) = (n \vee M_n)/\epsilon$. Let $\phi_t \equiv f(v_t) = (\phi_{s,t})_{1\le s\le \text{card}(F_{\alpha\bar{c},n})}$ be the vector of length $\text{card}(F_{\alpha\bar{c},n})$ and denote $\breve{\phi}_{s,t} = E(\phi_{s,t}|\mathcal{F}_{t-1}) - E(\phi_{s,t})$. Then,

$$B_{n2} \le \max_{1\le s\le \text{card}(F_{\alpha\bar{c},n})} n^{-1/2}\left|\sum_{t=1}^{n} \breve{\phi}_{s,t}\right|.$$

Thus, applying the tail probability bounds in Lemma B.4 in Chernozhukov et al. (2021) to the vector $\breve{\phi}_{s,t}$, we have with probability greater than $1-o(1)$,

$$\max_{1\le s\le \text{card}(F_{\alpha\bar{c},n})} n^{-1/2}\left|\sum_{t=1}^{n} \breve{\phi}_{s,t}\right| \lesssim_p (\log((n\vee M_n)/\epsilon))^{\nu+1/2}\|\breve{\phi}_{s,\cdot}\|_{\psi_\nu,0}. \tag{B.9}$$

To bound $\|\breve{\phi}_{s,\cdot}\|_{\psi_\nu,0}$ in (B.9), notice that by (B.7),

$$E|\breve{\phi}_{s,t}|^2 \le C|\mathbf{\Pi}_t^\top \bar{\boldsymbol{c}} - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t|^2 E\{\boldsymbol{\delta}^\top \mathbf{\Pi}_t \mathbf{\Pi}_t^\top \boldsymbol{\delta}\} \le C(r_n K^{-d})^2 \tilde{\varphi}(\breve{\mathfrak{m}}).$$

Following Wu (2005), the definition of $\|\breve{\phi}_{s,\cdot}\|_{2,0}$ implies that $\|\breve{\phi}_{s,\cdot}\|_{2,0} \le 2(E|\breve{\phi}_{s,t}|^2)^{1/2}$. Since $(E|\breve{\phi}_{s,t}|^{\mathfrak{p}})^{1/\mathfrak{p}} \le (E|\breve{\phi}_{s,t}|^2)^{1/2}$ when $\mathfrak{p} \ge 2$, we have $\|\breve{\phi}_{s,\cdot}\|_{\psi_\nu,0} = \sup_{q\ge 2} q^{-1/2}\|\breve{\phi}_{s,\cdot}\|_{q,0} \le \|\breve{\phi}_{s,\cdot}\|_{2,0} \le C(r_n K^{-d})\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})}$. Thus, (B.9) becomes to

$$B_{n2} \le \max_{1\le s\le \text{card}(F_{\alpha\bar{c},n})} n^{-1/2}\sum_{t=1}^{n} \breve{\phi}_{s,t} \lesssim_p (\log((n\vee M_n)/\epsilon))^{\nu+1/2}(r_n K^{-d})\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})} \le C\sqrt{\breve{\mathfrak{m}}\tilde{\varphi}(\breve{\mathfrak{m}})}.$$

As $P(B_{n1}+B_{n2} \ge x) \le P(B_{n1} \le x/2) + P(B_{n2} \le x/2)$, we complete the proof of Lemma B.5. □

Lemma B.6. *Under Assumption A1–A13, for any* $0 < \epsilon \le 1$ *and for some* $0 \le \nu < 1/2$, *we have*

$$\varepsilon_2(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}, n, M_n) \lesssim_p (\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon))^{\nu+1/2}\Big\{(\tilde{\varphi}(\breve{\mathfrak{m}}))^{1/2} + \|\max_{\boldsymbol{c}\in R(\tilde{\boldsymbol{\zeta}}_n,\breve{\mathfrak{m}})} \partial_{\boldsymbol{c}} E\{\psi_\tau(V_\cdot,\boldsymbol{c})|\mathcal{F}_\cdot\}\|_{\psi_\nu,0}\Big\}.$$

Proof. The proof of Lemma B.6 is similar to that of Lemma B.9 in Chernozhukov et al. (2021). In particular, consider the class of function $\tilde{\mathcal{F}}_c = \{v \mapsto \boldsymbol{\delta}^\top[\psi_\tau(v, \boldsymbol{c}) - \psi_\tau(v, \bar{\boldsymbol{c}})], \boldsymbol{c} \in R(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}), \boldsymbol{\delta} \in \boldsymbol{S}(\boldsymbol{c})\}$. By Assumption A10, the entropy of the function set $\tilde{\mathcal{F}}_c$ is given by $\text{ent}(\epsilon, \tilde{\mathcal{F}}_c) \leq C\breve{\mathfrak{m}} \log((n \vee M_n)/\epsilon)$, for some $C > 0$. Therefore, for any $f \in \tilde{\mathcal{F}}_c$, there exists a set $F_{c,n}$ such that $\min_{f' \in F_{c,n}} \|f - f'\|_{\mathcal{Q},2} \leq \bar{\epsilon}$, where $\bar{\epsilon} \equiv \epsilon \|2\bar{F}_c\|_{\mathcal{Q},2}$, $\bar{F}_c = \sup_{f \in \tilde{\mathcal{F}}_c} |f|$ and the cardinality of $F_{c,n}$ is $\text{card}(F_{c,n}) = ((n \vee M_n)/\epsilon)^{C\breve{\mathfrak{m}}}$. Then, we have

$$\sup_{f \in \tilde{\mathcal{F}}_c} \left| \frac{1}{n} \sum_{t=1}^{n} [f - \chi(f) - E\{f - \chi(f)\}] \right| \leq 2\bar{\epsilon},$$

where $\chi(f) \equiv \arg\min_{f' \in \tilde{\mathcal{F}}_c} \|f - f'\|_{\mathcal{Q},2}$. Define $E_n(f) = E_n(f(v_t)) \equiv n^{-1} \sum_{t=1}^{n} f(v_t)$. Hence, with probability $1 - o(1)$,

$$\begin{aligned}
\varepsilon_2(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}, n, M_n) = & \sup_{\boldsymbol{c} \in R(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}), \boldsymbol{\delta} \in \boldsymbol{S}(\boldsymbol{c})} n^{1/2} \left| \boldsymbol{\delta}^\top [S_n(\boldsymbol{c}) - S_n(\bar{\boldsymbol{c}}) - E\{S_n(\boldsymbol{c}) - S_n(\bar{\boldsymbol{c}})\}] \right| \\
= & n^{1/2} \sup_{f \in \tilde{\mathcal{F}}_c} \left| [E_n(f) - E_n\{\pi(f)\} - E(f) + E\{\pi(f)\}] + [E_n\{\pi(f)\} - E\{\pi(f)\}] \right| \\
\leq & 2n^{1/2}\bar{\epsilon} + n^{1/2} \max_{f \in F_{c,n}} \left| E_n(f) - E(f) \right| \\
\leq & 2n^{1/2}\bar{\epsilon} + n^{1/2} \max_{f \in F_{c,n}} \left| E_n(f) - E_n(E(f|\mathcal{F}_{t-1})) \right| \\
& + n^{1/2} \max_{f \in F_{c,n}} \left| E_n(E(f|\mathcal{F}_{t-1})) - E(f) \right| \\
\equiv & 2n^{1/2}\bar{\epsilon} + C_{n,1} + C_{n,2}.
\end{aligned} \tag{B.10}$$

Now, we look for the bounds for $C_{n,1}$. Consider the function set $F_{c,n}$, for each $f \in F_{c,n}$, let $\varphi_t \equiv f(v_t) = (\varphi_{s,t})_{1 \leq s \leq \text{card}(F_{c,n})}$ and $\tilde{\varphi}_t \equiv (\tilde{\varphi}_{s,t})_{1 \leq s \leq \text{card}(F_{c,n})}$ be vectors with length $\text{card}(F_{c,n}) = ((n \vee M_n)/\epsilon)^{C\breve{\mathfrak{m}}}$, where $\tilde{\varphi}_{s,t} \equiv \varphi_{s,t} - E(\varphi_{s,t}|\mathcal{F}_{t-1})$ form martingale differences. Again, following Wu (2005), the definition of $\|\tilde{\varphi}_{s,\cdot}\|_{2,0}$ implies that $\|\tilde{\varphi}_{s,\cdot}\|_{2,0} \leq 2(E|\tilde{\varphi}_{s,t}|^2)^{1/2} \lesssim 8(E|\varphi_{s,t}|^2)^{1/2}$. Moreover, by $|\boldsymbol{\delta}^\top[\psi_\tau(v, \boldsymbol{c}) - \psi_\tau(v, \bar{\boldsymbol{c}})]| \leq C|\boldsymbol{\delta}^\top \boldsymbol{\Pi}_t|$ for sufficiently large $C > 0$, we have $E|\varphi_{s,t}|^2 \leq \sup_{\boldsymbol{\delta} \in A_{M_n}^{\breve{\mathfrak{m}}}} E[(\boldsymbol{\delta}^\top \boldsymbol{\Pi}_t)^2] = \tilde{\varphi}(\breve{\mathfrak{m}})$. Since $(E|\tilde{\varphi}_{s,t}|^{\mathfrak{p}})^{1/\mathfrak{p}} \leq (E|\tilde{\varphi}_{s,t}|^2)^{1/2}$ when $\mathfrak{p} \geq 2$, we have $\|\tilde{\varphi}_{s,\cdot}\|_{\psi_\nu,0} = \sup_{q \geq 2} q^{-1/2} \|\tilde{\varphi}_{s,\cdot}\|_{q,0} \leq \|\tilde{\varphi}_{s,\cdot}\|_{2,0} \leq C\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})}$. Then, applying the tail probability bounds in

Lemma B.4 in Chernozhukov et al. (2021) to the vector $\tilde{\varphi}_{s,t}$, we have with probability greater than $1-o(1)$,

$$C_{n,1} \lesssim_p (\breve{\mathfrak{m}} \log((n \vee M_n)/\epsilon))^{\nu+1/2} \|\tilde{\varphi}_{s,\cdot}\|_{\psi_\nu,0} \leq C(\breve{\mathfrak{m}} \log((n \vee M_n)/\epsilon))^{\nu+1/2} (\tilde{\varphi}(\breve{\mathfrak{m}}))^{1/2}. \quad \text{(B.11)}$$

Next, we handle the term $C_{n,2}$. Again, for each $f \in F_{c,n}$, let $\breve{\varphi}_{s,t} = E(\varphi_{s,t}|\mathcal{F}_{t-1}) - E(\varphi_{s,t})$. Then,

$$C_{n,2} \leq \max_{1 \leq s \leq \text{card}(F_{c,n})} n^{-1/2} \left| \sum_{t=1}^{n} \breve{\varphi}_{s,t} \right|.$$

Moreover, for $1 \leq s \leq \text{card}(F_{c,n})$, there is a function $\mathfrak{g}$ corresponding to each $f \in F_{c,n}$ such that $\breve{\varphi}_{s,t} = \mathfrak{g}(v_t, \boldsymbol{c})$, where $\boldsymbol{c} \in R(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}})$. By the mean value theorem and the continuity of the function $\mathfrak{g}$, we have

$$\mathfrak{g}(v_t, \boldsymbol{c}) = \partial_{\boldsymbol{c}} \mathfrak{g}(v_t, \boldsymbol{c}^*)(\boldsymbol{c} - \bar{\boldsymbol{c}}),$$

where $\boldsymbol{c}^*$ is the corresponding point which joins the line segment between $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$. Then,

$$\max_{1 \leq s \leq \text{card}(F_{c,n})} n^{-1/2} \sum_{t=1}^{n} \breve{\varphi}_{s,t} = \max_{\boldsymbol{c}^* \in F_{c,n}^{\boldsymbol{c}}} n^{-1/2} \sum_{t=1}^{n} \partial_{\boldsymbol{c}} \mathfrak{g}(v_t, \boldsymbol{c}^*)(\boldsymbol{c} - \bar{\boldsymbol{c}}),$$

where $F_{c,n}^{\boldsymbol{c}}$ collects all the points of $\boldsymbol{c}$ according to $F_{c,n}$. Thus, we have

$$\begin{aligned} \max_{1 \leq s \leq \text{card}(F_{c,n})} n^{-1/2} \sum_{t=1}^{n} \breve{\varphi}_{s,t} \lesssim_p & (\breve{\mathfrak{m}} \log((n \vee M_n)/\epsilon))^{\nu+1/2} \|\partial_{\boldsymbol{c}} \mathfrak{g}(v_t, \boldsymbol{c}^*)\|_{\psi_\nu,0} \\ \lesssim & (\breve{\mathfrak{m}} \log((n \vee M_n)/\epsilon))^{\nu+1/2} \|\max_{\boldsymbol{c} \in R(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}})} \partial_{\boldsymbol{c}} E\{\psi_\tau(V_\cdot, \boldsymbol{c})|\mathcal{F}_\cdot\}\|_{\psi_\nu,0} \end{aligned} \quad \text{(B.12)}$$

As $P(C_{n,1} + C_{n,2} \geq x) \leq P(C_{n,1} \leq x/2) + P(C_{n,2} \leq x/2)$ and collecting the results from (B.11) and (B.12), we complete the proof of Lemma B.6. ☐

Lemma B.7. *Under Assumption A1–A13, we have*

$$\varepsilon_3(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}}, n, M_n) \lesssim_p \sqrt{n \tilde{\varphi}(\breve{\mathfrak{m}})} (\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})} \bar{f} \tilde{\boldsymbol{\zeta}}_n \wedge 1).$$

Proof. The proof is the same as that of Lemma 4 in Tang et al. (2013). We omit the proof here. □

Lemma B.8. *Let* $\Phi^{\mathbf{c}}_{\psi_\nu,0} \equiv \| \max_{\mathbf{c}\in R(\tilde{\boldsymbol{\zeta}}_n,\breve{\mathfrak{m}})} E\{\psi_\tau(V_\cdot, \boldsymbol{c})|\mathcal{F}_\cdot\}\|_{\psi_\nu,0}$ *and* $\mu(\breve{\mathfrak{m}}) = \sqrt{\varphi(\breve{\mathfrak{m}})}(\bar{f}\sqrt{\varphi(\breve{\mathfrak{m}})}\vee 1)$. *Suppose that Assumption A1–A13 hold and* $Y_1,\ldots,Y_n$ *are absolutely continuous conditional on* $Z_1,\ldots,Z_n$*,* $\mathbb{W}_1,\ldots,\mathbb{W}_n$*. Then, for some* $0\le\nu<1/2$*, we have*

$$\sqrt{\breve{\mathfrak{m}}} \lesssim_p \mu(\breve{\mathfrak{m}})\frac{n}{\lambda_{n,0}}(\tilde{\boldsymbol{\zeta}}_n\wedge 1) + \sqrt{\breve{\mathfrak{m}}}\frac{\sqrt{n\log((n\vee M_n)/\epsilon)}\{(\tilde{\varphi}(\breve{\mathfrak{m}}))^{1/2}+\Phi^{\mathbf{c}}_{\psi_\nu,0}\}}{\lambda_{n,0}}(\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon))^{\nu}.$$

Proof. Similar to Tang et al. (2013), four vectors of rank scores (dual variables) are defined to derive the proof:

1. the true rank scores, $a^*_{t,\tau} = \tau - I(Y_t < \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t)$ for $t=1,\ldots,n$;
2. the oracle rank scores, $\bar{a}_{t,\tau} = \tau - I(Y_t < \boldsymbol{\Pi}_t^\top\bar{\boldsymbol{c}})$ for $t=1,\ldots,n$;
3. the estimated rank scores, $a_{t,\tau} = \tau - I(Y_t < \boldsymbol{\Pi}_t^\top\tilde{\boldsymbol{c}})$ for $t=1,\ldots,n$;
4. the dual optimal rank scores, $\tilde{a}_{t,\tau}$, $t=1,\ldots,n$, that solve the dual program (B.5).

Let $\tilde{T} = \text{support}(\tilde{\boldsymbol{c}})$, and let $\boldsymbol{\Pi}_{t\tilde{T}}$ and $\tilde{\boldsymbol{c}}_{\tilde{T}}$ be the corresponding sub-vectors of $\boldsymbol{\Pi}_t$ and $\tilde{\boldsymbol{c}}$, respectively. From (B.6), we have that

$$\sqrt{\breve{\mathfrak{m}}} = \|\text{sign}(\tilde{\boldsymbol{c}}_{\tilde{T}})\|_2 = \left\|\frac{\sum_{t=1}^n \boldsymbol{\Pi}_{t\tilde{T}}\tilde{a}_{t,\tau}}{\lambda_{n,0}}\right\|_2. \tag{B.13}$$

Using the triangle inequality on (B.13),

$$\begin{aligned}\lambda_{n,0}\sqrt{\breve{\mathfrak{m}}} \le& \left\|\sum_{t=1}^n \boldsymbol{\Pi}_{t\tilde{T}}\{\tilde{a}_{t,\tau}-a_{t,\tau}\}\right\|_2 + \left\|\sum_{t=1}^n \boldsymbol{\Pi}_{t\tilde{T}}\{a_{t,\tau}-\bar{a}_{t,\tau}\}\right\|_2\\ &+\left\|\sum_{t=1}^n \boldsymbol{\Pi}_{t\tilde{T}}\{\bar{a}_{t,\tau}-a^*_{t,\tau}\}\right\|_2 + \left\|\sum_{t=1}^n \boldsymbol{\Pi}_{t\tilde{T}}a^*_{t,\tau}\right\|_2\end{aligned} \tag{B.14}$$

To bound the first component, following the proof of Lemma 5 in Tang et al. (2013), we observe that $\tilde{a}_{t,\tau}\ne a_{t,\tau}$ only if $Y_t = \boldsymbol{\Pi}_t^\top\tilde{\boldsymbol{c}}$. By Lemma B.3, the penalized quantile regression fit can interpolate at most $\breve{\mathfrak{m}}$ points with probability one. This implies that $n^{-1}\sum_{t=1}^n\{\tilde{a}_{t,\tau}-a_{t,\tau}\}^2 \le$

$\breve{\mathfrak{m}}/n$. Thus, by Cauchy–Schwarz inequality,

$$\begin{aligned}\Big\|\sum_{t=1}^n \mathbf{\Pi}_{t\tilde{T}}\{\tilde{a}_{t,\tau}-a_{t,\tau}\}\Big\|_2 \le & n \sup_{\boldsymbol{\delta}\in A^{\breve{\mathfrak{m}}}_{M_n}} n^{-1}\sum_{t=1}^n |\boldsymbol{\delta}^\top \mathbf{\Pi}_t||\tilde{a}_{t,\tau}-a_{t,\tau}| \\ \le & n\Big(\sup_{\boldsymbol{\delta}\in A^{\breve{\mathfrak{m}}}_{M_n}} n^{-1}\sum_{t=1}^n |\boldsymbol{\delta}^\top \mathbf{\Pi}_t|^2\Big)^{1/2}\Big(n^{-1}\sum_{t=1}^n |\tilde{a}_{t,\tau}-a_{t,\tau}|^2\Big)^{1/2} \\ \le & n\Big(\varphi(\breve{\mathfrak{m}})\frac{\breve{\mathfrak{m}}}{n}\Big)^{1/2} = \sqrt{n\breve{\mathfrak{m}}\varphi(\breve{\mathfrak{m}})}.\end{aligned} \tag{B.15}$$

For the second component, note that by Lemma B.6 and B.7,

$$\begin{aligned}\Big\|\sum_{t=1}^n \mathbf{\Pi}_{t\tilde{T}}\{a_{t,\tau}-\bar{a}_{t,\tau}\}\Big\|_2 \le & \Big\|\sum_{t=1}^n \Big(\mathbf{\Pi}_{t\tilde{T}}\{a_{t,\tau}-\bar{a}_{t,\tau}\} - E[\mathbf{\Pi}_{t\tilde{T}}\{a_{t,\tau}-\bar{a}_{t,\tau}\}]\Big)\Big\|_2 \\ & + \Big\|\sum_{t=1}^n E[\mathbf{\Pi}_{t\tilde{T}}\{a_{t,\tau}-\bar{a}_{t,\tau}\}]\Big\|_2 \\ \le & n^{1/2}\varepsilon_2(\tilde{\boldsymbol{\zeta}}_n,\breve{\mathfrak{m}},n,M_n) + n^{1/2}\varepsilon_3(\tilde{\boldsymbol{\zeta}}_n,\breve{\mathfrak{m}},n,M_n) \\ \le & \sqrt{n\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon)}(\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon))^{\nu}\{(\tilde{\varphi}(\breve{\mathfrak{m}}))^{1/2}+\Phi^{\mathbf{c}}_{\psi_\nu,0}\} \\ & + n\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})}(\sqrt{\tilde{\varphi}(\breve{\mathfrak{m}})}\bar{f}\tilde{\boldsymbol{\zeta}}_n\wedge 1).\end{aligned} \tag{B.16}$$

To bound the third component, following the same argument as in the proof of Lemma 5 in Tang et al. (2013) and by Lemma B.5, we have

$$\begin{aligned}\Big\|\sum_{t=1}^n \mathbf{\Pi}_{t\tilde{T}}\{\bar{a}_{t,\tau}-a^*_{t,\tau}\}\Big\|_2 \le & \Big\|\sum_{t=1}^n \Big(\mathbf{\Pi}_{t\tilde{T}}\{\bar{a}_{t,\tau}-a^*_{t,\tau}\} - E[\mathbf{\Pi}_{t\tilde{T}}\{\bar{a}_{t,\tau}-a^*_{t,\tau}\}]\Big)\Big\|_2 \\ & + \Big\|\sum_{t=1}^n E[\mathbf{\Pi}_{t\tilde{T}}\{\bar{a}_{t,\tau}-a^*_{t,\tau}\}]\Big\|_2 \\ \le & n^{1/2}\varepsilon_1(\breve{\mathfrak{m}},n,M_n) + n\sup_{\boldsymbol{\delta}\in A^{\breve{\mathfrak{m}}}_{M_n}}\Big|\boldsymbol{\delta}^\top[E\{M_n(\bar{\boldsymbol{c}})\}-E\{M^0\}]\Big| \\ \lesssim_p & \sqrt{n\breve{\mathfrak{m}}(\tilde{\varphi}(\breve{\mathfrak{m}})\vee\varphi(\breve{\mathfrak{m}}))}.\end{aligned} \tag{B.17}$$

As for the last part, we use lemma B.4 to obtain

$$\begin{aligned}\Big\|\sum_{t=1}^{n}\mathbf{\Pi}_{t\tilde{T}}a^*_{t,\tau}\Big\|_2 &\leq n^{1/2}\varepsilon_0(\breve{\mathfrak{m}},n,M_n)\\ &\lesssim_p \sqrt{n\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon)}(\log((n\vee M_n)/\epsilon))^{\nu}\breve{\mathfrak{m}}^{-1/2}\|\psi^0_{\tau,\cdot}\|_{\psi_\nu,0}.\end{aligned} \tag{B.18}$$

Then, combining (B.15)-(B.18), we have

$$\sqrt{\breve{\mathfrak{m}}}\lesssim_p \mu(\breve{\mathfrak{m}})\frac{n}{\lambda_{n,0}}(\tilde{\boldsymbol{\zeta}}_n\wedge 1)+\sqrt{\breve{\mathfrak{m}}}\frac{\sqrt{n\log((n\vee M_n)/\epsilon))}\{(\tilde{\varphi}(\breve{\mathfrak{m}}))^{1/2}+\Phi^{\boldsymbol{c}}_{\psi_\nu,0}\}}{\lambda_{n,0}}(\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon))^{\nu},$$

which gives us the result in lemma B.8. □

Lemma B.9. *Suppose that Assumption A1–A13 hold. Let* $Q_\tau(\boldsymbol{c})=E\{\rho_\tau(Y_t-\mathbf{\Pi}_t^\top\boldsymbol{c})\}$*,* $\hat{Q}_\tau(\boldsymbol{c})=n^{-1}\sum_{t=1}^n\rho_\tau(Y_t-\mathbf{\Pi}_t^\top\boldsymbol{c})$*. Then, we have*

$$\begin{aligned}&\left|\hat{Q}_\tau(\boldsymbol{c})-Q_\tau(\boldsymbol{c})-\big(\hat{Q}_\tau(\bar{\boldsymbol{c}})-Q_\tau(\bar{\boldsymbol{c}})\big)\right|\\ \lesssim_p&\frac{\tilde{\boldsymbol{\zeta}}_n}{\sqrt{n}}((\breve{\mathfrak{m}}+R_n)\log((n\vee M_n)/\epsilon))^{\nu+1/2}\{(\tilde{\varphi}(\breve{\mathfrak{m}}+R_n))^{1/2}+\Phi^{\boldsymbol{c}}_{\psi_\nu,0}\}\end{aligned} \tag{B.19}$$

uniformly over $\boldsymbol{c}\in R(\tilde{\boldsymbol{\zeta}}_n,\breve{\mathfrak{m}})$.

Proof. Note that $-\frac{(\boldsymbol{c}-\bar{\boldsymbol{c}})^\top}{\tilde{\boldsymbol{\zeta}}_n}S_n\left(\frac{\tilde{\boldsymbol{\zeta}}_n-z}{\tilde{\boldsymbol{\zeta}}_n}\boldsymbol{c}+\frac{z}{\tilde{\boldsymbol{\zeta}}_n}\bar{\boldsymbol{c}}\right)$ is the sub-gradient of $\hat{Q}_\tau\left(\frac{\tilde{\boldsymbol{\zeta}}_n-z}{\tilde{\boldsymbol{\zeta}}_n}\boldsymbol{c}+\frac{z}{\tilde{\boldsymbol{\zeta}}_n}\bar{\boldsymbol{c}}\right)$, where $\tilde{\boldsymbol{\zeta}}_n\geq\|\boldsymbol{c}-\bar{\boldsymbol{c}}\|_2$ and $\|\boldsymbol{c}-\bar{\boldsymbol{c}}\|_0\leq\breve{\mathfrak{m}}+R_n$. Therefore, we have

$$\left|\hat{Q}_\tau(\boldsymbol{c})-Q_\tau(\boldsymbol{c})-\big(\hat{Q}_\tau(\bar{\boldsymbol{c}})-Q_\tau(\bar{\boldsymbol{c}})\big)\right|\leq b_1+b_2,$$

where

$$b_1\equiv\left|\int_0^{\tilde{\boldsymbol{\zeta}}_n}\frac{(\boldsymbol{c}-\bar{\boldsymbol{c}})^\top}{\tilde{\boldsymbol{\zeta}}_n}\left(S_n\left(\frac{\tilde{\boldsymbol{\zeta}}_n-z}{\tilde{\boldsymbol{\zeta}}_n}\boldsymbol{c}+\frac{z}{\tilde{\boldsymbol{\zeta}}_n}\bar{\boldsymbol{c}}\right)-E\left\{S_n\left(\frac{\tilde{\boldsymbol{\zeta}}_n-z}{\tilde{\boldsymbol{\zeta}}_n}\boldsymbol{c}+\frac{z}{\tilde{\boldsymbol{\zeta}}_n}\bar{\boldsymbol{c}}\right)\right\}-[S_n(\bar{\boldsymbol{c}})-E\{S_n(\bar{\boldsymbol{c}})\}]\right)dz\right|,$$

and

$$b_2 \equiv \tilde{\boldsymbol{\zeta}}_n \left| \frac{(\boldsymbol{c} - \bar{\boldsymbol{c}})^\top}{\tilde{\boldsymbol{\zeta}}_n} [S_n(\bar{\boldsymbol{c}}) - E\{S_n(\bar{\boldsymbol{c}})\}] \right|.$$

By Lemma B.6,

$$\begin{aligned} b_1 \leq & n^{-1/2} \int_0^{\tilde{\boldsymbol{\zeta}}_n} \varepsilon_2(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}} + R_n, n, M_n) dz = \frac{\tilde{\boldsymbol{\zeta}}_n}{\sqrt{n}} \varepsilon_2(\tilde{\boldsymbol{\zeta}}_n, \breve{\mathfrak{m}} + R_n, n, M_n) \\ \lesssim_p & \frac{\tilde{\boldsymbol{\zeta}}_n}{\sqrt{n}} ((\breve{\mathfrak{m}} + R_n) \log((n \vee M_n)/\epsilon))^{\nu+1/2} \{(\tilde{\varphi}(\breve{\mathfrak{m}} + R_n))^{1/2} + \Phi^{\mathbf{c}}_{\psi_\nu,0}\}. \end{aligned}$$

As for b_2, define the class of function $\mathcal{F}'_{\bar{c}} = \{v \mapsto \boldsymbol{\delta}^\top \psi_\tau(v, \bar{\boldsymbol{c}}), \boldsymbol{\delta} \in A_{M_n}^{\breve{\mathfrak{m}}+R_n}\}$. Then, similar to the proof of Lemma B.4, one can easily show that

$$b_2 \lesssim_p \frac{\tilde{\boldsymbol{\zeta}}_n}{\sqrt{n}} (\log((n \vee M_n)/\epsilon))^{\nu+1/2} \max_{f \in \mathcal{F}'_{\bar{c}}} \|f(v_\cdot)\|_{\psi_\nu,0} = o(b_1).$$

Then, Lemma B.9 is proved. □

B.2 Proof of Theorem 1:

Proof. Recall that $\tilde{\boldsymbol{\zeta}}_n = \|\tilde{\boldsymbol{c}} - \bar{\boldsymbol{c}}\|_2$, $\breve{\mathfrak{m}} = \|\tilde{\boldsymbol{c}}\|_0$ and $m_0 = M_n \wedge \left(\frac{n}{\log(n \vee M_n)} \frac{\mathfrak{q}^2}{\mu^2} \right)$.

(I) We first show that $\breve{\mathfrak{m}} \leq m_0$ if $\aleph \geq \sqrt{2}$. Since $\breve{\mathfrak{m}} \leq M_n$ is trivial, we only need to verify the result for $m_0 = \frac{n}{\log(n \vee M_n)} \frac{\mathfrak{q}^2}{\mu^2}$.

By Lemma B.3, we have

$$\breve{\mathfrak{m}} \leq \bar{m} = \max \left\{ m : m \leq n \wedge M_n \wedge \frac{n^2 \varphi(m)}{\lambda_{n,0}^2} \right\}.$$

Suppose that $\bar{m} > m_0$ when $\aleph \geq \sqrt{2}$. Therefore, we have $\bar{m} = \ell m_0$ for some $\ell > 1$. By definition, $\bar{m}$ satisfies the inequality

$$\bar{m} \leq \frac{n^2 \varphi(\bar{m})}{\lambda_{n,0}^2}. \tag{B.20}$$

Since $\lambda_{n,0} = \aleph \sqrt{n \log(n \vee M_n) \varphi(m_0 + R_n)} (m_0 \log(n \vee M_n))^\nu \frac{\mu}{q}$ for some $0 \leq \nu < 1/2$ and $\varphi(m_0)$

is bounded, it is clear that $\lambda_{n,0} \geq \aleph\sqrt{n\log(n \vee M_n)\varphi(m_0)}\frac{\mu}{q}$. Inserting this bound on $\lambda_{n,0}$, $m_0 = \frac{n}{\log(n\vee M_n)}\frac{\mathfrak{q}^2}{\mu^2}$, and $\bar{m} = \ell m_0$ into (B.20), one can obtain that when $\aleph \geq \sqrt{2}$,

$$\bar{m} = \ell m_0 \leq \frac{n^2}{\aleph^2 n\log(n \vee M_n)}\frac{\varphi(\ell m_0)}{\varphi(m_0)}\frac{q^2}{\mu^2} < \frac{n}{\aleph^2\log(n \vee M_n)}2\ell\frac{q^2}{\mu^2} = \frac{2}{\aleph^2}\ell m_0 \leq \ell m_0,$$

which is a contradiction.

(II) We now prove that $\tilde{\boldsymbol{\zeta}}_n = o_p(1)$.

Define $\bar{T} \equiv \{l : \|\bar{\boldsymbol{c}}_{l,\tau}\|_1 > 0, l = 1, \ldots, m\} = \{1, 2, \ldots, r_n\}$. Let $\tilde{\boldsymbol{c}}_{\bar{T}}$ denote a vector whose $\bar{T}$ groups agree with that of $\tilde{\boldsymbol{c}}$, and whose remaining groups equal zero.

By definition of $\tilde{\boldsymbol{c}}$ and since $\|\tilde{\boldsymbol{c}}_{\bar{T}}\|_1 \leq \|\tilde{\boldsymbol{c}}\|_1$, we have

$$\begin{aligned}\hat{Q}_\tau(\tilde{\boldsymbol{c}}) - \hat{Q}_\tau(\bar{\boldsymbol{c}}) \leq& \frac{\lambda_{n,0}}{n}(\|\bar{\boldsymbol{c}}\|_1 - \|\tilde{\boldsymbol{c}}\|_1) \leq \frac{\lambda_{n,0}}{n}(\|\bar{\boldsymbol{c}}\|_1 - \|\tilde{\boldsymbol{c}}_{\bar{T}}\|_1)\\ \leq& \frac{\lambda_{n,0}}{n}\Big|\|\bar{\boldsymbol{c}}\|_1 - \|\tilde{\boldsymbol{c}}_{\bar{T}}\|_1\Big| \leq \frac{\lambda_{n,0}}{n}\|\bar{\boldsymbol{c}} - \tilde{\boldsymbol{c}}_{\bar{T}}\|_1\\ \leq& \frac{\lambda_{n,0}}{n}\sqrt{\sum_{l=1}^{r_n} R_{l,n}\|\bar{\boldsymbol{c}} - \tilde{\boldsymbol{c}}_{\bar{T}}\|_2} \leq \frac{\lambda_{n,0}}{n}\sqrt{R_n}\tilde{\boldsymbol{\zeta}}_n.\end{aligned}$$

Applying (B.19) yields

$$Q_\tau(\tilde{\boldsymbol{c}}) - Q_\tau(\bar{\boldsymbol{c}}) \lesssim_p \frac{\lambda_{n,0}}{n}\sqrt{R_n}\tilde{\boldsymbol{\zeta}}_n + \frac{\tilde{\boldsymbol{\zeta}}_n}{\sqrt{n}}((\breve{\mathfrak{m}} + R_n)\log((n \vee M_n)/\epsilon))^{\nu+1/2}\{(\tilde{\varphi}(\breve{\mathfrak{m}} + R_n))^{1/2} + \Phi^{\mathbf{c}}_{\psi_\nu,0}\}.$$

By Lemma B.2 and by the definition of $\tilde{\boldsymbol{\zeta}}_n$, we obtain,

$$\mathfrak{q}(\tilde{\boldsymbol{\zeta}}_n^2 \wedge \tilde{\boldsymbol{\zeta}}_n) \lesssim_p \frac{\lambda_{n,0}}{n}\sqrt{R_n}\tilde{\boldsymbol{\zeta}}_n + \frac{\tilde{\boldsymbol{\zeta}}_n}{\sqrt{n}}((\breve{\mathfrak{m}} + R_n)\log((n \vee M_n)/\epsilon))^{\nu+1/2}\{(\tilde{\varphi}(\breve{\mathfrak{m}} + R_n))^{1/2} + \Phi^{\mathbf{c}}_{\psi_\nu,0}\} \tag{B.21}$$

By Assumption A9, A11 and by the construction of $\lambda_{n,0}$, one can obtain that

(i)

$$\frac{\lambda_{n,0}\sqrt{R_n}}{n} = o_p(\mathfrak{q}),$$

(ii)

$$\frac{(R_n \log((n \vee M_n)/\epsilon))^{\nu+1/2}\{(\tilde{\varphi}(\breve{\mathfrak{m}}+R_n))^{1/2}+\Phi^{\mathbf{c}}_{\psi_\nu,0}\}}{\sqrt{n}}\frac{\mu}{\mathfrak{q}} \leq \frac{C\lambda_{n,0}\sqrt{R_n}}{n\aleph}\left(\frac{R_n}{m_0}\right)^{\nu} o(\aleph) = o_p(\mathfrak{q})$$

(iii)

$$\mu\frac{\sqrt{n\log((n \vee M_n)/\epsilon)}\{(\tilde{\varphi}(\breve{\mathfrak{m}}+R_n))^{1/2}+\Phi^{\mathbf{c}}_{\psi_\nu,0}\}}{\lambda_{n,0}}(\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon))^{\nu} = o_p(\mathfrak{q}).$$

for some constant $C > 0$ and by choosing ϵ such that $\left(\log((n\vee M_n)/\epsilon)/\log(n\vee M_n)\right)^{\nu+1/2} = o(\aleph)$. Result (iii), the fact that $\mu \geq \mathfrak{q}$, $\breve{\mathfrak{m}} \leq n \wedge M_n$, and the empirical sparseness in Lemma B.8 imply that

$$\sqrt{\breve{\mathfrak{m}}} \lesssim_p \mu\frac{n}{\lambda_{n,0}}(\tilde{\boldsymbol{\zeta}}_n \wedge 1) + \sqrt{\breve{\mathfrak{m}}}o_p(1), \tag{B.22}$$

which implies

$$\sqrt{\breve{\mathfrak{m}}} \lesssim_p \mu\frac{n}{\lambda_{n,0}}, \tag{B.23}$$

Using (B.23), $\breve{\mathfrak{m}} \leq n \wedge M_n$ and $\breve{\mathfrak{m}} > R_n$ in (B.21) gives

$$\begin{aligned} & I(\breve{\mathfrak{m}} > R_n)\mathfrak{q}(\tilde{\boldsymbol{\zeta}}_n^2 \wedge \tilde{\boldsymbol{\zeta}}_n) \\ \lesssim_p & \frac{\lambda_{n,0}}{n}\sqrt{R_n}\tilde{\boldsymbol{\zeta}}_n + \tilde{\boldsymbol{\zeta}}_n\mu\frac{\sqrt{n\log((n \vee M_n)/\epsilon)}\{(\tilde{\varphi}(\breve{\mathfrak{m}}+R_n))^{1/2}+\Phi^{\mathbf{c}}_{\psi_\nu,0}\}}{\lambda_{n,0}}(\breve{\mathfrak{m}}\log((n\vee M_n)/\epsilon))^{\nu} \\ = & \tilde{\boldsymbol{\zeta}}_n o_p(\mathfrak{q}), \end{aligned} \tag{B.24}$$

where the last equality follows by results (i) and (ii). On the other hand, when $\breve{\mathfrak{m}} \leq R_n$, (B.21) yields

$$\begin{aligned} I(\breve{\mathfrak{m}} \leq R_n)\mathfrak{q}(\tilde{\boldsymbol{\zeta}}_n^2 \wedge \tilde{\boldsymbol{\zeta}}_n) \lesssim_p & \frac{\lambda_{n,0}}{n}\sqrt{R_n}\tilde{\boldsymbol{\zeta}}_n + \tilde{\boldsymbol{\zeta}}_n\frac{(R_n \log((n \vee M_n)/\epsilon))^{\nu+1/2}\{(\tilde{\varphi}(\breve{\mathfrak{m}}+R_n))^{1/2}+\Phi^{\mathbf{c}}_{\psi_\nu,0}\}}{\sqrt{n}} \\ = & \tilde{\boldsymbol{\zeta}}_n o_p(\mathfrak{q}), \end{aligned} \tag{B.25}$$

where the last equality follows by results (i) and (ii) and $\mu \geq \mathfrak{q}$. Adding both sides of (B.24) and (B.25), we have

$$\mathfrak{q}(\tilde{\boldsymbol{\zeta}}_n^2 \wedge \tilde{\boldsymbol{\zeta}}_n) = \tilde{\boldsymbol{\zeta}}_n o_p(\mathfrak{q}). \tag{B.26}$$

Dividing both sides of (B.26) by $\mathfrak{q}$ and by $\tilde{\boldsymbol{\zeta}}_n$, we have $I(\tilde{\boldsymbol{\zeta}}_n > 0)(\tilde{\boldsymbol{\zeta}}_n \wedge 1) \lesssim_p I(\tilde{\boldsymbol{\zeta}}_n > 0)o_p(1)$, which implies that $\tilde{\boldsymbol{\zeta}}_n = o_p(1)$.

(III) Finally, we derive the rate of convergence of $\tilde{\boldsymbol{\zeta}}_n$ and the final bound of $\breve{\mathfrak{m}}$. Using the result $\tilde{\boldsymbol{\zeta}}_n = o_p(1)$ and (B.22), we have

$$\sqrt{\breve{\mathfrak{m}}} \lesssim_p \mu \frac{n\tilde{\boldsymbol{\zeta}}_n}{\lambda_{n,0}}. \tag{B.27}$$

When $\breve{\mathfrak{m}} \leq R_n$, the final bound of $\breve{\mathfrak{m}}$ is found. Otherwise, plugging (B.27) into (B.21), we have

$$\begin{aligned}
&\frac{\tilde{\boldsymbol{\zeta}}_n}{\sqrt{n}}((\breve{\mathfrak{m}} + R_n)\log((n \vee M_n)/\epsilon))^{\nu+1/2}\{(\tilde{\varphi}(\breve{\mathfrak{m}} + R_n))^{1/2} + \Phi^{\mathbf{c}}_{\psi_\nu,0}\} \\
\lesssim_p &\tilde{\boldsymbol{\zeta}}_n^2 \mu \frac{\sqrt{n\log((n \vee M_n)/\epsilon)}\{(\tilde{\varphi}(\breve{\mathfrak{m}} + R_n))^{1/2} + \Phi^{\mathbf{c}}_{\psi_\nu,0}\}}{\lambda_{n,0}}(\breve{\mathfrak{m}}\log((n \vee M_n)/\epsilon))^{\nu} = \tilde{\boldsymbol{\zeta}}_n^2 o_p(\mathfrak{q}).
\end{aligned}$$

This, in conjunction with $\tilde{\boldsymbol{\zeta}}_n = o_p(1)$ and (B.21), gives that

$$\mathfrak{q}\tilde{\boldsymbol{\zeta}}_n^2 \lesssim_p \frac{\lambda_{n,0}}{n}\sqrt{R_n}\tilde{\boldsymbol{\zeta}}_n + \tilde{\boldsymbol{\zeta}}_n^2 o_p(\mathfrak{q}),$$

or equivalently,

$$\tilde{\boldsymbol{\zeta}}_n \lesssim_p \frac{\lambda_{n,0}\sqrt{R_n}}{n\mathfrak{q}}. \tag{B.28}$$

Finally, inserting (B.28) into (B.27), we obtain the final bound of $\|\tilde{\boldsymbol{c}}\|_0$, which is $\|\tilde{\boldsymbol{c}}\|_0 = \breve{\mathfrak{m}} \lesssim_p (\mu/\mathfrak{q})^2 R_n$. □

B.3 Proof of Theorem 2:

We obtain from Theorem 1 that $\tilde{\boldsymbol{\zeta}}_n = o_p(1)$, $\breve{\mathfrak{m}} = \|\tilde{\boldsymbol{c}}\|_0 \lesssim_p R_n$, and all the nonzero $\bar{\boldsymbol{c}}_l$, $1 \leq l \leq r_n$, are selected with probability approaching one. Recall that $\tilde{T}_n = \{l : \|\tilde{\boldsymbol{c}}_l\| > 0, l =$

$1,\ldots,m\}$. Since $\|\tilde{\boldsymbol{c}}\|_0 \lesssim_p R_n$, we conclude that $\mathrm{card}(\tilde{T}_n)$ is bounded by $m_A < CR_n$, for some positive constant C. Without loss of generality, we define $\boldsymbol{\Pi}_{A,t} = (\boldsymbol{P}_{1t}^\top, \ldots, \boldsymbol{P}_{m_A t}^\top)^\top$ and denote $\boldsymbol{c}_A = (\boldsymbol{c}_1^\top, \ldots, \boldsymbol{c}_{m_A}^\top)^\top$ and $\bar{\boldsymbol{c}}_A = (\bar{\boldsymbol{c}}_1^\top, \ldots, \bar{\boldsymbol{c}}_{m_A}^\top)^\top$ as the corresponding sub-vectors. Therefore, we rewrite (11) in the main text as

$$Q_1(\boldsymbol{c}_A) = \frac{1}{n}\sum_{t=m+1}^{n} \rho_\tau \left\{Y_t - \boldsymbol{\Pi}_{A,t}^\top \boldsymbol{c}_A\right\} + \frac{\lambda_{n,1}}{n}\sum_{l=1}^{m_A} \tilde{\omega}_l \|\boldsymbol{c}_l\|_1. \tag{B.29}$$

Denote $\hat{\boldsymbol{c}}_A$ as the minimizer of $Q_1(\boldsymbol{c}_A)$. The definition of $\tilde{\omega}_l$ and the consistency of the first-stage estimator together indicate that there exists two positive constants C_1 and C_2 such that $\tilde{\omega}_l \leq C_1$ for $1 \leq l \leq r_n$, and $\tilde{\omega}_l \geq C_2 \tilde{\boldsymbol{\zeta}}_n^{-1}$ for $r_n + 1 \leq l \leq m_A$. By Assumption A9, one can verify that, the eigenvalues of $n^{-1}\sum_{t=m+1}^{n} \boldsymbol{\Pi}_{A,t}\boldsymbol{\Pi}_{A,t}^\top$ are uniformly bounded away from zero and from infinity, with probability approaching one.

(I) Proof of part (a)

Proof. To show the variable selection consistency, we only need to verify $\hat{\boldsymbol{c}}_l = 0$, for $l = r_n + 1, \ldots, m_A$, which suffices to prove that

$$\begin{cases} \frac{\partial Q_1(\boldsymbol{c}_A)}{\partial c_{l\mathfrak{u}}} < 0, & \text{if } c_{l\mathfrak{u}} < 0 \\ \frac{\partial Q_1(\boldsymbol{c}_A)}{\partial c_{l\mathfrak{u}}} > 0, & \text{if } c_{l\mathfrak{u}} > 0 \end{cases}$$

for $\mathfrak{u} = 1, \ldots, (1+\kappa)K^l$ and $l = r_n + 1, \ldots, m_A$. To this end, denote $\Pi_{A,\imath t}$ as the $\imath$th element of $\boldsymbol{\Pi}_{A,t}$ for $\imath \in \{1, \ldots, (1+\kappa)\sum_{l=r_n+1}^{m_A} K^l\}$, we need to show that

$$\left\| \sum_{t=m+1}^{n} \psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top \hat{\boldsymbol{c}}_A)\Pi_{A,\imath t} \right\| \leq \lambda_{n,1}\tilde{\omega}_l$$

for $l \in \{r_n + 1, \ldots, m_A\}$. Indeed, by Lemma B.1, we have

$$\left\| \sum_{t=m+1}^{n} \psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top \hat{\boldsymbol{c}}_A)\Pi_{A,\imath t} \right\| \leq R_n \max_{m+1\leq t\leq n} |\Pi_{A,\imath t}| = O(R_n).$$

Then, with $\tilde{\boldsymbol{\zeta}}_n R_n^{1/2} \to 0$ and by Assumption A11, one has

$$(\lambda_{n,1}\tilde{\omega}_l)^{-1}\left\| \sum_{t=m+1}^{n} \psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top \hat{\boldsymbol{c}}_A)\Pi_{A,\imath t}\right\| \le C R_n \lambda_{n,1}^{-1}\tilde{\boldsymbol{\zeta}}_n = C\tilde{\boldsymbol{\zeta}}_n R_n^{1/2}(R_n^{-1/2}\lambda_{n,1})^{-1} = o(1),$$

where the first inequality follows from $\tilde{\omega}_l \ge C_2\tilde{\boldsymbol{\zeta}}_n^{-1}$, for $r_n+1 \le l \le m_A$. Thus, $\left\| \sum_{t=m+1}^{n} \psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top \hat{\boldsymbol{c}}_A)\Pi_{A,\imath t}\right\|$ is dominated by $\lambda_{n,1}\tilde{\omega}_l$, for $l = r_n+1,\dots,m_A$. These complete the proof of part (a). □

(II) Proof of part (b)

Proof. Define $\boldsymbol{C}_n = \{\boldsymbol{c}_A : \|\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A\|_2 = C(R_n/n)^{1/2}\}$ for some sufficiently large $C > 0$. Using Knight's identity, we have

$$\begin{aligned} n\{Q_1(\boldsymbol{c}_A) - Q_1(\bar{\boldsymbol{c}}_A)\} = &\sum_{t=m+1}^{n} [\rho_\tau\{Y_t - \boldsymbol{\Pi}_{A,t}^\top \boldsymbol{c}_A\} - \rho_\tau\{Y_t - \boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A\}] \\ &+ \lambda_{n,1}\sum_{l=1}^{m_A}\tilde{\omega}_l(\|\boldsymbol{c}_l\|_1 - \|\bar{\boldsymbol{c}}_l\|_1) \\ \equiv & L_{n1} + L_{n2} + L_{n3} + L_{n4}, \end{aligned}$$

where

$$L_{n1} \equiv \sum_{t=m+1}^{n} \boldsymbol{\Pi}_{A,t}^\top(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)\psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top\bar{\boldsymbol{c}}_A) \equiv \sum_{t=m+1}^{n} L_{nt}^{(1)},$$

$$L_{n2} \equiv \sum_{t=m+1}^{n} E\left[\int_0^{\boldsymbol{\Pi}_{A,t}^\top(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)} I(Y_t - \boldsymbol{\Pi}_{A,t}^\top\bar{\boldsymbol{c}}_A \le z) - I(Y_t - \boldsymbol{\Pi}_{A,t}^\top\bar{\boldsymbol{c}}_A \le 0)dz\right],$$

$$\begin{aligned} L_{n3} \equiv \sum_{t=m+1}^{n} \Bigg\{ &\int_0^{\boldsymbol{\Pi}_{A,t}^\top(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)} I(Y_t - \boldsymbol{\Pi}_{A,t}^\top\bar{\boldsymbol{c}}_A \le z) - I(Y_t - \boldsymbol{\Pi}_{A,t}^\top\bar{\boldsymbol{c}}_A \le 0)dz \\ &- E\left[\int_0^{\boldsymbol{\Pi}_{A,t}^\top(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)} I(Y_t - \boldsymbol{\Pi}_{A,t}^\top\bar{\boldsymbol{c}}_A \le z) - I(Y_t - \boldsymbol{\Pi}_{A,t}^\top\bar{\boldsymbol{c}}_A \le 0)dz\right]\Bigg\} \end{aligned}$$

and

$$L_{n4} \equiv \lambda_{n,1} \sum_{l=1}^{m_A} \tilde{\omega}_l(\|\boldsymbol{c}_l\|_1 - \|\bar{\boldsymbol{c}}_l\|_1).$$

First, we derive the upper bound of L_{n1}. Notice that

$$\begin{aligned} E(L_{n1})^2 &= \sum_{t=m+1}^{n} Var(L_{nt}^{(1)}) + 2\sum_{\ell=1}^{n-1}(1-\frac{\ell}{n})Cov(L_{n1}^{(1)}, L_{n(\ell+1)}^{(1)}) \\ &\leq nVar(L_{n1}^{(1)}) + 2n\sum_{\ell=1}^{d_n-1}|Cov(L_{n1}^{(1)}, L_{n(\ell+1)}^{(1)})| + 2n\sum_{\ell=d_n}^{\infty}|Cov(L_{n1}^{(1)}, L_{n(\ell+1)}^{(1)})| \\ &\equiv J_1 + J_2 + J_3, \end{aligned}$$

with $d_n \to \infty$ specified later. First, we consider the last term, J_3, in the above equation. To this end, using Davydov's inequality (see, e.g., Corollary A.2 of Hall and Heyde (1980)), one has

$$|Cov(L_{n1}^{(1)}, L_{n(\ell+1)}^{(1)})| \leq C\alpha^{1-2/\delta}(\ell)[E|L_{n1}^{(1)}|^\delta]^{2/\delta}. \tag{B.30}$$

Notice that for any $k > 0$, $|\psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A)|^k = |\tau - I(Y_t < \boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A)|^k = |I(Y_t < \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t) - I(Y_t < \boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A)|^k$. Therefore, by Assumption A6, there exists a $C > 0$ such that

$$\begin{aligned} E[|\psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A)|^k | Z_t, \boldsymbol{W}_t] =& E[|I(Y_t < \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t) - I(Y_t < \boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A)|^k | Z_t, \boldsymbol{W}_t] \\ &\leq F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t | Z, \boldsymbol{W}) - F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A | Z, \boldsymbol{W}) \leq C r_n K^{-d}, \end{aligned}$$

which implies that

$$E|L_{n1}^{(1)}|^\delta = E[|\boldsymbol{\Pi}_{A,t}^\top(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)|^\delta |\psi_\tau(Y_t - \boldsymbol{\Pi}_{A,t}^\top \bar{\boldsymbol{c}}_A)|^\delta] \leq C(R_n/n^{1/2})^\delta r_n K^{-d}.$$

This, in conjunction with (B.30), gives that

$$
\begin{aligned}
J_3 \leq & Cn((R_n/n^{1/2})^{\delta} r_n K^{-d})^{2/\delta} \sum_{\ell=d_n}^{\infty} \alpha^{1-2/\delta}(\ell) \leq Cn(R_n^2/n) r_n^{2/\delta} K^{-2d/\delta} d_n^{-w} \sum_{\ell=d_n}^{\infty} \ell^w \alpha^{1-2/\delta}(\ell) \\
= & R_n \cdot o(R_n (r_n K)^{-1} K^{-2d/\delta} r_n^{1-2/\delta} d_n^{-w}) = o(R_n)
\end{aligned}
$$

by choosing d_n to satisfy $d_n^w r_n^{2/\delta - 1} = c$ and by Assumption A2. As for J_2, again by choosing sufficiently large $C > 0$, we use Assumptions A2 and A6 to obtain

$$
|Cov(L_{n1}^{(1)}, L_{n(\ell+1)}^{(1)})| \leq E|L_{n1}^{(1)} L_{n(\ell+1)}^{(1)}| + E|L_{n1}^{(1)}| E|L_{n(\ell+1)}^{(1)}| \leq C(R_n/n)
$$

It follows that $J_2 = O(d_n R_n)$ with $d_n = o(R_n)$. Analogously,

$$
\begin{aligned}
J_1 \leq nE(L_{n1}^{(1)})^2 = & nE[\mathbf{\Pi}_{A,t}^{\top}(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)\psi_{\tau}(Y_t - \mathbf{\Pi}_{A,t}^{\top}\bar{\boldsymbol{c}}_A)]^2 \\
= & nE[(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)^{\top} \mathbf{\Pi}_{A,t} \mathbf{\Pi}_{A,t}^{\top} (\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)(\psi_{\tau}(Y_t - \mathbf{\Pi}_{A,t}^{\top}\bar{\boldsymbol{c}}_A))^2] \\
= & O(R_n^2 r_n K^{-d}) = o(R_n),
\end{aligned}
$$

where the last equality holds due to Assumption A2 for $d \geq 2$. Thus, $L_{n1} = O_p((d_n R_n)^{1/2}) = o_p(R_n)$. Second, we derive the lower bound of L_{n2}. Indeed, let $b_{R,A,t} \equiv \mathbf{\Pi}_{A,t}^{\top}\bar{\boldsymbol{c}}_A - \boldsymbol{\alpha}_{\tau}^{\top}(\boldsymbol{Z}_t)\mathbb{W}_t$, we have

$$
\begin{aligned}
L_{n2} = & \sum_{t=m+1}^{n} \int_0^{\mathbf{\Pi}_{A,t}^{\top}(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)} F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_{\tau}^{\top}(\boldsymbol{Z}_t)\mathbb{W}_t + b_{R,A,t} + z|Z, \boldsymbol{W}) \\
& - F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_{\tau}^{\top}(\boldsymbol{Z}_t)\mathbb{W}_t + b_{R,A,t}|Z, \boldsymbol{W}) dz \\
= & \sum_{t=m+1}^{n} \int_0^{\mathbf{\Pi}_{A,t}^{\top}(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)} f_{Y|Z,\boldsymbol{W}}(0|Z, \boldsymbol{W}) z dz (1 + o_p(1)) \\
\geq & C \sum_{t=m+1}^{n} \frac{1}{2} \{\mathbf{\Pi}_{A,t}^{\top}(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)\}^2 = C\frac{1}{2}(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A)^{\top} \sum_{t=m+1}^{n} \mathbf{\Pi}_{A,t}\mathbf{\Pi}_{A,t}^{\top}(\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A) \\
\geq & Cn\|\boldsymbol{c}_A - \bar{\boldsymbol{c}}_A\|_2^2 = O(R_n),
\end{aligned}
$$

where the second equality follows from mean value theorem and Assumption A3.

Furthermore, we derive the upper bound of L_{n3}. Notice that

$$\begin{aligned} Var(L_{n3}) \leq & E\Bigg[\Bigg(\sum_{t=m+1}^{n}\int_{0}^{\boldsymbol{\Pi}_{A,t}^{\top}(\boldsymbol{c}_A-\bar{\boldsymbol{c}}_A)} I(Y_t-\boldsymbol{\Pi}_{A,t}^{\top}\bar{\boldsymbol{c}}_A\leq z)-I(Y_t-\boldsymbol{\Pi}_{A,t}^{\top}\bar{\boldsymbol{c}}_A\leq 0)dz\Bigg)^2\Bigg] \\ \leq & E\Bigg[\Bigg(\sum_{t=m+1}^{n}(\boldsymbol{\Pi}_{A,t}^{\top}(\boldsymbol{c}_A-\bar{\boldsymbol{c}}_A)\Bigg)^2\Bigg]. \end{aligned}$$

Then, similar to the derivation of bounding L_{n1}, one can show that $L_{n3}=o_p(R_n)$. We omit the detailed proof here. Finally, we consider L_{n4}. Clearly,

$$\begin{aligned} L_{n4} \geq & \lambda_{n,1}\sum_{l=1}^{r_n}\tilde{\omega}_l(\|\boldsymbol{c}_l\|_1-\|\bar{\boldsymbol{c}}_l\|_1)\geq -\lambda_{n,1}\sum_{l=1}^{r_n}\tilde{\omega}_l(\|\boldsymbol{c}_l-\bar{\boldsymbol{c}}_l\|_1) \\ \geq & -C\lambda_{n,1}R_n^{1/2}\sum_{l=1}^{r_n}\|\boldsymbol{c}_l-\bar{\boldsymbol{c}}_l\|\geq -C\lambda_{n,1}R_n^{1/2}\|\boldsymbol{c}_A-\bar{\boldsymbol{c}}_A\| \\ = & CR_n n^{-1/2}\lambda_{n,1}=o(R_n) \end{aligned}$$

by Assumption A11. Combining the above results yields $P(\inf_{\boldsymbol{c}_A\in\boldsymbol{C}_n} n\{Q_1(\boldsymbol{c}_A)-Q_1(\bar{\boldsymbol{c}}_A)\} > CR_n)\to 1$. By the convexity of $Q_1(\boldsymbol{c}_A)-Q_1(\bar{\boldsymbol{c}}_A)$, we have that for any $\epsilon>0$, there exists a constant $C>0$ such that $P\big(\|\hat{\boldsymbol{c}}_A-\bar{\boldsymbol{c}}_A\|\leq C(R_n/n)^{1/2}\big)>1-\epsilon$. Therefore, the proof of (b) is completed. □

(III) Proof of part (c)

Proof. The proof of part (c) follows by combining part (b) with $\max_{m+1\leq t\leq n}\|\boldsymbol{\Pi}_{A,t}\|=O(R_n^{1/2})$, because

$$\max_{m+1\leq t\leq n}|\hat{q}_{\tau,t}-q_{\tau,t}|\leq \max_{m+1\leq t\leq n}\|\boldsymbol{\Pi}_{A,t}\|\|\hat{\boldsymbol{c}}_A-\bar{\boldsymbol{c}}_A\|=O_p(R_n/n^{1/2}).$$

□

Appendix C: Mathematical Proofs of Theorem 3

In this section, we give certain lemmas with their detailed proofs that are useful for proving the Theorem 3 in the main article. Of course, notations and assumptions that are used here are the same as those in the main article. Also note that C and M are denoted as generic constants that may vary across occurrences.

C.1 Some Lemmas for Proving Theorem 3:

Lemma C.1. *Let $\hat{\beta}$ be the minimizer of the function $\sum_{t=1}^{n} \omega_t \rho_\tau(Y_t - X_t^\top \beta)$, where $\omega_t > 0$. Then, $\| \sum_{t=1}^{n} \omega_t X_t \psi_\tau(Y_t - X_t^\top \hat{\beta})\| \leq \dim(X) \max_{t \leq n} \|\omega_t X_t\|$.*

Proof. The proof follows from that of Lemma A.2 in Ruppert and Carroll (1980). See also Lemma A.2 in Cai and Xu (2008). □

To obtain Bahadur results given in Lemma C.8 (below), we need to introduce some notations. In Lemmas C.2 - C.7, τ is dropped from $\boldsymbol{c}_\tau$ for simplicity and we use the notation E_t to represent the conditional expectation $E\{\cdot|Z, \boldsymbol{W}\} \equiv E\{\cdot|\mathcal{F}_{t-1}\}$. Let $\boldsymbol{\Pi}_{a,t} = (\boldsymbol{P}_{1t}^\top, \ldots, \boldsymbol{P}_{r_n t}^\top)^\top$ be the submatrix consisting of the first r_n compositions of $\boldsymbol{\Pi}_t = (\boldsymbol{P}_{1t}^\top, \ldots, \boldsymbol{P}_{mt}^\top)^\top$ corresponding to the active covariates. Without loss of generality, we set the first r_n compositions of $\hat{\boldsymbol{c}}_A$ be non-zero, that is, $\hat{\boldsymbol{c}}_A = (\hat{\boldsymbol{c}}_1^\top, \ldots, \hat{\boldsymbol{c}}_{r_n}^\top, \boldsymbol{0}^\top)^\top \equiv (\hat{\boldsymbol{c}}_a^\top, \boldsymbol{0}^\top)^\top$. Similarly, let $\bar{\boldsymbol{c}}_A = (\bar{\boldsymbol{c}}_1^\top, \ldots, \bar{\boldsymbol{c}}_{r_n}^\top, \boldsymbol{0}^\top)^\top \equiv (\bar{\boldsymbol{c}}_a^\top, \boldsymbol{0}^\top)^\top$ have the same definition as that in the Appendix B. In addition, let $Y_t^* \equiv Y_t - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t$ and $b_{R,t} \equiv \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t - \boldsymbol{\Pi}_{a,t}^\top \bar{\boldsymbol{c}}_a$. The oracle property in Theorem 2 implies that $\hat{\boldsymbol{c}}_A = (\hat{\boldsymbol{c}}_a^\top, \boldsymbol{0}^\top)^\top$ is a minimizer of (B.29) in the Appendix B. Then, define

$$
\begin{aligned}
G_n(\boldsymbol{c}_a) =& \frac{1}{n} \sum_{t=m+1}^{n} \{\tau - I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top(\boldsymbol{c}_a - \bar{\boldsymbol{c}}_a) - b_{R,t})\}\boldsymbol{\Pi}_{a,t} + \frac{\lambda_{n,1,\omega}(\boldsymbol{c}_a)}{n} \\
\equiv& G_{0n}(\boldsymbol{c}_a) + \frac{\lambda_{n,1,\omega}(\boldsymbol{c}_a)}{n},
\end{aligned}
$$

where

$$G_{0n}(\boldsymbol{c}_a) = \frac{1}{n}\sum_{t=m+1}^{n}\{\tau - I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top(\boldsymbol{c}_a - \bar{\boldsymbol{c}}_a) - b_{R,t})\}\boldsymbol{\Pi}_{a,t}$$

and $\lambda_{n,1,\omega}(\boldsymbol{c}_a) = (\lambda_{n,1}\tilde{\omega}_1\text{sgn}(\boldsymbol{c}_1^\top), \ldots, \lambda_{n,1}\tilde{\omega}_{r_n}\text{sgn}(\boldsymbol{c}_{r_n}^\top))^\top$. Similarly, define

$$G_n^*(\boldsymbol{c}_a) = \frac{1}{n}\sum_{t=m+1}^{n}\{\tau - F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t + \boldsymbol{\Pi}_{a,t}^\top(\boldsymbol{c}_a - \bar{\boldsymbol{c}}_a) - b_{R,t}|Z, \boldsymbol{W})\}\boldsymbol{\Pi}_{a,t}$$

and $\tilde{G}_n(\boldsymbol{c}_a) = G_n(\boldsymbol{c}_a) - G_n^*(\boldsymbol{c}_a)$.

Lemma C.2. *Under Assumption A1–A13, we have*

$$\|\tilde{G}_n(\bar{\boldsymbol{c}}_a)\| = O_p((R_n/n)^{1/2}).$$

Proof. Notice that by Assumption A11 and by the fact that $\tilde{\omega}_l \le C$ for $1 \le l \le r_n$ and some constant $C > 0$, we have $\|n^{-1}\lambda_{n,1,\omega}(\bar{\boldsymbol{c}}_a)\| = o(n^{-1/2})$. Then, by triangle inequality, $\|\tilde{G}_n(\bar{\boldsymbol{c}}_a)\| \le \|G_{0n}(\bar{\boldsymbol{c}}_a) - G_n^*(\bar{\boldsymbol{c}}_a)\| + o(n^{-1/2})$. Following the proof of Lemma A.3 in Horowitz and Lee (2005), one can show that $\|G_{0n}(\bar{\boldsymbol{c}}_a) - G_n^*(\bar{\boldsymbol{c}}_a)\| = O_p((R_n/n)^{1/2})$ and the lemma is proved. □

Lemma C.3. *Under Assumption A1–A13, we have*

$$\|G_n(\hat{\boldsymbol{c}}_a)\| = O(R_n^{3/2}/n).$$

Proof. Note that

$$\begin{aligned}
\|G_n(\hat{\boldsymbol{c}}_a)\| \le& \|G_{0n}(\hat{\boldsymbol{c}}_a)\| + n^{-1}\|\lambda_{n,1,\omega}(\hat{\boldsymbol{c}}_a)\| \\
\le& \left\|\frac{1}{n}\sum_{t=m+1}^{n}\{\tau - I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) - b_{R,t})\}\boldsymbol{\Pi}_{a,t}\right\| + n^{-1}\|\lambda_{n,1,\omega}(\hat{\boldsymbol{c}}_a)\| \\
=& \left\|\frac{1}{n}\sum_{t=m+1}^{n}\{\tau - I(Y_t < \boldsymbol{\Pi}_{a,t}^\top\hat{\boldsymbol{c}}_a)\}\boldsymbol{\Pi}_{a,t}\right\| + n^{-1}\|\lambda_{n,1,\omega}(\hat{\boldsymbol{c}}_a)\| \\
\le& n^{-1}R_n \max_{m+1\le t\le n}\|\boldsymbol{\Pi}_{a,t}\| + n^{-1}\|\lambda_{n,1,\omega}(\hat{\boldsymbol{c}}_a)\| \quad \text{(by Lemma B.1 in Appendix B)} \\
=& O(R_n^{3/2}/n),
\end{aligned}$$

where we use the fact that $\|n^{-1}\lambda_{n,1,\omega}(\hat{\boldsymbol{c}}_a)\| = o(n^{-1/2})$. □

Lemma C.4. *Under Assumption A1–A13, we have*

$$\sup_{\|\hat{\boldsymbol{c}}_a-\bar{\boldsymbol{c}}_a\|\leq(R_n/n)^{1/2}}\|\tilde{G}_n(\hat{\boldsymbol{c}}_a)-\tilde{G}_n(\bar{\boldsymbol{c}}_a)\|=o_p((R_n/n)^{1/2}).$$

Proof. Define $\hat{\boldsymbol{\vartheta}}\equiv\hat{\boldsymbol{c}}_a-\bar{\boldsymbol{c}}_a$ and $\eta_t(\hat{\boldsymbol{\vartheta}})=\boldsymbol{\Pi}_{a,t}\{I(Y_t^*<\boldsymbol{\Pi}_{a,t}^\top\hat{\boldsymbol{\vartheta}}-b_{R,t})-I(Y_t^*<-b_{R,t})\}$. Clearly,

$$\begin{aligned}E_t\{\eta_t(\hat{\boldsymbol{\vartheta}})\}=&\boldsymbol{\Pi}_{a,t}\{F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t+\boldsymbol{\Pi}_{a,t}^\top\hat{\boldsymbol{\vartheta}}-b_{R,t}|Z,\boldsymbol{W})\\&+F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t-b_{R,t}|Z,\boldsymbol{W})\}.\end{aligned}$$

Notice that

$$\begin{aligned}\tilde{G}_n(\hat{\boldsymbol{c}}_a)-\tilde{G}_n(\bar{\boldsymbol{c}}_a)=&\frac{1}{n}\sum_{t=m+1}^n\{\tau-I(Y_t^*<\boldsymbol{\Pi}_{a,t}^\top(\hat{\boldsymbol{c}}_a-\bar{\boldsymbol{c}}_a)-b_{R,t})\}\boldsymbol{\Pi}_{a,t}\\&-\frac{1}{n}\sum_{t=m+1}^n\{\tau-F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t+\boldsymbol{\Pi}_{a,t}^\top(\hat{\boldsymbol{c}}_a-\bar{\boldsymbol{c}}_a)-b_{R,t}|Z,\boldsymbol{W})\}\boldsymbol{\Pi}_{a,t}\\&-\frac{1}{n}\sum_{t=m+1}^n\{\tau-I(Y_t^*<-b_{R,t})\}\boldsymbol{\Pi}_{a,t}\\&+\frac{1}{n}\sum_{t=m+1}^n\{\tau-F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t-b_{R,t}|Z,\boldsymbol{W})\}\boldsymbol{\Pi}_{a,t}\\&+\frac{\lambda_{n,1,\omega}(\hat{\boldsymbol{c}}_a)}{n}-\frac{\lambda_{n,1,\omega}(\bar{\boldsymbol{c}}_a)}{n}\\=&\frac{1}{n}\sum_{t=m+1}^n[\eta_t(\hat{\boldsymbol{\vartheta}})-E_t\{\eta_t(\hat{\boldsymbol{\vartheta}})\}]+o(n^{-1/2}),\end{aligned}$$

where the last equality holds due to Assumption A11. To finish the proof, it suffices to show that, for any $a\in\{a\in\mathbb{R}^{R_n}:\|a\|=1\}$,

$$\sup_{\|\hat{\boldsymbol{\vartheta}}\|\leq(R_n/n)^{1/2}}\left|\sum_{t=m+1}^n a^\top[\eta_t(\hat{\boldsymbol{\vartheta}})-E_t\{\eta_t(\hat{\boldsymbol{\vartheta}})\}]\right|=o_p((nR_n)^{1/2}).$$

Similar to the proof in Xiao and Koenker (2009), covering the ball $\{\|\hat{\boldsymbol{\vartheta}}\|\leq C(R_n/n)^{1/2}\}$ with cubes $\mathcal{C}=\{\mathcal{C}_k\}$, where $\mathcal{C}_k$ is a cube with center $\hat{\boldsymbol{\vartheta}}_k$ and side length $C(R_n/n^5)^{1/2}$, so that $N(n)=\#(\mathcal{C})=(2n^2)^{R_n}$. Therefore, because for $\hat{\boldsymbol{\vartheta}}\in\mathcal{C}_k$, $\|\hat{\boldsymbol{\vartheta}}-\hat{\boldsymbol{\vartheta}}_k\|\leq C(R_n/n^{5/2})$ and $I(Y_{\mathfrak{s}}^*<x)$ is nondecreasing in x,

$$
\begin{aligned}
&\sup_{\|\hat{\boldsymbol{\vartheta}}\|\le C(R_n/n)^{1/2}} \left| \sum_{t=m+1}^{n} a^\top[\eta_t(\hat{\boldsymbol{\vartheta}}) - E_t\{\eta_t(\hat{\boldsymbol{\vartheta}})\}] \right| \\
\le & \max_{1\le k\le N(n)} \left| \sum_{t=m+1}^{n} a^\top[\eta_t(\hat{\boldsymbol{\vartheta}}_k) - E_t\{\eta_t(\hat{\boldsymbol{\vartheta}}_k)\}] \right| \\
& + \max_{1\le k\le N(n)} \left| \sum_{t=m+1}^{n} |(a^\top \boldsymbol{\Pi}_{a,t})| \{b_{nt}(\hat{\boldsymbol{\vartheta}}_k) - E_t(b_{nt}(\hat{\boldsymbol{\vartheta}}_k))\} \right| \\
& + \max_{1\le k\le N(n)} \left| \sum_{t=m+1}^{n} |(a^\top \boldsymbol{\Pi}_{a,t})| \{E_t(d_{nt}(\hat{\boldsymbol{\vartheta}}_k))\} \right| \\
\equiv & M_1 + M_2 + M_3,
\end{aligned}
$$

where $b_{nt}(\hat{\boldsymbol{\vartheta}}_k) = I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t}) - I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t} + C(R_n/n^{5/2})\|\boldsymbol{\Pi}_{a,t}\|)$ and $d_{nt}(\hat{\boldsymbol{\vartheta}}_k) = I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t} + C(R_n/n^{5/2})\|\boldsymbol{\Pi}_{a,t}\|) - I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t} - C(R_n/n^{5/2})\|\boldsymbol{\Pi}_{a,t}\|)$. The analyses of M_2 and M_3 are similar to those in Welsh (1989) and Xiao and Koenker (2009), so that our focus here is only on M_1. Notice, for any $\flat > 0$, $|I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t}) - I(Y_t^* < -b_{R,t})|^\flat = I(d_{3t} < Y_t \le d_{4t})$, where $d_{3t} = \min(c_{2t}, c_{2t} + c_{3t})$ and $d_{4t} = \max(c_{2t}, c_{2t} + c_{3t})$ with $c_{2t} = -b_{R,t}$ and $c_{3t} = \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k$. Therefore, by Assumption A6, there exists a $C > 0$ such that $E\{|I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t}) - I(Y_t^* < -b_{R,t})|^\flat | Z_t, \boldsymbol{W}_t\} = F_{Y|Z,\boldsymbol{W}}(d_{4t}) - F_{Y|Z,\boldsymbol{W}}(d_{3t}) \le C|\boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k| \le C(R_n/n)^{1/2}\|\boldsymbol{\Pi}_{a,t}\|$, which implies that

$$
\begin{aligned}
E_t[a^\top \eta_t(\hat{\boldsymbol{\vartheta}}_k)]^2 =& E_t[|a^\top \boldsymbol{\Pi}_{a,t}|^2 |I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t}) - I(Y_t^* < -b_{R,t})|^2] \\
\le & C(R_n/n)^{1/2} R_n^{1/2} \left[\frac{1}{n} \sum_{t=m+1}^{n} a^\top \boldsymbol{\Pi}_{a,t} \boldsymbol{\Pi}_{a,t}^\top a \right] \le C((R_n/n)^{1/2} R_n^{1/2}).
\end{aligned}
$$

where the last inequality holds due to the boundedness of eigenvalues of $n^{-1}\sum_{t=m+1}^{n} \boldsymbol{\Pi}_{a,t}\boldsymbol{\Pi}_{a,t}^\top$. Thus, we have

$$
\mathcal{W}_n^2 = \sum_{t=m+1}^{n} E_t[a^\top \{\eta_t(\hat{\boldsymbol{\vartheta}}_k) - E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}]^2 \le \sum_{t=m+1}^{n} E_t[a^\top \eta_t(\hat{\boldsymbol{\vartheta}}_k)]^2 = O((nR_n)^{1/2} R_n^{1/2})
$$

and

$$\mathcal{S}_n^2 = \sum_{t=m+1}^{n} [a^\top \{\eta_t(\hat{\boldsymbol{\vartheta}}_k) - E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}]^2 = O_p((nR_n)^{1/2} R_n^{1/2}).$$

Also, notice that $\xi_t(\hat{\boldsymbol{\vartheta}}_k) = \{\eta_t(\hat{\boldsymbol{\vartheta}}_k) - E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}$ is a martingale difference sequence. Therefore, let $L = (nR_n)^{1/2}$. Thus, we have

$$\begin{aligned}
& P\left[\max_{1\le k\le N(n)} \left|\frac{1}{\sqrt{nR_n}} \sum_{t=m+1}^{n} \{a^\top \{\eta_t(\hat{\boldsymbol{\vartheta}}_k) - E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}\}\right| > \epsilon\right] \\
\le\ & N(n) \max_k P\left[\left|\frac{1}{\sqrt{nR_n}} \sum_{t=m+1}^{n} \{a^\top \{\eta_t(\hat{\boldsymbol{\vartheta}}_k) - E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}\}\right| > \epsilon\right] \\
\le\ & N(n) \max_k P\left[\left|\sum_{t=m+1}^{n} a^\top \xi_t(\hat{\boldsymbol{\vartheta}}_k)\right| > \sqrt{nR_n}\epsilon, \mathcal{W}_n^2 + \mathcal{S}_n^2 \le L\right] \\
& + N(n) \max_k P\left[\left|\sum_{t=m+1}^{n} a^\top \xi_t(\hat{\boldsymbol{\vartheta}}_k)\right| > \sqrt{nR_n}\epsilon, \mathcal{W}_n^2 + \mathcal{S}_n^2 > L\right] \equiv I_1 + I_2. \qquad \text{(C.1)}
\end{aligned}$$

For I_1, by exponential inequality for martingale difference sequences (see, e.g., Bercu and Touati, 2008), we have

$$\begin{aligned}
& N(n) \max_k P\left[\left|\sum_{t=m+1}^{n} a^\top \xi_t(\hat{\boldsymbol{\vartheta}}_k)\right| > \sqrt{nR_n}\epsilon, \mathcal{W}_n^2 + \mathcal{S}_n^2 \le L\right] \\
\le & 2N(n) \exp\left(-\frac{(nR_n)\epsilon^2}{2L}\right).
\end{aligned}$$

For I_2, because $P[\mathcal{W}_n^2 + \mathcal{S}_n^2 > L] \le P[\mathcal{W}_n^2 > L/2] + P[\mathcal{S}_n^2 > L/2]$ and each term can be bounded exponentially under Assumptions A1, A7 and A8. Thus, $M_1 = o_p((nR_n)^{1/2})$. This implies that $\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a) = o_p(n^{-1}(nR_n)^{1/2}) = o_p((R_n/n)^{1/2})$. □

Lemma C.5. *Under Assumption A1–A13, we have*

$$\sup_{\|\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a\| \le (R_n/n)^{1/2}} \|G_n(\hat{\boldsymbol{c}}_a) - G_n^*(\hat{\boldsymbol{c}}_a)\| = O_p((R_n/n)^{1/2}).$$

Proof. Using triangle inequality, write

$$\sup_{\|\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a\| \le (R_n/n)^{1/2}} \|G_n(\hat{\boldsymbol{c}}_a) - G_n^*(\hat{\boldsymbol{c}}_a)\| \le \sup_{\|\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a\| \le (R_n/n)^{1/2}} \|\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)\| + \|\tilde{G}_n(\bar{\boldsymbol{c}}_a)\|.$$

Then, the desired result follows immediately from Lemma C.2 and Lemma C.4. □

Define $\boldsymbol{\Phi}_{n,a} = n^{-1}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t}\boldsymbol{\Pi}_{a,t}^\top$, $\boldsymbol{\Phi}_a = E[f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t}\boldsymbol{\Pi}_{a,t}^\top]$ and $\zeta_R = \max_{m+1\le t\le n}\|\boldsymbol{\Pi}_{a,t}\|$. Note that $\zeta_R = O(R_n^{1/2})$. In addition, by Assumption A9, the smallest eigenvalue of $\boldsymbol{\Phi}_a$ is bounded away from 0, and the largest eigenvalue of $\boldsymbol{\Phi}_a$ is bounded from above. Using Davydov's inequality and similar to the proof strategy of Lemma 4 in Horowitz and Mammen (2004), we can argue that the $\|\boldsymbol{\Phi}_{n,a} - \boldsymbol{\Phi}_a\|^2 = O_p(R_n^2/n) = o_p(1)$.

Lemma C.6. *Under Assumption A1–A13, we have*

$$G_n^*(\hat{\boldsymbol{c}}_a) = -\boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) + \frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} + \boldsymbol{R}_n^*,$$

where $\boldsymbol{R}_n^* = O(\zeta_R\|\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a\|^2) + O(\zeta_R r_n^2 K^{-2d})$.

Proof. Using a first-order Taylor series expansion and by Assumption A9, for sufficiently large constant $C > 0$, we have

$$\begin{aligned}
G_n^*(\hat{\boldsymbol{c}}_a) =& \frac{1}{n}\sum_{t=m+1}^{n}\{\tau - F_{Y|Z,\boldsymbol{W}}(\boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t + \boldsymbol{\Pi}_{a,t}^\top(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) - b_{R,t}|Z, \boldsymbol{W})\}\boldsymbol{\Pi}_{a,t} \\
=& \frac{1}{n}\sum_{t=m+1}^{n}\{F_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t)|Z, \boldsymbol{W}) - F_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t) + \boldsymbol{\Pi}_{a,t}^\top(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) - b_{R,t}|Z, \boldsymbol{W})\}\boldsymbol{\Pi}_{a,t} \\
=& -\boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) + \frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} \\
&+ C\max_{m+1\le t\le n}\|\boldsymbol{\Pi}_{a,t}\| n^{-1}\sum_{t=m+1}^{n}\{\boldsymbol{\Pi}_{a,t}^\top(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) - b_{R,t}\}^2
\end{aligned}$$

$$
\begin{aligned}
&= -\boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) + \frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} \\
&\quad + C\zeta_R\left\{(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a)^\top \boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) + \max_{m+1\le t\le n} b_{R,t}^2\right\} \\
&= -\boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) + \frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} \\
&\quad + C\zeta_R \lambda_{\max}(\boldsymbol{\Phi}_{n,a})(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a)^\top(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) + C\zeta_R \max_{m+1\le t\le n} b_{R,t}^2 \\
&= -\boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) + \frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} \\
&\quad + O(\zeta_R \|\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a\|^2) + O(\zeta_R r_n^2 K^{-2d}).
\end{aligned}
$$

Then, the proof is finished. □

Lemma C.7. *Under Assumption A1–A13, we have*

$$
\left\|\frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t}\right\| = O(r_n K^{-d})
$$

Proof. This can be verified by direct calculation and by Assumption A3. □

Lemma C.8. *(Bahadur representation) Under Assumptions A1 – A13, one has,*

$$
\begin{aligned}
\hat{\boldsymbol{\vartheta}} \equiv \hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a =& n^{-1}\boldsymbol{\Phi}_a^{-1}\sum_{t=m+1}^{n} \psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t} \\
&+ n^{-1}\boldsymbol{\Phi}_a^{-1}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} + \boldsymbol{R}_n + o_p(n^{-1/2}),
\end{aligned}
$$

where $\boldsymbol{R}_n$ satisfies

$$
\|\boldsymbol{R}_n\| = O_p((R_n/n)^{1/2} + R_n^{1/2} r_n^2 K^{-2d}).
$$

Proof. Write

$$
G_n(\hat{\boldsymbol{c}}_a) = \tilde{G}_n(\bar{\boldsymbol{c}}_a) + [\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)] + G_n^*(\hat{\boldsymbol{c}}_a). \tag{C.2}
$$

By Lemma C.6, (C.2) can be rewritten as

$$\begin{aligned}\boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) = &- G_n(\hat{\boldsymbol{c}}_a) + \tilde{G}_n(\bar{\boldsymbol{c}}_a) + [\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)] \\ &+ \frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} + \boldsymbol{R}_n^*.\end{aligned} \tag{C.3}$$

Applying Lemma C.3 and Lemma C.4, we have

$$\boldsymbol{\Phi}_{n,a}(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) = \tilde{G}_n(\bar{\boldsymbol{c}}_a) + \frac{1}{n}\sum_{t=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_t, \boldsymbol{W}_t))\boldsymbol{\Pi}_{a,t} b_{R,t} + \boldsymbol{R}_n,$$

where $\boldsymbol{R}_n$ satisfies

$$\|\boldsymbol{R}_n\| \leq \|G_n(\hat{\boldsymbol{c}}_a)\| + \|\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)\| + \|\boldsymbol{R}_n^*\| = O_p((R_n/n)^{1/2} + R_n^{1/2} r_n^2 K^{-2d}).$$

Define

$$\bar{G}_n(\bar{\boldsymbol{c}}_a) = n^{-1}\sum_{t=m+1}^{n}\{\tau - I(Y_t^* < 0)\}\boldsymbol{\Pi}_{a,t} \equiv n^{-1}\sum_{t=m+1}^{n}\psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t}.$$

Using arguments similar to those in the proof of Lemma C.2, we have

$$E[\|\tilde{G}_n(\bar{\boldsymbol{c}}_a) - \bar{G}_n(\bar{\boldsymbol{c}}_a)\|^2 | Z, \boldsymbol{W}] \leq C n^{-1} R_n \max_t |b_{R,t}| + \|n^{-1}\lambda_{n,1,\omega}(\bar{\boldsymbol{c}}_a)\|^2.$$

Hence, by Markov's inequality,

$$\|\tilde{G}_n(\bar{\boldsymbol{c}}_a) - \bar{G}_n(\bar{\boldsymbol{c}}_a)\| = o_p(n^{-1/2}).$$

Then, the lemma follows from $\|\boldsymbol{\Phi}_{n,a} - \boldsymbol{\Phi}_a\|^2 = O_p(R_n^2/n) = o_p(1)$. □

Lemma C.9. *For some $0 < M < \infty$ and let $\mathfrak{L}_n = (R_n/n)^{1/2}$, define $K_{n\mathfrak{L}} = \{(\boldsymbol{\theta}, \boldsymbol{\vartheta}) : \|\boldsymbol{\vartheta}\| \leq \mathfrak{L}_n, \|\boldsymbol{\theta}\| \leq M\}$, let $V_n(\boldsymbol{\vartheta})$ and $V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ be vectors that satisfy (i) $-\boldsymbol{\theta}^\top V_n(\lambda\boldsymbol{\theta}, \boldsymbol{\vartheta}) \geq -\boldsymbol{\theta}^\top V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta})$*

for $\lambda \geq 1$ *and* $\|\boldsymbol{\vartheta}\| \leq \mathfrak{L}_n$*, and (ii)*

$$\sup_{(\boldsymbol{\theta},\boldsymbol{\vartheta})\in K_{n\mathfrak{L}}} \|V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) + V_n(\boldsymbol{\vartheta}) + D\boldsymbol{\theta} - A_n\| = o_p(1),$$

where $\|A_n\| = O_p(1)$ *and* D *is a positive-definite matrix. Suppose that* $\boldsymbol{\theta}_n$ *and* $\boldsymbol{\vartheta}_n$ *are vectors such that* $\|V_n(\boldsymbol{\theta}_n,\boldsymbol{\vartheta}_n)\| = o_p(1)$ *and* $\|V_n(\boldsymbol{\vartheta}_n)\| = O_p(1)$*. Then, one has* $\|\boldsymbol{\theta}_n\| = O_p(1)$ *and* $\boldsymbol{\theta}_n = D^{-1}(A_n - V_n(\boldsymbol{\vartheta}_n)) + o_p(1)$*.*

Proof. The proof follows from Koenker and Zhao (1996) and Conditions (i) and (ii) that $V_n(\boldsymbol{\theta}_n,\boldsymbol{\vartheta}_n) + V_n(\boldsymbol{\vartheta}_n) + D\boldsymbol{\theta}_n - A_n = o_p(1)$. This completes the proof. □

To show Lemmas C.10 and C.11 later, τ is dropped from $\boldsymbol{g}_\tau(z_0)$. For the notational convenience again, define $b_n = (nh)^{-1/2}$ and $\hat{q}_t - q_t = \boldsymbol{\Pi}_{a,t}^\top(\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a) - b_{R,t} \equiv \boldsymbol{\Pi}_{a,t}^\top\hat{\boldsymbol{\vartheta}} - b_{R,t}$. Furthermore, let $\boldsymbol{\theta}_0 = b_n^{-1}(\Theta_0 - \boldsymbol{g}(z_0))$ and $\boldsymbol{\theta}_1 = hb_n^{-1}(\Theta_1 - \boldsymbol{g}^{(1)}(z_0))$. Then, $\boldsymbol{\theta} = b_n^{-1}\boldsymbol{H}\begin{pmatrix}\Theta_0 - \boldsymbol{g}(z_0)\\ \Theta_1 - \boldsymbol{g}^{(1)}(z_0)\end{pmatrix}$, where $\boldsymbol{H} = \text{diag}\{I_{2\kappa+1}, hI_{2\kappa+1}\}$. For convenience of analysis, we rewrite $\hat{\boldsymbol{X}}_t \equiv \boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}}) \equiv \boldsymbol{X}_t(q_t + \boldsymbol{\Pi}_{a,t}^\top\hat{\boldsymbol{\vartheta}} - b_{R,t})$ because it contains $\hat{q}_t$. Similarly, $\boldsymbol{X}_t(\boldsymbol{\vartheta}) \equiv \boldsymbol{X}_t(q_t + \boldsymbol{\Pi}_{a,t}^\top\boldsymbol{\vartheta} - b_{R,t})$, $\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) \equiv \boldsymbol{X}_t^*(q_t + \boldsymbol{\Pi}_{a,t}^\top\boldsymbol{\vartheta} - b_{R,t})$ and $\hat{\boldsymbol{X}}_t^* \equiv \boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}) \equiv \boldsymbol{X}_t^*(q_t + \boldsymbol{\Pi}_{a,t}^\top\hat{\boldsymbol{\vartheta}} - b_{R,t})$, where $\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) = \begin{pmatrix}\boldsymbol{X}_t(\boldsymbol{\vartheta})\\ z_{th}\boldsymbol{X}_t(\boldsymbol{\vartheta})\end{pmatrix}$ and $\boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}) = \begin{pmatrix}\boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}})\\ z_{th}\boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}})\end{pmatrix}$ and $z_{th} = (Z_t - z_0)/h$. Of course, $\boldsymbol{X}_t^*(0) \equiv \boldsymbol{X}_t^* = \begin{pmatrix}\boldsymbol{X}_t\\ z_{th}\boldsymbol{X}_t\end{pmatrix}$. Hence, $\partial\boldsymbol{X}_t(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta} = \boldsymbol{\Upsilon}_{a,t}$, where $\boldsymbol{\Upsilon}_{a,t}^\top = (0_{1\times R_n}^\top, \boldsymbol{\Pi}_{a,t}, \ldots, \boldsymbol{\Pi}_{a,t}, 0_{\kappa\times R_n}^\top)$. Next, denote $v_t^*(\boldsymbol{\vartheta}) = Y_t - \boldsymbol{X}_t^\top(\boldsymbol{\vartheta})[\boldsymbol{g}(z_0) + \boldsymbol{g}^{(1)}(z_0)(Z_t - z_0)]$, $v_t^*(0) = Y_t - \boldsymbol{X}_t^\top[\boldsymbol{g}(z_0) + \boldsymbol{g}^{(1)}(z_0)(Z_t - z_0)]$ and $v_{nt}^* = v_{nt}^*(\boldsymbol{\theta},\boldsymbol{\vartheta}) = v_t^*(\boldsymbol{\vartheta}) - b_n\boldsymbol{\theta}^\top\boldsymbol{X}_t^*(\boldsymbol{\vartheta})$. In addition, define $\Gamma^*(Z_t) = E[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t))\boldsymbol{X}_t^*\boldsymbol{g}_\tau(z_0)^\top\boldsymbol{\Upsilon}_{a,t}|Z_t]$ and $\Gamma(Z_t) = E[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t))\boldsymbol{X}_t\boldsymbol{g}_\tau(z_0)^\top\boldsymbol{\Upsilon}_{a,t}|Z_t]$. Again, let $A_m = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq M\}$ for some $0 < M < \infty$ and $B_m = \{\boldsymbol{\vartheta} : \|\boldsymbol{\vartheta}\| \leq \mathfrak{L}_n\}$ for $\mathfrak{L}_n = (R_n/n)^{1/2}$, Therefore,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{n} \rho_\tau(v_t^*(\hat{\boldsymbol{\vartheta}}) - b_n\boldsymbol{\theta}^\top\boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}))K(z_{th}) \equiv \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}).$$

Now, define vector functions of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$

$$V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = b_n \sum_{t=1}^{n} \psi_\tau(v_t^*(\boldsymbol{\vartheta}) - b_n \boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta})) \boldsymbol{X}_t^*(\boldsymbol{\vartheta}) K(z_{th}),$$

and

$$V_n(\boldsymbol{\vartheta}) = b_n \sum_{t=1}^{n} \Gamma^*(Z_t) \boldsymbol{\vartheta} K(z_{th}),$$

where $\psi_\tau(x) = \tau - I(x < 0)$. In the next three lemmas, we show that $V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ and $V_n(\boldsymbol{\vartheta})$ satisfy Lemma C.9, so that we can derive the local Bahadur representation for $\hat{\boldsymbol{\theta}}$.

Lemma C.10. *Under the assumptions in Theorem 3, one has*

$$\sup_{\boldsymbol{\vartheta} \in B_m, \boldsymbol{\theta} \in A_m} \| V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta}) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta})] \| = o_p(1).$$

Proof. For any $\boldsymbol{\theta} \in A_m$ and for any $\boldsymbol{\vartheta} \in B_m$, we have

$$\begin{aligned}
& V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta}) \\
=& b_n \sum_{t=1}^{n} [\psi_\tau(v_t^*(\boldsymbol{\vartheta}) - b_n \boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}))] \boldsymbol{X}_t^*(\boldsymbol{\vartheta}) K(z_{th}) \\
& + b_n \sum_{t=1}^{n} [\psi_\tau(v_t^*(\boldsymbol{\vartheta}))] (\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \boldsymbol{X}_t^*) K(z_{th}) \\
& + b_n \sum_{t=1}^{n} [\psi_\tau(v_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(0))] \boldsymbol{X}_t^* K(z_{th}) + b_n \sum_{t=1}^{n} \Gamma^*(Z_t) \boldsymbol{\vartheta} K(z_{th}) \\
=& b_n \sum_{t=1}^{n} V_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}) + b_n \sum_{t=1}^{n} U_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}) + b_n \sum_{t=1}^{n} W_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}) + b_n \sum_{t=1}^{n} R_{nt}(\boldsymbol{\vartheta}),
\end{aligned}$$

where $V_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}))] \boldsymbol{X}_t^*(\boldsymbol{\vartheta}) K(z_{th}) = \left(V_{nt}^{(1)\top}, V_{nt}^{(2)\top}\right)^\top$, $U_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}))] (\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \boldsymbol{X}_t^*) K(z_{th}) = \left(U_{nt}^{(1)\top}, U_{nt}^{(2)\top}\right)^\top$, $W_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = [\psi_\tau(v_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(0))] \boldsymbol{X}_t^* \times K(z_{th}) = \left(W_{nt}^{(1)\top}, W_{nt}^{(2)\top}\right)^\top$, and $R_{nt}(\boldsymbol{\vartheta}) = \Gamma^*(Z_t) \boldsymbol{\vartheta} K(z_{th}) = \left(R_{nt}^{(1)\top}, R_{nt}^{(2)\top}\right)^\top$ with $V_{nt}^{(1)} = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}))] \boldsymbol{X}_t(\boldsymbol{\vartheta}) K(z_{th})$, $V_{nt}^{(2)} = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}))] \boldsymbol{X}_t(\boldsymbol{\vartheta}) z_{th} K(z_{th})$, $U_{nt}^{(1)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}))] (\boldsymbol{X}_t(\boldsymbol{\vartheta}) - \boldsymbol{X}_t) K(z_{th})$, and $U_{nt}^{(2)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}))] (\boldsymbol{X}_t(\boldsymbol{\vartheta}) - \boldsymbol{X}_t) z_{th} K(z_{th})$. In addition, $W_{nt}^{(1)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(0))]\boldsymbol{X}_t K(z_{th})$, $W_{nt}^{(2)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(0))]\boldsymbol{X}_t z_{th} K(z_{th})$, $R_{nt}^{(1)} = \Gamma(Z_t)\boldsymbol{\vartheta} K(z_{th})$ and $R_{nt}^{(2)} = \Gamma(Z_t)\boldsymbol{\vartheta} z_{th} K(z_{th})$. Thus,

$$
\begin{aligned}
&\|V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta}) - E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta})]\| \\
=&\left\| b_n \begin{pmatrix} \sum_{t=1}^n (V_{nt}^{(1)} - EV_{nt}^{(1)}) \\ \sum_{t=1}^n (V_{nt}^{(2)} - EV_{nt}^{(2)}) \end{pmatrix} \right\| + \left\| b_n \begin{pmatrix} \sum_{t=1}^n (U_{nt}^{(1)} - EU_{nt}^{(1)}) \\ \sum_{t=1}^n (U_{nt}^{(2)} - EU_{nt}^{(2)}) \end{pmatrix} \right\| \\
&+ \left\| b_n \begin{pmatrix} \sum_{t=1}^n (W_{nt}^{(1)} - EW_{nt}^{(1)}) \\ \sum_{t=1}^n (W_{nt}^{(2)} - EW_{nt}^{(2)}) \end{pmatrix} \right\| + \left\| b_n \begin{pmatrix} \sum_{t=1}^n (R_{nt}^{(1)} - ER_{nt}^{(1)}) \\ \sum_{t=1}^n (R_{nt}^{(2)} - ER_{nt}^{(2)}) \end{pmatrix} \right\| \\
\le& b_n \|\sum_{t=1}^n (V_{nt}^{(1)} - EV_{nt}^{(1)})\| + b_n \|\sum_{t=1}^n (V_{nt}^{(2)} - EV_{nt}^{(2)})\| \\
&+ b_n \|\sum_{t=1}^n (U_{nt}^{(1)} - EU_{nt}^{(1)})\| + b_n \|\sum_{t=1}^n (U_{nt}^{(2)} - EU_{nt}^{(2)})\| \\
&+ b_n \|\sum_{t=1}^n (W_{nt}^{(1)} - EW_{nt}^{(1)})\| + b_n \|\sum_{t=1}^n (W_{nt}^{(2)} - EW_{nt}^{(2)})\| \\
&+ b_n \|\sum_{t=1}^n (R_{nt}^{(1)} - ER_{nt}^{(1)})\| + b_n \|\sum_{t=1}^n (R_{nt}^{(2)} - ER_{nt}^{(2)})\| \\
\equiv& V_n^{(1)} + V_n^{(2)} + U_n^{(1)} + U_n^{(2)} + W_n^{(1)} + W_n^{(2)} + R_n^{(1)} + R_n^{(2)}.
\end{aligned}
$$

As for $V_n^{(1)}$, it is easy to see that

$$
V_n^{(1)} \equiv b_n \|\sum_{t=1}^n (V_{nt}^{(1)} - EV_{nt}^{(1)})\| \le \sum_{i=1}^{2\kappa+1} \|b_n \sum_{t=1}^n (V_{nt}^{(1i)} - EV_{nt}^{(1i)})\| = \sum_{i=1}^{2\kappa+1} \|V_n^{(1i)}\|,
$$

where $V_{nt}^{(1i)} = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}))] X_{it}(\boldsymbol{\vartheta}) K(z_{th})$, and $V_n^{(1i)} = b_n \sum_{t=1}^n (V_{nt}^{(1i)} - EV_{nt}^{(1i)})$. Now, we consider the variance of $V_n^{(1i)}$; that is,

$$
\begin{aligned}
E(V_n^{(1i)})^2 &= \frac{1}{nh} E\Big\{ \sum_{t=1}^n (V_{nt}^{(1i)} - EV_{nt}^{(1i)}) \Big\}^2 \\
&= \frac{1}{nh} \Big[\sum_{t=1}^n Var(V_{nt}^{(1i)}) + 2 \sum_{\ell=1}^{n-1} (1 - \frac{\ell}{n}) Cov(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)}) \Big] \\
&\le \frac{1}{h} Var(V_{n1}^{(1i)}) + \frac{2}{h} \sum_{\ell=1}^{d_n-1} |Cov(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)})| + \frac{2}{h} \sum_{\ell=d_n}^{\infty} |Cov(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)})| \\
&\equiv J_4 + J_5 + J_6
\end{aligned}
$$

with $d_n \to \infty$ specified later. First, we consider the last term, J_6, in the above equation. To this end, using Davydov's inequality (see, e.g., Corollary A.2 of Hall and Heyde (1980)), one has

$$|Cov(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)})| \leq C\alpha^{1-2/\delta}(\ell)[E|V_{n1}^{(1i)}|^\delta]^{2/\delta}. \tag{C.4}$$

Notice that for any $k > 0$, $|\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}))|^k = I(r_{3t} < Y_t \leq r_{4t})$, where $r_{3t} = \min(p_{2t}, p_{2t} + p_{3t})$ and $r_{4t} = \max(p_{2t}, p_{2t} + p_{3t})$ with $p_{2t} = [\boldsymbol{g}_\tau(z_0) + \boldsymbol{g}_\tau^{(1)}(z_0)(Z_t - z_0)]^\top \boldsymbol{X}_t(\boldsymbol{\vartheta})$ and $p_{3t} = \frac{1}{\sqrt{nh}}\boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta})$. Therefore, by Assumption A6, there exists a $C > 0$ such that

$$E\{|\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}))|^k | Z_t, \boldsymbol{X}_t\} = F_{Y|Z,\boldsymbol{X}}(r_{4t}) - F_{Y|Z,\boldsymbol{X}}(r_{3t}) \leq C b_n |\boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta})|,$$

which implies by Assumption A9 that

$$\begin{aligned} E[V_{n1}^{(1i)}|^\delta &= E[|\psi_\tau(v_{n1}^*) - \psi_\tau(v_1^*(\boldsymbol{\vartheta}))|^\delta |X_{i1}(\boldsymbol{\vartheta})|^\delta K^\delta(z_{1h})] \\ &\leq C b_n E[|\boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta})||X_{i1}(\boldsymbol{\vartheta})|^\delta K^\delta(z_{1h})]. \end{aligned}$$

Notice that since $\|\boldsymbol{\vartheta}\| \leq \mathfrak{L}_n$, by mean value theorem and triangle inequality, one can choose a sufficiently large $C > 0$, such that $\|\boldsymbol{X}_t^*(\boldsymbol{\vartheta})\| \leq C\|\boldsymbol{X}_t^*\|$. Then,

$$\begin{aligned} E|V_{n1}^{(1i)}|^\delta &= E[|\psi_\tau(v_{n1}^*) - \psi_\tau(v_1^*(\boldsymbol{\vartheta}))|^\delta |X_{i1}(\boldsymbol{\vartheta})|^\delta K^\delta(z_{1h})] \\ &\leq C b_n E[|\boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta})||X_{i1}(\boldsymbol{\vartheta})|^\delta K^\delta(z_{1h})] \leq C b_n E[|\boldsymbol{\theta}^\top \boldsymbol{X}_t^*||X_{1i}|^\delta K^\delta(z_{1h})] \leq C b_n h. \end{aligned}$$

This, in conjunction with (C.4), gives that

$$J_6 \leq C b_n^{2/\delta} h^{2/\delta - 1} \sum_{\ell=d_n}^{\infty} \alpha^{1-2/\delta}(\ell) \leq C b_n^{2/\delta} h^{2/\delta-1} d_n^{-w} \sum_{\ell=d_n}^{\infty} \ell^w \alpha^{1-2/\delta}(\ell) = o(b_n^{2/\delta} h^{2/\delta-1} d_n^{-w}) = o(1).$$

As for J_5, again by choosing sufficiently large $C > 0$, we use Assumptions A6 and A12 to obtain

$$\begin{aligned} |Cov(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)})| &\leq E|V_{n1}^{(1i)} V_{n(\ell+1)}^{(1i)}| + E|V_{n1}^{(1i)}|E|V_{n(\ell+1)}^{(1i)}| \\ &\leq C E|X_{1i} X_{(\ell+1)i}| K(z_{1h}) K(z_{(\ell+1)h}) + Ch^2 \leq Ch^2. \end{aligned}$$

It follows that $J_5 = o(1)$ by $d_n h \to 0$. Analogously,

$$J_4 = h^{-1} Var(V_{n1}^{(1i)}) \leq h^{-1} E(V_{n1}^{(1i)})^2 = O(b_n).$$

Thus, $V_{n1}^{(1i)} = o_p(1)$. So that $V_n^{(1)} = o_p(1)$. Similarly, it can be shown that $V_n^{(2)} = o_p(1)$. For $U_n^{(1)}$, also notice that

$$U_n^{(1)} \equiv b_n \| \sum_{t=1}^{n} (U_{nt}^{(1)} - EU_{nt}^{(1)}) \| \leq \sum_{i=1}^{2\kappa+1} \| b_n \sum_{t=1}^{n} (U_{nt}^{(1i)} - EU_{nt}^{(1i)}) \| = \sum_{i=1}^{2\kappa+1} \| U_n^{(1i)} \|,$$

where $U_{nt}^{(1i)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}))](X_{ti}(\boldsymbol{\vartheta}) - X_{ti})K(z_{th})$ and $U_n^{(1i)} = b_n \sum_{t=1}^n (U_{nt}^{(1i)} - EU_{nt}^{(1i)})$. By mean value theorem, there exists $\boldsymbol{\vartheta}' \in (0, \boldsymbol{\vartheta})$, such that

$$\begin{aligned} E|U_{n1}^{(1i)}|^\delta &= E[|\psi_\tau(v_1^*(\boldsymbol{\vartheta}))|^\delta |X_{1i}(\boldsymbol{\vartheta}) - X_{1i}|^\delta K^\delta(z_{1h})] \\ &\leq CE[|X_{1i}(\boldsymbol{\vartheta}) - X_{1i}|^\delta K^\delta(z_{1h})] \leq CE\left[\left|\left(\frac{\partial X_{1i}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}'} \boldsymbol{\vartheta}\right)\right|^\delta K^\delta(z_{1h})\right] \leq C(R_n/n^{1/2})^\delta h \end{aligned}$$

by the boundedness of $\psi_\tau(\cdot)$. Then, it can be shown that $U_{n1}^{(1i)} = o_p(1)$ so that $U_n^{(1)} = o_p(1)$. Similarly, one can also prove that $U_n^{(2)} = o_p(1)$. As for $W_{nt}^{(1)}$, notice that for any $k > 0$, $|\psi_\tau(v_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(0))|^k = I(c_{3t} < Y_t \leq c_{4t})$, where $c_{3t} = \min(d_{2t}, d_{3t})$ and $c_{4t} = \max(d_{2t}, d_{3t})$ with $d_{2t} = [\boldsymbol{g}_\tau(z_0) + \boldsymbol{g}_\tau^{(1)}(z_0)(Z_t - z_0)]^\top \boldsymbol{X}_t(\boldsymbol{\vartheta})$ and $d_{3t} = [\boldsymbol{g}_\tau(z_0) + \boldsymbol{g}_\tau^{(1)}(z_0)(Z_t - z_0)]^\top \boldsymbol{X}_t$. Therefore, by Assumption A6, there exists a $C > 0$ such that

$$E\{|\psi_\tau(v_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(0))|^k | Z_t, \boldsymbol{X}_t\} = F_{Y|Z,\boldsymbol{X}}(c_{4t}) - F_{Y|Z,\boldsymbol{X}}(c_{3t}) \leq C\left|\left(\frac{\partial X_{1i}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}'}\right)\boldsymbol{\vartheta}\right|,$$

which implies by Assumption A9 that

$$E|W_{n1}^{(1i)}|^\delta = E[|\psi_\tau(v_t^*(\boldsymbol{\vartheta})) - \psi_\tau(v_t^*(0))|^\delta |X_{i1}|^\delta K^\delta(z_{1h})] \leq C(R_n/n^{1/2})^\delta h.$$

Then, it is not hard to show that $W_{nt}^{(1)} = o_p(1)$ and $W_{nt}^{(2)} = o_p(1)$. Similarly, one can also obtain that $R_{nt}^{(1)} = o_p(1)$ and $R_{nt}^{(2)} = o_p(1)$. Thus, it follows that for any fixed $\boldsymbol{\theta} \in A_m$ and for any fixed $\boldsymbol{\vartheta} \in B_m$,

$$\|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta}) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta})]\| = o_p(1). \tag{C.5}$$

Next, to show that the above result holds uniformly in A_m and B_m, we use the Bickel's (1975) chaining approach to show that

$$\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta}) - E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta})]\| = o_p(1).$$

Now, we decompose A_m and B_m into cubes, respectively, based on the grid $(j_1\hbar M,\ldots, j_{2(2\kappa+1)}\hbar M)$ and $(i_1\Bbbk\mathfrak{L},\ldots,i_{2(2\kappa+1)}\Bbbk\mathfrak{L})$, where $j_k = 0,\pm1,\ldots,\pm[1/\hbar]+1$, $i_k = 0,\pm1,\ldots,\pm[1/\Bbbk]+1$, $[\cdot]$ denotes taking integer part of $\cdot$, and $\hbar$ and $\Bbbk$ are fixed positive small numbers. Denote $D(\boldsymbol{\theta})$ and $D(\boldsymbol{\vartheta})$ the lower vertex of cubes that contain $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$, respectively. Then,

$$\begin{aligned}
&\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta}) - E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta})]\|\\
\le&\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|V_n(D(\boldsymbol{\theta}),0) - V_n(0,0) - E[V_n(D(\boldsymbol{\theta}),0) - V_n(0,0)]\|\\
&+\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|V_n(D(\boldsymbol{\theta}),\boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}),0) - E[V_n(D(\boldsymbol{\theta}),\boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}),0)]\|\\
&+\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}),\boldsymbol{\vartheta}) - E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}),\boldsymbol{\vartheta})]\|\\
&+\sup_{\boldsymbol{\vartheta}\in B_m} \|V_n(\boldsymbol{\vartheta}) - E[V_n(\boldsymbol{\vartheta})]\|\\
\equiv& H_1 + H_2 + H_3 + H_4.
\end{aligned}$$

Notice that following the way in Xu (2005), it is not hard to show that $H_4 = o_p(1)$. We only need to focus on H_1, H_2 and H_3. To this end, for H_1, since $\boldsymbol{X}_t \equiv \boldsymbol{X}_t(0)$, it follows easily from (C.5) that

$$H_1 \equiv \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|V_n(D(\boldsymbol{\theta}),0) - V_n(0,0) - E[V_n(D(\boldsymbol{\theta}),0) - V_n(0,0)]\| = o_p(1).$$

As for the first term of H_3, notice that

$$\begin{aligned}
&\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}),\boldsymbol{\vartheta})\|\\
=&b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[\psi_\tau(v_{nt}^*(\boldsymbol{\theta},\boldsymbol{\vartheta})) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}),\boldsymbol{\vartheta}))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\|\\
\le&b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I(v_{nt}^*(D(\boldsymbol{\theta}),\boldsymbol{\vartheta}) < 0) - I(v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta})) < 0)]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\|\\
&+b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I(v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta})) < 0) - I(v_{nt}^*(\boldsymbol{\theta},\boldsymbol{\vartheta}) < 0)]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\|
\end{aligned}$$

$$
\begin{aligned}
&\le 2b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\| \\
&\le 2b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}](\boldsymbol{X}_t^*(\boldsymbol{\vartheta})-\boldsymbol{X}_t^*(D(\boldsymbol{\vartheta})))K(z_{th})\| \\
&\quad + 2b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}]\boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}))K(z_{th})\| \\
&\le 2b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}](\boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}))+\mathfrak{L}_n)K(z_{th})\| \\
&\le 2Cb_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}} - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}] \\
&\quad \times \boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}))K(z_{th})\| \\
&\quad + 2Cb_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}]\boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}))K(z_{th})\| \\
&\le 2Cb_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}} - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}] \\
&\quad \times \boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}))K(z_{th})\| + (2C/h)\max\{\hbar,\Bbbk\}\|E[\boldsymbol{X}_t^*K(z_{th})]\| \\
&\le 2Cb_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}} - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}] \\
&\quad \times \boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}))K(z_{th})\| + 2C\max\{\hbar,\Bbbk\},
\end{aligned} \tag{C.6}
$$

where the fourth inequality follows from the Lipschitz continuity. Since the number of the elements in $\{D(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| \le M\}$ and $\{D(\boldsymbol{\vartheta}) : \|\boldsymbol{\vartheta}\| \le \mathfrak{L}_n\}$ are finite, one can easily show that

$$
\begin{aligned}
&2Cb_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}} - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}] \\
&\times \boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}))K(z_{th})\| = o_p(1)
\end{aligned}
$$

by following the same steps as in (C.5). Let $\max\{\hbar,\Bbbk\} \to 0$. Then, it follows that the first term of H_3 is $o_p(1)$. As for the second term of H_3, in the same way as in (C.6),

$$
\begin{aligned}
&\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}),\boldsymbol{\vartheta})]\| \\
&= b_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta},\boldsymbol{\vartheta})) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}),\boldsymbol{\vartheta}))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\}\| \\
&\le 2nb_n \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|E[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}),D(\boldsymbol{\vartheta}))|<\frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\| \le C\max\{\hbar,\Bbbk\}.
\end{aligned}
$$

When $\max\{\hbar, \Bbbk\} \to 0$, one has

$$\sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta})]\| = o(1).$$

Thus, $H_3 = o_p(1)$. For the first term of H_2, notice that

$$\begin{aligned}
&\sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}), 0)]\| \\
=&b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}))\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))\boldsymbol{X}_t^*]K(z_{th})\| \\
\leq& b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta})) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta})))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\| \\
&+ b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}))) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\| \\
&+ b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))](\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \boldsymbol{X}_t^*)K(z_{th})\| \equiv H_{21} + H_{22} + H_{23}.
\end{aligned}$$

It is easy to see that by following the same deduction as in (C.6), one can derive $H_{21} = o_p(1)$ and $H_{22} = o_p(1)$. Also, notice that for H_{23}, by mean value theorem,

$$\begin{aligned}
H_{23} \equiv& b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))](\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \boldsymbol{X}_t^*))K(z_{th})\| \\
\leq& C(R_n/n^{1/2})b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n}[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))]K(z_{th})\|,
\end{aligned}$$

and the last term can be vanished in probability in the same way as processing $U_n^{(1)}$ and $U_n^{(2)}$. Therefore, the first term of H_2 is $o_p(1)$. For the second term of H_2,

$$\begin{aligned}
&\sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|E\{V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}) - V_n(D(\boldsymbol{\theta}), 0)\}\| \\
=&b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n} E[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}))\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))\boldsymbol{X}_t^*]K(z_{th})\| \\
\leq& b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n} E[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta})) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\| \\
&+ b_n \sup_{\boldsymbol{\vartheta}\in B_m, \boldsymbol{\theta}\in A_m} \|\sum_{t=1}^{n} E[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))](\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \boldsymbol{X}_t^*)K(z_{th})\| \equiv H'_{21} + H'_{22}.
\end{aligned}$$

Now, we consider H'_{22}. Notice that

$$
\begin{aligned}
H'_{22} &\equiv \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|b_n \sum_{t=1}^{n} E\{[\psi_\tau(v^*_{nt}(D(\boldsymbol{\theta}),0))](\boldsymbol{X}^*_t(\boldsymbol{\vartheta}) - \boldsymbol{X}^*_t)K(z_{th})\}\| \\
&= \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|b_n \sum_{t=1}^{n} E\{[\tau - F_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t) + hz_{th}\boldsymbol{g}^{(1)}_\tau(z_0)^\top \boldsymbol{X}_t \\
&\quad + b_n D(\boldsymbol{\theta})^\top \boldsymbol{X}^*_t | Z_t, \boldsymbol{X}_t)](\boldsymbol{X}^*_t(\boldsymbol{\vartheta}) - \boldsymbol{X}^*_t)K(z_{th})\}\| \\
&= \sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|b_n \sum_{t=1}^{n} E\{[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t) + hz_{th}\boldsymbol{g}^{(1)}_\tau(z_0)^\top \boldsymbol{X}_t \\
&\quad + \Im\Pi(h,z_0,Z_t,\boldsymbol{X}_t)|Z_t,\boldsymbol{X}_t)](\boldsymbol{X}^*_t(\boldsymbol{\vartheta}) - \boldsymbol{X}^*_t)\} \\
&\quad \times \Pi(h,z_0,Z_t,\boldsymbol{X}_t)K(z_{th})\|,
\end{aligned}
$$

where $\Pi(h,z_0,Z_t,\boldsymbol{X}_t) = q_\tau(Z_t,\boldsymbol{X}_t) - q_\tau(z_0,\boldsymbol{X}_t) - hz_{th}\boldsymbol{g}^{(1)}_\tau(z_0)^\top \boldsymbol{X}_t - b_n D(\boldsymbol{\theta})^\top \boldsymbol{X}^*_t$. An application of Taylor expansion of $q_\tau(Z_t,\boldsymbol{X}_t)$ at $(z_0,\boldsymbol{X}_t)$ leads to

$$
\Pi(h,z_0,Z_t,\boldsymbol{X}_t) = \frac{\boldsymbol{g}^{(2)}_\tau(z_0+\zeta hz_{th})^\top}{2} h^2 z^2_{th}\boldsymbol{X}_t - b_n D(\boldsymbol{\theta})^\top \boldsymbol{X}^*_t = O_p(h^2).
$$

Therefore, it results in that by mean value theorem, there exists $\boldsymbol{\vartheta}' \in (0,\boldsymbol{\vartheta})$, such that

$$
\begin{aligned}
&\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|b_n \sum_{t=1}^{n} E\{[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t) + hz_{th}\boldsymbol{g}^{(1)}_\tau(z_0)^\top \boldsymbol{X}_t \\
&\quad + \Im\Pi(h,z_0,Z_t,\boldsymbol{X}_t)|Z_t,\boldsymbol{X}_t)](\boldsymbol{X}^*_t(\boldsymbol{\vartheta}) - \boldsymbol{X}^*_t)\}\Pi(h,z_0,Z_t,\boldsymbol{X}_t)K(z_{th})\| \\
\le &\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|b_n \sum_{t=1}^{n} E\{[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t) + hz_{th}\boldsymbol{g}^{(1)}_\tau(z_0)^\top \boldsymbol{X}_t \\
&\quad + \Im\Pi(h,z_0,Z_t,\boldsymbol{X}_t)|Z_t,\boldsymbol{X}_t)] \left(\left.\frac{\partial \boldsymbol{X}^*_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}'} \right) \boldsymbol{\vartheta}\} \\
&\quad \times \Pi(h,z_0,Z_t,\boldsymbol{X}_t)K(z_{th})\| = o(1).
\end{aligned}
$$

In the same way as in analyzing (C.6), it can be easily shown that $H'_{21} = o_p(1)$. So, $H_2 = o_p(1)$. The proof of Lemma C.10 is completed. □

Lemma C.11. *Under the assumptions in Theorem 3, one has*

$$
\sup_{\boldsymbol{\vartheta}\in B_m,\boldsymbol{\theta}\in A_m} \|E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) - V_n(0,0) + V_n(\boldsymbol{\vartheta})] + f_z(z_0)\Omega^*_1(z_0)\boldsymbol{\theta}\| = o(1),
$$

where $\Omega^*_1(z_0) = \mathrm{diag}\{\Omega^*(z_0), \mu_2\Omega^*(z_0)\}$.

Proof. Notice that

$$E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta})-V_n(0,0)+V_n(\boldsymbol{\vartheta})] = E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta})-V_n(\boldsymbol{\theta},0)+V_n(\boldsymbol{\vartheta})]+E[V_n(\boldsymbol{\theta},0)-V_n(0,0)] \equiv R_1+R_2.$$

For R_2, since the deduction is the same as that in Cai and Xu (2008), we only need to focus on R_1. Indeed,

$$\begin{aligned}
R_1 \equiv\ & b_n\sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta},\boldsymbol{\vartheta}))\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \psi_\tau(v_{nt}^*(\boldsymbol{\theta},0))\boldsymbol{X}_t^*]K(z_{th})\} + E[V_n(\boldsymbol{\vartheta})] \\
=\ & b_n\sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta},\boldsymbol{\vartheta})) - \psi_\tau(v_{nt}^*(0,\boldsymbol{\vartheta}))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
& +b_n\sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(0,0)) - \psi_\tau(v_{nt}^*(\boldsymbol{\theta},0))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
& +b_n\sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(0,\boldsymbol{\vartheta})) - \psi_\tau(v_{nt}^*(0,0))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
& +b_n\sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta},0))](\boldsymbol{X}_t^*(\boldsymbol{\vartheta}) - \boldsymbol{X}_t^*)K(z_{th})\} + b_n\sum_{t=1}^{n} E\{\Gamma^*(Z_t)\boldsymbol{\vartheta}K(z_{th})\} \\
\equiv\ & R_{11}+R_{12}+R_{13}+R_{14}+R_{15}.
\end{aligned}$$

Here, R_{14} can be vanished in the same way as that in proving Lemma C.10. We first consider R_{11} as follows

$$\begin{aligned}
R_{11} \equiv & b_n\sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta},\boldsymbol{\vartheta})) - \psi_\tau(v_{nt}^*(0,\boldsymbol{\vartheta}))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
= & b_n\sum_{t=1}^{n} E\{[F_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t(\boldsymbol{\vartheta})) + hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)^\top\boldsymbol{X}_t(\boldsymbol{\vartheta})|Z_t,\boldsymbol{X}_t) \\
& - F_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t(\boldsymbol{\vartheta})) + hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)^\top\boldsymbol{X}_t(\boldsymbol{\vartheta}) \\
& + b_n\boldsymbol{\theta}^\top\boldsymbol{X}_t^*(\boldsymbol{\vartheta})|Z_t,\boldsymbol{X}_t)]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
= & -\frac{1}{h}E\{[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t(\boldsymbol{\vartheta})) + hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)^\top\boldsymbol{X}_t(\boldsymbol{\vartheta}) \\
& + \eth b_n\boldsymbol{\theta}^\top\boldsymbol{X}_t^*(\boldsymbol{\vartheta})|Z_t,\boldsymbol{X}_t)]\boldsymbol{\theta}^\top\boldsymbol{X}_t^*(\boldsymbol{\vartheta})\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
= & -\frac{1}{h}E\{[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t(\boldsymbol{\vartheta}))|Z_t,\boldsymbol{X}_t)]\boldsymbol{\theta}^\top\boldsymbol{X}_t^*\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} + o(1).
\end{aligned}$$

In the same way, one can easily show by Assumption A6 that

$$
\begin{aligned}
R_{11}+R_{12} =& \frac{1}{h}E\{[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t)|Z_t,\boldsymbol{X}_t)-f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t(\boldsymbol{\vartheta}))|Z_t,\boldsymbol{X}_t)] \\
&\times\boldsymbol{\theta}^\top\boldsymbol{X}_t^*\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\}+o(1) \\
\le& C\frac{1}{h}E\{\boldsymbol{g}_\tau(z_0)^\top(\boldsymbol{X}_t-\boldsymbol{X}_t(\boldsymbol{\vartheta}))\boldsymbol{\theta}^\top\boldsymbol{X}_t^*\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\}+o(1)=o(1).
\end{aligned}
$$

As for R_{13} and R_{15}, by applying mean value theorem, there exists $\boldsymbol{\vartheta}'\in(0,\boldsymbol{\vartheta})$ such that

$$
\begin{aligned}
R_{13}\equiv& b_n\sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(0,\boldsymbol{\vartheta}))-\psi_\tau(v_{nt}^*(0,0))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
=& b_n\sum_{t=1}^n E\{[F_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t)+hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)^\top\boldsymbol{X}_t|Z_t,\boldsymbol{X}_t) \\
&-F_{Y|Z,\boldsymbol{X}}(q_\tau(z_0,\boldsymbol{X}_t(\boldsymbol{\vartheta}))+hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)^\top\boldsymbol{X}_t(\boldsymbol{\vartheta})|Z_t,\boldsymbol{X}_t)]\boldsymbol{X}_t^*(\boldsymbol{\vartheta})K(z_{th})\} \\
=&-b_n\sum_{t=1}^n E\{[f_{Y|Z,\boldsymbol{X}}(\tilde{\boldsymbol{X}}_t^\top(\boldsymbol{g}_\tau(z_0)+hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0))|Z_t,\boldsymbol{X}_t)] \\
&\times\boldsymbol{X}_t^*(\boldsymbol{\vartheta})(\boldsymbol{X}_t(\boldsymbol{\vartheta})-\boldsymbol{X}_t)^\top[\boldsymbol{g}_\tau(z_0)+hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)]K(z_{th})\} \\
=&-b_n\sum_{t=1}^n E\{\Gamma^*(Z_t)\boldsymbol{\vartheta}K(z_{th})\}+o(h)
\end{aligned}
$$

by some simple calculations, where $\tilde{\boldsymbol{X}}_t\equiv\boldsymbol{X}_t+C\boldsymbol{\vartheta}$. This implies that $R_{13}+R_{15}=o(1)$. Thus, one has

$$
\|E[V_n(\boldsymbol{\theta},\boldsymbol{\vartheta})-V_n(0,0)+V_n(\boldsymbol{\vartheta})]+f_z(z_0)\Omega_1^*(z_0)\boldsymbol{\theta}\|=o(1). \tag{C.7}
$$

Similar to the proof of Lemma A.3 in Xu (2005), one can prove that (C.7) holds uniformly in A_m and B_m with the details omitted. These complete the proof of Lemma C.11. □

Lemma C.12. *Let* $B_t=[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^*-\psi_\tau(Y_t^*)\Gamma^*(Z_t)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,t}]K(z_{th})$. *Then, under the assumptions in Theorem 3, one has*

$$
E[B_1]=\frac{h^3f_z(z_0)}{2}\begin{pmatrix}\mu_2\Omega^*(z_0)\boldsymbol{g}_\tau^{(2)}(z_0)\\ 0\end{pmatrix}+o(h^3),
$$

and

$$
Var[B_1]=h\tau(1-\tau)f_z(z_0)\begin{pmatrix}\nu_0 & 0\\ 0 & \nu_2\end{pmatrix}\otimes\Big\{\Omega(z_0)-H_1(z_0)+H_2(z_0)\Big\}+o(h),
$$

where $H_1(z_0) = E[\boldsymbol{X}_1\boldsymbol{\Pi}_{a,1}^\top | Z_1 = z_0]\boldsymbol{\Phi}_a^{-1}\Gamma^\top(z_0) + \Gamma(z_0)\boldsymbol{\Phi}_a^{-1}E[\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^\top | Z_1 = z_0]$ *and* $H_2(z_0) = \Gamma(z_0)\boldsymbol{\Phi}_a^{-1}\boldsymbol{D}(z_0)\boldsymbol{\Phi}_a^{-1}\Gamma^\top(z_0)$. *Then,*

$$Var\left\{\frac{1}{\sqrt{nh}}\sum_{t=1}^n B_t\right\} = \tau(1-\tau)f_z(z_0)\begin{pmatrix}\nu_0 & 0\\ 0 & \nu_2\end{pmatrix} \otimes \left\{\Omega(z_0) - H_1(z_0) + H_2(z_0)\right\} + o(1).$$

Proof. This proof is similar to the proof of Lemma A.4 in Cai and Xu (2008). First, we calculate $E[B_1]$ to obtain

$$\begin{aligned} E[B_1] &= E\{[\psi_\tau(v_1^*(0))\boldsymbol{X}_1^* - \psi_\tau(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}]K(z_{1h})\} \\ &= E\{\psi_\tau(v_1^*(0))\boldsymbol{X}_1^* K(z_{1h})\} - E\{\psi_\tau(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}K(z_{1h})\} \equiv Q_1 + Q_2. \end{aligned}$$

Similar to the proof of Lemma 3.5 in Xu (2005), one can easily obtain that

$$Q_1 = \frac{h^3}{2} f_z(z_0)\{\begin{pmatrix}\mu_2\\ 0\end{pmatrix} \otimes \Omega^*(z_0)\}\boldsymbol{g}_\tau^{(2)}(z_0) + o(h^3) \tag{C.8}$$

with the detail omitted. For Q_2, recall that $Y_t^* \equiv Y_t - \boldsymbol{\alpha}_\tau^\top(\boldsymbol{Z}_t)\mathbb{W}_t$. Then,

$$Q_2 \equiv -E\{\psi_\tau(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}K(z_{1h})\} = 0$$

As for $E[B_1 B_1^\top]$, we have

$$\begin{aligned} E[B_1B_1^\top] =& E\Bigg(\Bigg\{\psi_\tau^2(v_1^*(0))\boldsymbol{X}_1^*\boldsymbol{X}_1^{*\top} - [\psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\boldsymbol{X}_1^*\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1) \\ &+ \psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}] \\ &+ \psi_\tau^2(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)\Bigg\}K^2(z_{1h})\Bigg) \\ =& E\{\psi_\tau^2(v_1^*(0))\boldsymbol{X}_1^*\boldsymbol{X}_1^{*\top}K^2(z_{1h})\} - E\{[\psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\boldsymbol{X}_1^*\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1) \\ &+ \psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}]K^2(z_{1h})\} \\ &+ E\{\psi_\tau^2(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)K^2(z_{1h})\} \\ \equiv& P^{(1)} + P^{(2)} + P^{(3)}. \end{aligned}$$

For $P^{(1)}$, one has

$$P^{(1)} \equiv \tau(1-\tau)E\{\boldsymbol{X}_1^*\boldsymbol{X}_1^{*\top}K^2(z_{1h})\}+o(h^2) = h\tau(1-\tau)f_z(z_0)\begin{pmatrix}\nu_0 & 0\\ 0 & \nu_2\end{pmatrix}\otimes\Omega(z_0)(1+o(1))+o(h^2). \tag{C.9}$$

Similarly,

$$\begin{aligned}
P^{(3)} \equiv\ & E[\psi_\tau^2(Y_1^*)\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)K^2(z_{1h})]\\
=&\tau(1-\tau)E\{\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)K^2(z_{1h})\}+o(h^2)\\
=&\tau(1-\tau)E\{\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}E[\boldsymbol{\Pi}_{a,1}\boldsymbol{\Pi}_{a,1}^\top|Z_1]\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)K^2(z_{1h})\}+o(h^2)\\
=&h_2\tau(1-\tau)f_z(z_0)\begin{pmatrix}\nu_0 & 0\\ 0 & \nu_2\end{pmatrix}\otimes\Big\{\Gamma(z_0)\boldsymbol{\Phi}_a^{-1}\boldsymbol{D}(z_0)\boldsymbol{\Phi}_a^{-1}\Gamma^\top(z_0)\Big\}(1+o(1))+o(h^2)\\
=&h\tau(1-\tau)f_z(z_0)\Big\{\begin{pmatrix}\nu_0 & 0\\ 0 & \nu_2\end{pmatrix}\otimes H_2(z_0)\Big\}(1+o(1))+o(h^2).
\end{aligned} \tag{C.10}$$

As for $P^{(2)}$, one has

$$\begin{aligned}
P^{(2)} \equiv & -E\{\psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)[\boldsymbol{X}_1^*\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)\\
&+\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}]K^2(z_{1h})\}\\
=&-E\{[\tau-I_{\{v_1^*(0)<0\}}][\tau-I_{\{Y_1^*<0\}}][\boldsymbol{X}_1^*\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)\\
&+\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}]K^2(z_{1h})\}\\
=&-E\{[\tau^2-\tau(I_{\{Y_1^*<0\}}+I_{\{v_1^*(0)<0\}})+I_{\{Y_1^*<0\}}][\boldsymbol{X}_1^*\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)\\
&+\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}]K^2(z_{1h})\}\\
=&-E\{[(\tau-1)(\tau-I_{\{Y_1^*<0\}})+\tau(\tau-I_{\{v_1^*(0)<0\}})][\boldsymbol{X}_1^*\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)\\
&+\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}]K^2(z_{1h})\}\\
&-\tau(1-\tau)E\{[\boldsymbol{X}_1^*\boldsymbol{\Pi}_{a,1}^\top\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1)+\Gamma^*(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}]K^2(z_{1h})\}\\
\equiv&P^{(21)}+P^{(22)}.
\end{aligned}$$

It can be shown that $P^{(21)} = o(h^2)$, using the same idea in proving Lemma 3.5 in Xu (2005). We now focus on evaluating $P^{(22)}$. A simple algebra gives that

$$
\begin{aligned}
P^{(22)} \equiv & -\tau(1-\tau)E\{[\boldsymbol{X}_1^{*}\boldsymbol{\Pi}_{a,1}^{\top}\boldsymbol{\Phi}_a^{-1}\Gamma^{*\top}(Z_1) + \Gamma^{*}(Z_1)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{*\top}]K^2(z_{1h})\} \\
= & -\tau(1-\tau)E\Bigg\{ \begin{pmatrix} \boldsymbol{X}_1\boldsymbol{\Pi}_{a,1}^{\top}\boldsymbol{\Phi}_a^{-1} \\ z_{1h}\boldsymbol{X}_1\boldsymbol{\Pi}_{a,1}^{\top}\boldsymbol{\Phi}_a^{-1} \end{pmatrix} \begin{pmatrix} \Gamma^{\top}(Z_1) & z_{1h}\Gamma^{\top}(Z_1) \end{pmatrix} K^2(z_{1h}) \Bigg\} \\
& -\tau(1-\tau)E\Bigg\{ \begin{pmatrix} \Gamma(Z_1) \\ z_{1h}\Gamma(Z_1) \end{pmatrix} \begin{pmatrix} \boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{\top} & z_{1h}\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{\top} \end{pmatrix} K^2(z_{1h}) \Bigg\} \\
= & -\tau(1-\tau)E\Bigg\{ \begin{pmatrix} 1 & z_{1h} \\ z_{1h} & z_{1h}^2 \end{pmatrix} \otimes E[\boldsymbol{X}_1\boldsymbol{\Pi}_{a,1}^{\top}|Z_1]\boldsymbol{\Phi}_a^{-1}\Gamma^{\top}(Z_1)K^2(z_{1h}) \Bigg\} \\
& -\tau(1-\tau)E\Bigg\{ \begin{pmatrix} 1 & z_{1h} \\ z_{1h} & z_{1h}^2 \end{pmatrix} \otimes \Gamma(Z_1)\boldsymbol{\Phi}_a^{-1}E[\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{\top}|Z_1]K^2(z_{1h}) \Bigg\} \\
= & -h\tau(1-\tau)f_z(z_0) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \Big\{ E[\boldsymbol{X}_1\boldsymbol{\Pi}_{a,1}^{\top}|Z_1 = z_0]\boldsymbol{\Phi}_a^{-1}\Gamma^{\top}(z_0) \\
& + \Gamma(z_0)\boldsymbol{\Phi}_a^{-1}E[\boldsymbol{\Pi}_{a,1}\boldsymbol{X}_1^{\top}|Z_1 = z_0] \Big\}(1+o(1)) \\
= & -h\tau(1-\tau)f_z(z_0)\Bigg\{ \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes H_1(z_0) \Bigg\}(1+o(1)).
\end{aligned}
$$

Therefore,

$$
P^{(2)} = -h\tau(1-\tau)f_z(z_0)\Bigg\{ \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes H_1(z_0) \Bigg\}(1+o(1)) + o(h^2). \tag{C.11}
$$

Next, it is shown that the last part of lemma holds true. Notice that

$$
\begin{aligned}
& Var\Bigg\{ \frac{1}{\sqrt{nh_2}} \sum_{t=1}^{n} B_t \Bigg\} = \frac{1}{h}[Var(B_1) + 2\sum_{\ell=1}^{n-1}(1-\frac{\ell}{n})Cov(B_1, B_{\ell+1})] \\
\leq & \frac{1}{h}Var(B_1) + \frac{2}{h}\sum_{\ell=1}^{e_n-1}|Cov(B_1, B_{\ell+1})| + \frac{2}{h}\sum_{\ell=e_n}^{\infty}|Cov(B_1, B_{\ell+1})| \equiv G_1 + G_2 + G_3.
\end{aligned}
$$

By (C.8), (C.9), (C.10), (C.11) and Assumption A11,

$$G_1 \to \tau(1-\tau) f_z(z_0) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \Big\{ \Omega(z_0) - H_1(z_0) + H_2(z_0) \Big\}.$$

Now it remains to show that $|G_2| = o(1)$ and $|G_3| = o(1)$. First, we consider G_3. To this end, by using Davydov's inequality (see, e.g., Corollary A.2 of Hall and Heyde (1980)) and the boundedness of $\psi_\tau(\cdot)$, one has

$$|Cov(B_1, B_{\ell+1})| \le C\alpha^{1-2/\delta}(\ell)[E|B_1|^\delta]^{2/\delta} \le Ch^{2/\delta}\alpha^{1-2/\delta}(\ell),$$

which gives

$$G_3 \le Ch^{2/\delta-1} \sum_{\ell=e_n}^{\infty} \alpha^{1-2/\delta}(\ell) \le Ch^{2/\delta-1} e_n^{-w} \sum_{\ell=e_n}^{\infty} \ell^w \alpha^{1-2/\delta}(\ell) = o(h^{2/\delta-1} e_n^{-w}) = o(1),$$

by choosing e_n to satisfy $e_n^w h^{1-2/\delta} = c$. As for G_2, following the proof of Lemma 3.5 in Xu (2005), one has $|G_2| = o(1)$. These prove Lemma C.12. □

C.2 Proof of Theorem 3:

Proof. Following Cai and Xu (2008), $\|V_n(0,0)\| = O_p(1)$. Thus, by Lemmas C.10, C.11 and C.12, $V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ satisfies Condition (ii) in Lemma C.9; that is, $\|A_n\| = O_p(1)$ and $\sup_{\|\boldsymbol{\theta}\| \le M, \|\boldsymbol{\vartheta}\| \le \mathfrak{L}_n} \|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}) + V_n(\boldsymbol{\vartheta}) + D\boldsymbol{\theta} - A_n\| = o_p(1)$ with $D = f_z(z_0)\Omega_1^*(z_0)$ and $A_n = V_n(0,0)$. It remains to show that $\|V_n(\hat{\boldsymbol{\vartheta}})\| = O_p(1)$.

First, write $\|V_n(\hat{\boldsymbol{\vartheta}})\|$ as

$$\begin{aligned} \|V_n(\hat{\boldsymbol{\vartheta}})\| =& b_n \Big\| \sum_{t=1}^n \Gamma^*(Z_t) \hat{\boldsymbol{\vartheta}} K(z_{th}) \Big\| = \left\| b_n \begin{pmatrix} \sum_{t=1}^n R_{nt}^{(1)}(\hat{\boldsymbol{\vartheta}}) \\ \sum_{t=1}^n R_{nt}^{(2)}(\hat{\boldsymbol{\vartheta}}) \end{pmatrix} \right\| \\ \le& b_n \| \sum_{t=1}^n R_{nt}^{(1)}(\hat{\boldsymbol{\vartheta}}) \| + b_n \| \sum_{t=1}^n R_{nt}^{(2)}(\hat{\boldsymbol{\vartheta}}) \| \\ \equiv& R_n^{(1)}(\hat{\boldsymbol{\vartheta}}) + R_n^{(2)}(\hat{\boldsymbol{\vartheta}}), \end{aligned}$$

where $R_{nt}^{(1)}(\hat{\boldsymbol{\vartheta}}) = \Gamma(Z_t)\hat{\boldsymbol{\vartheta}}K(z_{th})$ and $R_{nt}^{(2)}(\hat{\boldsymbol{\vartheta}}) = \Gamma(Z_t)\hat{\boldsymbol{\vartheta}}z_{th}K(z_{th})$. For $R_n^{(1)}(\hat{\boldsymbol{\vartheta}})$, by Lemma C.7, for some constant $C > 0$,

$$R_n^{(1)}(\hat{\boldsymbol{\vartheta}}) \equiv b_n \| \sum_{t=1}^{n} R_{nt}^{(1)}(\hat{\boldsymbol{\vartheta}}) \| = \left\| b_n \sum_{t=1}^{n} [\Gamma(Z_t)\hat{\boldsymbol{\vartheta}}]K(z_{th}) \right\|.$$

Now, we consider the convergence of $\|b_n \sum_{t=1}^{n}[\Gamma(Z_t)\hat{\boldsymbol{\vartheta}}]K(z_{th})\|$. Recall that $\boldsymbol{R}_n = -G_n(\hat{\boldsymbol{c}}_a) + [\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)] + \boldsymbol{R}_n^*$, where $\boldsymbol{R}_n^* = O(\zeta_R\|\hat{\boldsymbol{c}}_a - \bar{\boldsymbol{c}}_a\|^2) + O(\zeta_R r_n^2 K^{-2d})$. By Lemma C.8,

$$\begin{aligned}
&\left\| b_n \sum_{t=1}^{n} [\Gamma(Z_t)\hat{\boldsymbol{\vartheta}}]K(z_{th}) \right\| \\
\le & b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1} \frac{1}{n} \sum_{\mathfrak{s}=m+1}^{n} \psi_\tau(Y_\mathfrak{s}^*)\boldsymbol{\Pi}_{a,\mathfrak{s}}K(z_{th}) \\
& + b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1} \frac{1}{n} \sum_{\mathfrak{s}=m+1}^{n} f_{Y|Z,\boldsymbol{W}}(q_\tau(Z_\mathfrak{s}, \boldsymbol{W}_\mathfrak{s}))\boldsymbol{\Pi}_{a,\mathfrak{s}} b_{R,\mathfrak{s}} K(z_{th}) + b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1}\boldsymbol{R}_n K(z_{th}) \\
\equiv & T^{(1)} + T^{(2)} + T^{(3)}.
\end{aligned}$$

We first focus on $T^{(3)}$. Indeed,

$$\begin{aligned}
T^{(3)} \equiv & b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1}\boldsymbol{R}_n K(z_{th}) \\
= & - b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1} G_n(\hat{\boldsymbol{c}}_a) K(z_{th}) + b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1}[\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)]K(z_{th}) \\
& + b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1}\boldsymbol{R}_n^* K(z_{th}) \\
= & T^{(31)} + T^{(32)} + T^{(33)}.
\end{aligned}$$

For $T^{(32)}$, notice that for some $C > 0$,

$$\begin{aligned}
T^{(32)} \equiv & b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1}[\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)]K(z_{th}) \\
\le & C b_n \max_{1\le t\le n} \|\boldsymbol{\Pi}_{a,t}\| \boldsymbol{\Phi}_a^{-1} \sum_{t=1}^{n} [\tilde{G}_n(\hat{\boldsymbol{c}}_a) - \tilde{G}_n(\bar{\boldsymbol{c}}_a)]K(z_{th}) \\
\le & C h b_n R_n^{1/2} \boldsymbol{\Phi}_a^{-1} \sum_{t=m+1}^{n} [\eta_t(\hat{\boldsymbol{\vartheta}}) - E_t\{\eta_t(\hat{\boldsymbol{\vartheta}})\}],
\end{aligned}$$

where the last inequality follows from $n^{-1}\sum_{t=1}^{n} K(z_{th}) = O(h)$. Similar to the derivation in proving Lemma C.4, to finish the proof, it suffices to show that, for any $a \in \{a \in \mathbb{R}^{R_n} : \|a\| = 1\}$,

$$\sup_{\|\hat{\boldsymbol{\vartheta}}\| \le (R_n/n)^{1/2}} \left| \sum_{t=m+1}^{n} a^\top [\eta_t(\hat{\boldsymbol{\vartheta}}) - E_t\{\eta_t(\hat{\boldsymbol{\vartheta}})\}] \right| = o_p((hb_n R_n^{1/2})^{-1}).$$

Similar to the proof in Xiao and Koenker (2009), covering the ball $\{\|\hat{\boldsymbol{\vartheta}}\| \le C(R_n/n)^{1/2}\}$ with cubes $\mathcal{C} = \{\mathcal{C}_k\}$, where $\mathcal{C}_k$ is a cube with center $\hat{\boldsymbol{\vartheta}}_k$ and side length $C(R_n/n^5)^{1/2}$, so that $N(n) = \#(\mathcal{C}) = (2n^2)^{R_n}$. Therefore, because for $\hat{\boldsymbol{\vartheta}} \in \mathcal{C}_k$, $\|\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}_k\| \le C(R_n/n^{5/2})$ and $I(Y_t^* < x)$ is nondecreasing in x,

$$\begin{aligned}
&\sup_{\|\hat{\boldsymbol{\vartheta}}\| \le C(R_n/n)^{1/2}} \left| \sum_{t=m+1}^{n} a^\top [\eta_t(\hat{\boldsymbol{\vartheta}}) - E_t\{\eta_t(\hat{\boldsymbol{\vartheta}})\}] \right| \\
\le& \max_{1 \le k \le N(n)} \left| \sum_{t=m+1}^{n} a^\top [\eta_t(\hat{\boldsymbol{\vartheta}}_k) - E_t\{\eta_t(\hat{\boldsymbol{\vartheta}}_k)\}] \right| \\
&+ \max_{1 \le k \le N(n)} \left| \sum_{t=m+1}^{n} |(a^\top \boldsymbol{\Pi}_{a,t})| \{b_{nt}(\hat{\boldsymbol{\vartheta}}_k) - E_t(b_{nt}(\hat{\boldsymbol{\vartheta}}_k))\} \right| \\
&+ \max_{1 \le k \le N(n)} \left| \sum_{t=m+1}^{n} |(a^\top \boldsymbol{\Pi}_{a,t})| \{E_t(d_{nt}(\hat{\boldsymbol{\vartheta}}_k))\} \right| \\
\equiv& M_4 + M_5 + M_6,
\end{aligned}$$

where $b_{nt}(\hat{\boldsymbol{\vartheta}}_k) = I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t}) - I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t} + C(R_n/n^{5/2})\|\boldsymbol{\Pi}_{a,t}\|)$ and $d_{nt}(\hat{\boldsymbol{\vartheta}}_k) = I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t} + C(R_n/n^{5/2})\|\boldsymbol{\Pi}_{a,t}\|) - I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t} - C(R_n/n^{5/2})\|\boldsymbol{\Pi}_{a,t}\|)$. The analyses of M_5 and M_6 are similar to those in Welsh (1989) and Xiao and Koenker (2009), so that our focus here is only on M_4. Notice, for any $\flat > 0$, $|I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t}) - I(Y_t^* < -b_{R,t})|^\flat = I(d_{3t} < Y_t \le d_{4t})$, where $d_{3t} = \min(c_{2t}, c_{2t} + c_{3t})$ and $d_{4t} = \max(c_{2t}, c_{2t} + c_{3t})$ with $c_{2t} = -b_{R,t}$ and $c_{3t} = \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k$. Therefore, by Assumption A6, there exists a $C > 0$ such that $E\{|I(Y_t^* < \boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k - b_{R,t}) - I(Y_t^* < -b_{R,t})|^\flat | Z_t, \boldsymbol{W}_t\} = F_{Y|Z,\boldsymbol{W}}(d_{4t}) - F_{Y|Z,\boldsymbol{W}}(d_{3t}) \le C|\boldsymbol{\Pi}_{a,t}^\top \hat{\boldsymbol{\vartheta}}_k| \le C(R_n/n)^{1/2}\|\boldsymbol{\Pi}_{a,t}\|$, which implies that

$$E_t[a^\top \eta_t(\hat{\boldsymbol{\vartheta}}_k)]^2 \le C((R_n/n)^{1/2} R_n^{1/2}).$$

where the inequality holds due to the boundedness of eigenvalues of $n^{-1}\sum_{t=m+1}^{n}\mathbf{\Pi}_{a,t}\mathbf{\Pi}_{a,t}^{\top}$. Thus, we have

$$\mathcal{W}_n^2=\sum_{t=m+1}^{n}E_t[a^\top\{\eta_t(\hat{\boldsymbol{\vartheta}}_k)-E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}]^2\le\sum_{t=m+1}^{n}E_t[a^\top\eta_t(\hat{\boldsymbol{\vartheta}}_k)]^2=O((n/R_n)^{1/2}R_n^{3/2})$$

and

$$\mathcal{S}_n^2=\sum_{t=m+1}^{n}[a^\top\{\eta_t(\hat{\boldsymbol{\vartheta}}_k)-E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}]^2=O_p((n/R_n)^{1/2}R_n^{3/2}).$$

Also, notice that $\xi_t(\hat{\boldsymbol{\vartheta}}_k)=\{\eta_t(\hat{\boldsymbol{\vartheta}}_k)-E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}$ is a martingale difference sequence. Therefore, let $\mathcal{M}=(n/R_n)^{1/2}$. Thus, we have

$$\begin{aligned}
&P\left[\max_{1\le k\le N(n)}\left|hb_nR_n^{1/2}\sum_{t=m+1}^{n}\{a^\top\{\eta_t(\hat{\boldsymbol{\vartheta}}_k)-E(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}\}\right|>\epsilon\right]\\
\le\;& N(n)\max_k P\left[\left|hb_nR_n^{1/2}\sum_{t=m+1}^{n}\{a^\top\{\eta_t(\hat{\boldsymbol{\vartheta}}_k)-E_t(\eta_t(\hat{\boldsymbol{\vartheta}}_k))\}\}\right|>\epsilon\right]\\
\le\;& N(n)\max_k P\left[\left|\sum_{t=m+1}^{n}a^\top\xi_t(\hat{\boldsymbol{\vartheta}}_k)\right|>(hb_nR_n^{1/2})^{-1}\epsilon,\mathcal{W}_n^2+\mathcal{S}_n^2\le\mathcal{M}\right]\\
&+N(n)\max_k P\left[\left|\sum_{t=m+1}^{n}a^\top\xi_t(\hat{\boldsymbol{\vartheta}}_k)\right|>(hb_nR_n^{1/2})^{-1}\epsilon,\mathcal{W}_n^2+\mathcal{S}_n^2>\mathcal{M}\right]\equiv I_1+I_2. \quad (C.12)
\end{aligned}$$

For I_1, by exponential inequality for martingale difference sequences (see, e.g., Bercu and Touati, 2008), we have

$$\begin{aligned}
&N(n)\max_k P\left[\left|\sum_{t=m+1}^{n}a^\top\xi_t(\hat{\boldsymbol{\vartheta}}_k)\right|>(hb_nR_n^{1/2})^{-1}\epsilon,\mathcal{W}_n^2+\mathcal{S}_n^2\le\mathcal{M}\right]\\
\le&2N(n)\exp\left(-\frac{(hb_nR_n^{1/2})^{-2}\epsilon^2}{2\mathcal{M}}\right)=2N(n)\exp\left(-\frac{n^{1/2}\epsilon^2}{2R_n^{1/2}h}\right).
\end{aligned}$$

For I_2, because $P[\mathcal{W}_n^2+\mathcal{S}_n^2>\mathcal{M}]\le P[\mathcal{W}_n^2>\mathcal{M}/2]+P[\mathcal{S}_n^2>\mathcal{M}/2]$ and each term can be bounded exponentially under Assumptions A1, A7 and A8. Thus, $M_4 = o_p((hb_n R_n^{1/2})^{-1})$. This implies that $T^{(32)} = o_p(1)$. By Lemma C.3 and using Davydov's inequality, one has $T^{(31)} = o_p(1)$. Similarly, $T^{(33)} = o_p(1)$. These give us $T^{(3)} = o_p(1)$. Also, by Lemma C.7 and applying Davydov's inequality, it is not hard to show that $T^{(2)} = o_p(1)$. Now, we restrict our attention on $T^{(1)}$. Notice that by boundedness of $\psi_\tau(\cdot)$,

$$\begin{aligned}
T^{(1)} \equiv & b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1} \frac{1}{n} \sum_{\mathfrak{s}=m+1}^{n} \psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{\Pi}_{a,\mathfrak{s}} K(z_{th}) \\
= & b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1} \frac{1}{n} \sum_{\mathfrak{s}=m+1}^{n} \{\psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t} + \psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{\Pi}_{a,\mathfrak{s}} - \psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t}\} K(z_{th}) \\
= & b_n \sum_{t=1}^{n} \Gamma(Z_t)\boldsymbol{\Phi}_a^{-1} \psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t} K(z_{th}) + o_p(1).
\end{aligned}$$

Then, by the definition of Y_t^* and similar to the proof of Lemma C.12, we have $E[T^{(1)}] = 0$ and $Var[T^{(1)}] = O(1)$. These imply that $\|V_n(\hat{\boldsymbol{\vartheta}})\| = O_p(1)$.

To show $\|V_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}})\| = o_p(1)$, it follows from Lemma C.1 and mean value theorem that

$$\begin{aligned}
\|V_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}})\| = & b_n \left\| \sum_{t=1}^{n} [\psi_\tau(v_t^*(\hat{\boldsymbol{\vartheta}}) - b_n \hat{\boldsymbol{\theta}}^\top \boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}))] \boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}) K(z_{th}) \right\| \\
& \leq b_n \dim(\boldsymbol{X}^*(\hat{\boldsymbol{\vartheta}})) \max_{1 \leq t \leq n} \|\boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}) K(z_{th})\| \\
& \leq b_n \dim(\boldsymbol{X}^*) \max_{1 \leq t \leq n} \|\boldsymbol{X}_t^* K(z_{th})\| + C b_n \dim(\boldsymbol{X}^*) \max_{1 \leq t \leq n} \left\| \left(\frac{\partial \boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}})}{\partial \hat{\boldsymbol{\vartheta}}} \bigg|_{\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}'} \right) K(z_{th}) \right\| \\
& = o_p(1),
\end{aligned}$$

where $\hat{\boldsymbol{\theta}}$ is the minimizer of $J(\boldsymbol{\theta})$. Finally, because $\psi_\tau(x)$ is an increasing function of x; then $-\boldsymbol{\theta}^\top V_n(\lambda\boldsymbol{\theta}, \boldsymbol{\vartheta}) = a_n \sum_{t=1}^{n} \psi_\tau[v_t^*(\boldsymbol{\vartheta}) + \lambda a_n(-\boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta}))](-\boldsymbol{\theta}^\top \boldsymbol{X}_t^*(\boldsymbol{\vartheta})) K(z_{th})$ is an increasing function of λ. Thus, Condition (i) in Lemma C.9 is satisfied. Then, it follows from Lemma C.8, Lemmas C.10 and C.11 that

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} =& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh}f_z(z_0)}\sum_{t=1}^{n}[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \Gamma^*(Z_t)\hat{\boldsymbol{\vartheta}}]K(z_{th}) + o_p(1) \\
=& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh}f_z(z_0)}\sum_{t=1}^{n}\left[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \Gamma^*(Z_t)\boldsymbol{\Phi}_a^{-1}\frac{1}{n}\sum_{\mathfrak{s}=m+1}^{n}\psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{\Pi}_{a,\mathfrak{s}}\right]K(z_{th}) + o_p(1) \\
=& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh}f_z(z_0)}\sum_{t=1}^{n}\Bigg[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \Gamma^*(Z_t)\boldsymbol{\Phi}_a^{-1} \\
&\times \frac{1}{n}\sum_{\mathfrak{s}=m+1}^{n}\{\psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t} + \psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{\Pi}_{a,\mathfrak{s}} - \psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t}\}\Bigg]K(z_{th}) + o_p(1) \\
=& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh}f_z(z_0)}\sum_{t=1}^{n}\left[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \psi_\tau(Y_t^*)\Gamma^*(Z_t)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,t}\right]K(z_{th}) \\
&- \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh}f_z(z_0)}\sum_{t=1}^{n}\left[\Gamma^*(Z_t)\boldsymbol{\Phi}_a^{-1}\frac{1}{n}\sum_{\mathfrak{s}=m+1}^{n}\{\psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{\Pi}_{a,\mathfrak{s}} - \psi_\tau(Y_t^*)\boldsymbol{\Pi}_{a,t}\}\right]K(z_{th}) + o_p(1).
\end{aligned}
$$

Here, by using Davydov's inequality to control the variance, the second part of last equality can be asymptotically vanished. Then,

$$
\hat{\boldsymbol{\theta}} = \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh}f_z(z_0)}\sum_{t=1}^{n}\left[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \psi_\tau(Y_t^*)\Gamma^*(Z_t)\boldsymbol{\Phi}_a^{-1}\boldsymbol{\Pi}_{a,t}\right]K(z_{th}) + o_p(1),
$$

Therefore, following the proof of Theorem 1 in Cai and Xu (2008), the theorem is proved. □

C.3 Proof of Consistency of $\hat{\Sigma}_\tau(z_0)$

Proof. We first focus on $\hat{\Gamma}(z_0)$ in Section 2.4 of the main text. Notice that

$$
\begin{aligned}
\hat{\Gamma}(z_0) =& \frac{1}{n}\sum_{t=1}^{n} w_{2t}\hat{\boldsymbol{X}}_t\hat{\boldsymbol{g}}_\tau^\top(z_0)\boldsymbol{\Pi}_{a,t}K_h(Z_t - z_0) \\
=& \frac{1}{n}\sum_{t=1}^{n} w_{2t}(\hat{\boldsymbol{X}}_t - \boldsymbol{X}_t)(\hat{\boldsymbol{g}}_\tau(z_0) - \boldsymbol{g}_\tau(z_0))^\top\boldsymbol{\Pi}_{a,t}K_h(Z_t - z_0) \\
&+ \frac{1}{n}\sum_{t=1}^{n} w_{2t}\boldsymbol{X}_t(\hat{\boldsymbol{g}}_\tau(z_0) - \boldsymbol{g}_\tau(z_0))^\top\boldsymbol{\Pi}_{a,t}K_h(Z_t - z_0) \\
&+ \frac{1}{n}\sum_{t=1}^{n} w_{2t}(\hat{\boldsymbol{X}}_t - \boldsymbol{X}_t)\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{\Pi}_{a,t}K_h(Z_t - z_0) + \frac{1}{n}\sum_{t=1}^{n} w_{2t}\boldsymbol{X}_t\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{\Pi}_{a,t}K_h(Z_t - z_0) \\
\equiv& S^{(1)} + S^{(2)} + S^{(3)} + S^{(4)}.
\end{aligned}
$$

We first consider $S^{(3)}$. By Taylor's expansion and Theorem 2(c), we have

$$
\begin{aligned}
E[w_{2t}|Z_t, \boldsymbol{X}_t] &= (F_{Y|Z,\boldsymbol{X}}(\hat{\boldsymbol{g}}_\tau^\top(z_0)\hat{\boldsymbol{X}}_t + \delta_{2n}) - F_{Y|Z,\boldsymbol{X}}(\hat{\boldsymbol{g}}_\tau^\top(z_0)\hat{\boldsymbol{X}}_t - \delta_{2n}))/(2\delta_{2n}) \\
&= f_{Y|Z,\boldsymbol{X}}(\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{X}_t) + o_p(1).
\end{aligned}
$$

On the other hand, by applying mean value theorem, there exists $\hat{\boldsymbol{\vartheta}}' \in (0, \hat{\boldsymbol{\vartheta}})$ such that

$$
\hat{\boldsymbol{X}}_t \equiv \boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}}) = \boldsymbol{X}_t + \left(\frac{\partial \boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}})}{\partial \hat{\boldsymbol{\vartheta}}}\bigg|_{\hat{\boldsymbol{\vartheta}}=\hat{\boldsymbol{\vartheta}}'}\right)\hat{\boldsymbol{\vartheta}} = \boldsymbol{X}_t + \boldsymbol{\Upsilon}_{a,t}\hat{\boldsymbol{\vartheta}}.
$$

Therefore, by Theorem 2(c) and Assumption A2,

$$
\begin{aligned}
E[S^{(3)}] &= E\left[f_{Y|Z,\boldsymbol{X}}(\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{X}_t)\boldsymbol{\Upsilon}_{a,t}\hat{\boldsymbol{\vartheta}}\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{\Pi}_{a,t}K_h(Z_t - z_0)\right] + o(1) \\
&= O(R_n^{3/2}/n^{1/2}) = o(1).
\end{aligned}
$$

Similar to the proof of $Var[B_1]$ in Lemma C.12 and by Theorem 2(c), it can be shown that $Var[S^{(3)}] = o(1)$. Therefore, $S^{(3)} = o_p(1)$. Similarly, we can show that $S^{(1)} = o_p(1)$ and

$S^{(2)} = o_p(1)$. Now, we only need to focus on $S^{(4)}$. Indeed,

$$
\begin{aligned}
E[S^{(4)}] &= E[f_{Y|Z,\boldsymbol{X}}(\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{X}_t)\boldsymbol{X}_t\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{\Pi}_{a,t}K_h(Z_t - z_0)] + o(1) \\
&= \int f_{Y|Z,\boldsymbol{X}}(\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{X}_t)\boldsymbol{X}_t\boldsymbol{g}_\tau^\top(z_0)\boldsymbol{\Pi}_{a,t}K(z)f_z(z_0 + hz)dz + o(1) \to f_z(z_0)\Gamma(z_0).
\end{aligned}
$$

Again, similar to the proof of $Var[B_1]$ in Lemma C.12, it is shown that $Var[S^{(4)}] = o(1)$. This yields that $\hat{\Gamma}(z_0) = f_z(z_0)\Gamma(z_0) + o_p(1)$ in Section 2.4. The consistency of $\hat{\boldsymbol{\Phi}}_a$, $\hat{\Omega}(z_0)$, $\hat{\Omega}^*(z_0)$, $\hat{H}_1(z_0)$ and $\hat{H}_2(z_0)$ can be derived in similar ways. □

Appendix D: Mathematical Proof for Stationarity and α-Mixing

In this section, we show that the model (1) in the main article can generate a strictly stationary and α-mixing process. Throughout this section, $0_{a\times b}$ stands for a $(a \times b)$ matrix of zeros and I_a is a $(a \times a)$ identity matrix. Next, we define $\psi(\cdot) = \|\cdot\|$, where $\|\cdot\|$ is the Euclidean norm. For a random vector Z and random matrix A, we denote $\|Z\|_{\psi,2} = [E\|Z\|^2]^{1/2}$ and $\|A\|_{\psi,2} = \sup_{z\neq 0} \|Az\|_{\psi,2}/\|z\|$. In addition, for $1 \le i \le \kappa$, let $\mathcal{F}^b_{i,a}$ be the σ-algebra generated by $\{(Y_{it}, Z_{it})\}^b_{t=a}$. Then, a stationary process $\{(Y_{it}, Z_{it})\}^{\infty}_{t=-\infty}$ is said to be α-mixing (strongly mixing) if the mixing coefficient $\alpha(t)$ defined by

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}^0_{i,-\infty}, B \in \mathcal{F}^{\infty}_{i,t}\}$$

converges to zero as $t \to \infty$.

To study the probabilistic properties of model (1) in the main article, $\mathbb{Y}_t$ and $\boldsymbol{q}_{\tau,t}$ in (1) need to be jointly introduced in a vector autoregression process. To proceed, for convenience of presentation, let $\kappa = \kappa_1$ and $Z_t = Z_{it}$ in (1) in the main article, denote U_{it} $(1 \le i \le \kappa,\ 1 \le t \le n)$ as independent and identically distributed (i.i.d.) standard uniform random variables on the set of $[0, 1]$. Then, we consider following equation system of functional-coefficient VAR models for dynamic quantiles, given by

$$Y_{it} = \gamma_{i0}(U_{it}, Z_t) + \sum_{s=1}^{q} \boldsymbol{\gamma}^{\top}_{i,s}(U_{it}, Z_t)\boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}^{\top}_{i,l}(U_{it}, Z_t)\mathbb{Y}_{t-l}, \tag{D.1}$$

and

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_t) + \sum_{s=1}^{q} \boldsymbol{\gamma}^{\top}_{i,s,\tau}(Z_t)\boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}^{\top}_{i,l,\tau}(Z_t)\mathbb{Y}_{t-l} \tag{D.2}$$

for some p and q, where Y_{it}, $\boldsymbol{q}_{\tau,t}$ and $\mathbb{Y}_t$ in (D.1) and (D.2) have the same definition as that in (1) and equation (D.2) is the same as (1) with $Z_t = Z_{it}$. In addition, $\gamma_{i0}(\cdot, \cdot)$ in (D.1) is a scalar and measurable function of U_{it} and Z_t (from $\mathbb{R}^2$ to $\mathbb{R}$), both $\boldsymbol{\gamma}_{i,s}(\cdot, \cdot) = (\gamma_{si1}(\cdot, \cdot), \ldots, \gamma_{si\kappa}(\cdot, \cdot))^{\top}$ and $\boldsymbol{\beta}_{i,l}(\cdot, \cdot) = (\beta_{li1}(\cdot, \cdot), \ldots, \beta_{li\kappa}(\cdot, \cdot))^{\top}$ in (D.1) are $\kappa \times 1$ vectors of measurable functions from $\mathbb{R}^2$ to $\mathbb{R}$. Following the same argument in Koenker and Xiao (2006), by assuming that the right side of (D.1) is monotonically increasing in U_{it}, the conditional quantile function of Y_{it} given $(Z_t, \{\boldsymbol{q}_{\tau,t-s}\}_{s=1}^{q}, \{\mathbb{Y}_{t-l}\}_{l=1}^{p})$ becomes (D.2). Note that (D.1) is called a Skorohod representation for Y_{it}, see Durrett (1996) for the definition of Skorohod representation.

Now, we can rewrite the system formed by (D.1) and (D.2) into an autoregression process of order 1 as follows

$$\mathbb{X}_t = \boldsymbol{\mu}(Z_t) + \boldsymbol{A}_{U_t}(Z_t)\mathbb{X}_{t-1} + \boldsymbol{D}_{U_t}(Z_t), \tag{D.3}$$

where $\mathbb{X}_t = (\mathbb{Y}_t^\top, \ldots, \mathbb{Y}_{t-p+1}^\top, \boldsymbol{q}_{\tau,t}^\top, \ldots, \boldsymbol{q}_{\tau,t-q+1}^\top)^\top$ and $\boldsymbol{A}_{U_t}(Z_t)$ is a $\kappa(p+q) \times \kappa(p+q)$ matrix as follows:

$$\boldsymbol{A}_{U_t}(Z_t) = \begin{pmatrix} \boldsymbol{\Gamma}_{\beta,U_t}(Z_t) & \boldsymbol{\Gamma}_{U_t}(Z_t) \\ [I_{\kappa(p-1)}, 0_{\kappa(p-1)\times\kappa}] & 0_{\kappa(p-1)\times\kappa q} \\ \boldsymbol{\Gamma}_{\beta,\tau}(Z_t) & \boldsymbol{\Gamma}_{\tau}(Z_t) \\ 0_{\kappa(q-1)\times\kappa p} & [I_{\kappa(q-1)}, 0_{\kappa(q-1)\times\kappa}] \end{pmatrix}.$$

Here, for $s = 1, \ldots, q$ and $l = 1, \ldots, p$, $\boldsymbol{\Gamma}_{\beta,U_t}(Z_t) = (\boldsymbol{\Gamma}_{\beta,1,U_t}(Z_t), \ldots, \boldsymbol{\Gamma}_{\beta,p,U_t}(Z_t))$, where $\boldsymbol{\Gamma}_{\beta,l,U_t}(Z_t)$ $= (\beta_{lij}(U_{it}, Z_t))_{1\le i\le\kappa,1\le j\le\kappa}$ is a $\kappa \times \kappa$ matrix. In addition, $\boldsymbol{\Gamma}_{U_t}(Z_t) = (\boldsymbol{\Gamma}_{1,U_t}(Z_t), \ldots, \boldsymbol{\Gamma}_{q,U_t}(Z_t))$, where $\boldsymbol{\Gamma}_{s,U_t}(Z_t) = (\gamma_{sij}(U_{it}, Z_t))_{1\le i\le\kappa,1\le j\le\kappa}$ is a $\kappa\times\kappa$ matrix. Similarly, $\boldsymbol{\Gamma}_{\beta,\tau}(Z_t) = (\boldsymbol{\Gamma}_{\beta,1,\tau}(Z_t), \ldots,$ $\boldsymbol{\Gamma}_{\beta,p,\tau}(Z_t))$, where $\boldsymbol{\Gamma}_{\beta,l,\tau}(Z_t) = (\beta_{lij,\tau}(Z_t))_{1\le i\le\kappa,1\le j\le\kappa}$ is a $\kappa\times\kappa$ matrix. Also, $\boldsymbol{\Gamma}_{\tau}(Z_t) = (\boldsymbol{\Gamma}_{1,\tau}(Z_t), \ldots,$ $\boldsymbol{\Gamma}_{q,\tau}(Z_t))$, where $\boldsymbol{\Gamma}_{s,\tau}(Z_t) = (\gamma_{sij,\tau}(Z_t))_{1\le i\le\kappa,1\le j\le\kappa}$ is a $\kappa \times \kappa$ matrix. Furthermore, $\boldsymbol{\mu}(Z_t) =$ $(E_U^\top(\boldsymbol{\gamma}_0(U_t, Z_t)), 0, \ldots, 0, \boldsymbol{\gamma}_{0,\tau}^\top(Z_t), 0, \ldots, 0)^\top$, where $E_U(\boldsymbol{\gamma}_0(U_t, Z_t)) = (E_U(\gamma_{10}(U_{1t}, Z_t)), \ldots,$ $E_U(\gamma_{\kappa 0}(U_{\kappa t}, Z_t)))^\top$ and $\boldsymbol{\gamma}_{0,\tau}(Z_t) = (\gamma_{10,\tau}(Z_t), \ldots, \gamma_{\kappa 0,\tau}(Z_t))^\top$. Here, $E_U(\cdot)$ is denoted as taking expectation on U_{it} for any fixed Z_t, and $\gamma_{i0}(U_{it}, Z_t)$ and $\gamma_{i0,\tau}(Z_t)$ are defined in a similar way as foregoing functional coefficients, respectively. Finally, $\boldsymbol{D}_{U_t}(Z_t) = (\check{\gamma}_{10}(U_{1t}, Z_t), \ldots, \check{\gamma}_{\kappa 0}(U_{\kappa t}, Z_t),$ $0_{1\times\kappa(p+q-1)})^\top$, where $\check{\gamma}_{i0}(U_{it}, Z_t) = \gamma_{i0}(U_{it}, Z_t) - E_U(\gamma_{i0}(U_{it}, Z_t))$.

Remark D.1. *Notice that when setting Z_t as a smoothing variable, the equations corresponding to ($\kappa p+1$)-th, …, ($\kappa p+\kappa$)-th rows of* (D.3) *are exactly* (D.2) *and the model (1) in the main article, while the ith row of* (D.3) *with $i = 1, \ldots, \kappa$ is equation* (D.1)*. Given these relations, one can conclude that $\mathbb{Y}_t$ and $\boldsymbol{q}_{\tau,t}$ jointly follow a VAR process of order 1 in* (D.3)*, which is similar* *to the nonparametric additive models in Cai and Masry (2000) and the generalized polynomial random coefficient autoregressive (RCA) models in Carrasco and Chen (2002).*

Now, denote $\lambda_{\max}(\boldsymbol{A}_{U_t})$ as the largest eigenvalue in absolute value of following matrix $\boldsymbol{A}_{U_t}$:

$$\boldsymbol{A}_{U_t} = \begin{pmatrix} \boldsymbol{\Gamma}_{\beta,1,U_t} & \boldsymbol{\Gamma}_{\beta,2,U_t} & \dots & \boldsymbol{\Gamma}_{\beta,p-1,U_t} & \boldsymbol{\Gamma}_{\beta,p,U_t} & \boldsymbol{\Gamma}_{1,U_t} & \boldsymbol{\Gamma}_{2,U_t} & \dots & \boldsymbol{\Gamma}_{q-1,U_t} & \boldsymbol{\Gamma}_{q,U_t} \\ I_\kappa & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\ 0_{\kappa\times\kappa} & I_\kappa & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & I_\kappa & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\ \boldsymbol{\Gamma}_{\beta,1} & \boldsymbol{\Gamma}_{\beta,2} & \dots & \boldsymbol{\Gamma}_{\beta,p-1} & \boldsymbol{\Gamma}_{\beta,p} & \boldsymbol{\Gamma}_{1} & \boldsymbol{\Gamma}_{2} & \dots & \boldsymbol{\Gamma}_{q-1} & \boldsymbol{\Gamma}_{q} \\ 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & I_\kappa & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\ 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & I_\kappa & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & I_\kappa & 0_{\kappa\times\kappa} \end{pmatrix},$$

where

$$\boldsymbol{\Gamma}_{\beta,l,U_t} = \begin{pmatrix} \beta_{l11}(U_{1t}) & \beta_{l12}(U_{1t}) & \dots & \beta_{l1\kappa}(U_{1t}) \\ \beta_{l21}(U_{2t}) & \beta_{l22}(U_{2t}) & \dots & \beta_{l2\kappa}(U_{2t}) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{l\kappa1}(U_{\kappa t}) & \beta_{l\kappa2}(U_{\kappa t}) & \dots & \beta_{l\kappa\kappa}(U_{\kappa t}) \end{pmatrix}, \quad \boldsymbol{\Gamma}_{s,U_t} = \begin{pmatrix} \gamma_{s11}(U_{1t}) & \gamma_{s12}(U_{1t}) & \dots & \gamma_{s1\kappa}(U_{1t}) \\ \gamma_{s21}(U_{2t}) & \gamma_{s22}(U_{2t}) & \dots & \gamma_{s2\kappa}(U_{2t}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{s\kappa1}(U_{\kappa t}) & \gamma_{s\kappa2}(U_{\kappa t}) & \dots & \gamma_{s\kappa\kappa}(U_{\kappa t}) \end{pmatrix},$$

$$\boldsymbol{\Gamma}_{\beta,l} = \begin{pmatrix} \beta_{l11,\tau} & \beta_{l12,\tau} & \dots & \beta_{l1\kappa,\tau} \\ \beta_{l21,\tau} & \beta_{l22,\tau} & \dots & \beta_{l2\kappa,\tau} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{l\kappa1,\tau} & \beta_{l\kappa2,\tau} & \dots & \beta_{l\kappa\kappa,\tau} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Gamma}_{s} = \begin{pmatrix} \gamma_{s11,\tau} & \gamma_{s12,\tau} & \dots & \gamma_{s1\kappa,\tau} \\ \gamma_{s21,\tau} & \gamma_{s22,\tau} & \dots & \gamma_{s2\kappa,\tau} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{s\kappa1,\tau} & \gamma_{s\kappa2,\tau} & \dots & \gamma_{s\kappa\kappa,\tau} \end{pmatrix},$$

with each entry being defined in the Assumption D later. Then, following assumptions are needed to guarantee that process $\{\mathbb{X}_t\}$ in model (D.3) is strictly stationary and α-mixing.

Assumption D.

D1: *Let* $\{\mathbb{X}_t\}$ *be a* ϕ*-irreducible and aperiodic Markov chain. For* $i = 1, \ldots, \kappa$, $j = 1, \ldots, \kappa$, $l = 1, \ldots, p$ *and* $s = 1, \ldots, q$, *each entry of* $\mathbf{\Gamma}_{s,U_t}(Z_t)$ *and* $\mathbf{\Gamma}_{\beta,l,U_t}(Z_t)$ *in* (D.1) *is bounded such that* $|\gamma_{sij}(U_{it}, \cdot)| \leq \gamma_{sij}(U_{it})$ *and* $|\beta_{lij}(U_{it}, \cdot)| \leq \beta_{lij}(U_{it})$, $\beta_{lij}(U_{it})$ *and* $\gamma_{sij}(U_{it})$ *are unknown measurable functions of* U_{it} *from* $[0, 1]$ *to* $\mathbb{R}$*; Similarly, each entry of* $\mathbf{\Gamma}_{s,\tau}(Z_t)$ *and* $\mathbf{\Gamma}_{\beta,l,\tau}(Z_t)$ *in* (D.2) *is bounded such that* $|\gamma_{sij,\tau}(\cdot)| \leq \gamma_{sij,\tau}$ *and* $|\beta_{lij,\tau}(\cdot)| \leq \beta_{lij,\tau}$. *Furthermore,* $E\{[\lambda_{\max}(\boldsymbol{A}_{U_t})]^2\} < 1$.

D2: *For* $i = 1, \ldots, \kappa$, $\check{\gamma}_{i0}(U_{it}, Z_t)$ *in* $\boldsymbol{D}_{U_t}(Z_t)$ *is bounded such that* $|\check{\gamma}_{i0}(U_{it}, \cdot)| \leq \check{\gamma}_{i0}(U_{it})$, *where* $\{\check{\gamma}_{i0}(U_{it})\}$ *are i.i.d. random variables with mean 0 and finite variance. In addition, denote* $\boldsymbol{D}_{U_t} = (\check{\gamma}_{10}(U_{1t}), \ldots, \check{\gamma}_{\kappa 0}(U_{\kappa t}), 0_{1\times\kappa(p+q-1)})^\top$, *then,* $E\|\boldsymbol{D}_{U_t}\|^2 < \infty$ *and* $E\|\boldsymbol{\mu}(Z_t)\| < \infty$.

Remark D.2. *The* ϕ*-irreducibility and aperiodicity in Assumption D1 are key assumptions for deriving geometric ergodicity and subsequently,* α*-mixing property. The conditions that imply* ϕ*-irreducibility and aperiodicity of nonlinear time series have been studied extensively in literature. For example, Chan and Tong (1985) showed that under some mild conditions, a simple nonparametric autoregressive process is a* ϕ*-irreducible and aperiodic Markov chain. In addition, Pham (1986) obtained conditions for random coefficient autoregressive (RCA) models to be* ϕ*-irreducible. In this article, we simply impose the assumptions of* ϕ*-irreducibility and aperiodicity on* $\{\mathbb{X}_t\}$, *which are common settings among literature, see, for example, Chen and Tsay (1993). It is of particular interest to explore the conditions under which* $\{\mathbb{X}_t\}$ *is* ϕ*-irreducibility and aperiodicity and we leave this as a future topic. Moreover, the moment conditions* $E\{[\lambda_{\max}(\boldsymbol{A}_{U_t})]^2\} < 1$ *in Assumption D1 is used to bound the random matrices* $\boldsymbol{A}_{U_t}(Z_t)$, *which is similar to the condition in Carrasco and Chen (2002). We stress that we are not seeking to achieve the weakest possible regularity conditions for probabilistic properties of model* (D.3), *but instead focusing on constructing varying interdependences among conditional quantiles.*

Proposition D.1. *Under Assumption D, if* $\mathbb{X}_0$ *is initialized from the invariant measure, then,* $\{\mathbb{X}_t\}$ *defined in* (D.3) *is a strictly stationary and* α*-mixing process.*

To prove Proposition D.1, we first need to prove following lemma.

Lemma D.3. *Under Assumption D, for any* $\mathbb{W} = (w_1, \ldots, w_{\kappa(p+q)})^\top$*, we have* $\|\boldsymbol{A}_{U_t}(Z_t)\mathbb{W}\|_{\psi,2} \leq \|\boldsymbol{A}_{U_t}|\mathbb{W}|\|_{\psi,2}$*. Here,* $\boldsymbol{A}_{U_t}(Z_t)$ *is defined in* (D.3)*,* $\boldsymbol{A}_{U_t}$ *is defined previously and* $|\mathbb{W}| = (|w_1|, \ldots, |w_{\kappa(p+q)}|)^\top$.

Proof. Similar to the proof of Lemma A.1 in Chen and Tsay (1993), let $\boldsymbol{A}_{U_t}(Z_t)\mathbb{W} = (d_1, \ldots, d_{\kappa(p+q)})^\top$ and $\boldsymbol{A}_{U_t}|\mathbb{W}| = (g_1, \ldots, g_{\kappa(p+q)})^\top$. Then, for $\iota = \kappa+1, \ldots, \kappa p$ and for $\iota = \kappa p + \kappa + 1, \ldots, \kappa(p+q)$, we have $|d_\iota| = g_\iota$. For $\iota = 1, \ldots, \kappa$ and for $\iota' = \kappa p + 1, \ldots, \kappa p + \kappa$, by Assumption D,

$$
\begin{aligned}
|d_\iota| = &|\beta_{1\iota 1}(U_{\iota t}, Z_t)w_1 + \cdots + \beta_{p\iota\kappa}(U_{\iota t}, Z_t)w_{\kappa p} + \gamma_{1\iota 1}(U_{\iota t}, Z_t)w_{\kappa p+1} + \cdots + \\
&\gamma_{q\iota\kappa}(U_{\iota t}, Z_t)w_{\kappa(p+q)}| \\
\leq &|\beta_{1\iota 1}(U_{\iota t}, Z_t)w_1| + \cdots + |\beta_{p\iota\kappa}(U_{\iota t}, Z_t)w_{\kappa p}| + |\gamma_{1\iota 1}(U_{\iota t}, Z_t)w_{\kappa p+1}| + \cdots + \\
&|\gamma_{q\iota\kappa}(U_{\iota t}, Z_t)w_{\kappa(p+q)}| \\
\leq &|\beta_{1\iota 1}(U_{\iota t})w_1| + \cdots + |\beta_{p\iota\kappa}(U_{\iota t})w_{\kappa p}| + |\gamma_{1\iota 1}(U_{\iota t})w_{\kappa p+1}| + \cdots + |\gamma_{q\iota\kappa}(U_{\iota t})w_{\kappa(p+q)}| = g_\iota,
\end{aligned}
$$

and

$$
\begin{aligned}
|d_{\iota'}| = &|\beta_{1(\iota'-\kappa p)1,\tau}(Z_t)w_1 + \cdots + \beta_{p(\iota'-\kappa p)\kappa,\tau}(Z_t)w_{\kappa p} + \gamma_{1(\iota'-\kappa p)1,\tau}(Z_t)w_{\kappa p+1} + \cdots + \\
&\gamma_{q(\iota'-\kappa p)\kappa,\tau}(Z_t)w_{\kappa(p+q)}| \\
\leq &|\beta_{1(\iota'-\kappa p)1,\tau}(Z_t)w_1| + \cdots + |\beta_{p(\iota'-\kappa p)\kappa,\tau}(Z_t)w_{\kappa p}| + |\gamma_{1(\iota'-\kappa p)1,\tau}(Z_t)w_{\kappa p+1}| \\
&+ \cdots + |\gamma_{q(\iota'-\kappa p)\kappa,\tau}(Z_t)w_{\kappa(p+q)}| \\
\leq &|\beta_{1(\iota'-\kappa p)1,\tau}w_1| + \cdots + |\beta_{p(\iota'-\kappa p)\kappa,\tau}w_{\kappa p}| + |\gamma_{1(\iota'-\kappa p)1,\tau}w_{\kappa p+1}| + \cdots + \\
&|\gamma_{q(\iota'-\kappa p)\kappa,\tau}w_{\kappa(p+q)}| = g_{\iota'}.
\end{aligned}
$$

Hence, $\|\boldsymbol{A}_{U_t}(Z_t)\mathbb{W}\|_{\psi,2} \leq \|\boldsymbol{A}_{U_t}|\mathbb{W}|\|_{\psi,2}$. □

Proof of Proposition D.1:

Proof. By Proposition 3 in Carrasco and Chen (2002) and Lemma 2 in Pham (1986), Assumption D1 implies $\|\boldsymbol{A}_{U_t}\|_{\psi,2} < 1$ for all $U_{it} \in [0,1]$. Then, we can find $0 < \delta < 1$ and $\varrho > 0$, such that

$\| \prod_{\jmath=0}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \|_{\psi,2} < 1 - \delta$. Consequently, by Assumption D2 and Lemma D.3, for some constant $C > 0$,

$$
\begin{aligned}
E(\|\mathbb{X}_{t+\varrho}\| | \mathbb{X}_t = \mathbb{X}) = & E\left(\left\| \prod_{\jmath=0}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}}(Z_{t+\jmath}) \mathbb{X}_t + \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}}(Z_{t+\imath}) \right] \boldsymbol{D}_{U_{t+\jmath}}(Z_{t+\jmath}) \right\| \Bigg| \mathbb{X}_t = \mathbb{X} \right) \\
& + E\left(\left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}}(Z_{t+\imath}) \right] \boldsymbol{\mu}(Z_{t+\jmath}) \right\| \Bigg| \mathbb{X}_t = \mathbb{X} \right) \\
\leq & \left[\left\| \prod_{\jmath=0}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} |\mathbb{X}| \right\|_{\psi,2} \right] + C \cdot E\left(\left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] |\boldsymbol{D}_{U_{t+\jmath}}| \right\| \Bigg| \mathbb{X}_t = \mathbb{X} \right) \\
& + C \cdot E\left(\left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] \right\| \right) \\
\leq & \left[\left\| \prod_{\jmath=0}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right\|_{\psi,2} \right] \|\mathbb{X}\| + C \cdot E \left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] |\boldsymbol{D}_{U_{t+\jmath}}| \right\| \\
& + C \cdot E\left(\left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] \right\| \right)
\end{aligned}
$$

$$
\leq (1-\delta)\|\mathbb{X}\| + C \cdot E \left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] |\boldsymbol{D}_{U_{t+\jmath}}| \right\| + C \cdot E\left(\left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] \right\| \right),
$$

where each element of $\boldsymbol{D}_{U_t} = (\check{\gamma}_{10}(U_{1t}), \ldots, \check{\gamma}_{\kappa 0}(U_{\kappa t}), 0_{1\times\kappa(p+q-1)})^\top$ is defined in Assumption D2 and the first inequality follows from Jensen's inequality. Notice that $E \left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] \right\|$ is bounded and by Assumption D2, $E\|\boldsymbol{D}_{U_t}\|$ is bounded, so that $E \left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] |\boldsymbol{D}_{U_{t+\jmath}}| \right\|$ is bounded and the bound does not depend on $\mathbb{X}$ and Z_t. Thus, we can find a sufficiently large $M > 0$ such that when $\|\mathbb{X}\| > M$,

$$
(1-\delta)\|\mathbb{X}\| + C \cdot E \left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] |\boldsymbol{D}_{U_{t+\jmath}}| \right\| + C \cdot E\left(\left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] \right\| \right) \leq (1-\delta_1)\|\mathbb{X}\|,
$$

where $0 < \delta_1 < 1$. Hence, the compact set $K = \{\mathbb{X} : \|\mathbb{X}\| \leq M\}$ satisfies that when $\mathbb{X} \notin K$, $E(\|\mathbb{X}_{t+\varrho}\| | \mathbb{X}_t = \mathbb{X}) < (1-\delta_1)\|\mathbb{X}\|$. By Lemmas 1.1 and 1.2 in Chen and Tsay (1993), $\{\mathbb{X}_t\}$ is geometrically ergodic. If $\mathbb{X}_0$ is initialized from the invariant measure, then, by the results of Pham (1986), $\{\mathbb{X}_t\}$ is strictly stationary and α-mixing. $\square$

References

Belloni, A., and Chernozhukov, V. (2011), l_1-penalized Quantile Regression in High-dimensional Sparse Models, *Annals of Statistics*, 39, 82-130.

Bercu, B., and Touati, A. (2008), Exponential Inequalities for Self-normalized Martingales With Applications, *The Annals of Applied Probability*, 18(4), 1848-1869.

Bertsimas, D., and Tsitsiklis, J. (1997), *Introduction to Linear Optimization.* Athena Scientific, Belmont, MA.

Bickel, P.J. (1975), One-Step Huber Estimates in the Linear Model, *Journal of the American Statistical Association*, 70(350), 428-434.

Cai, Z., and Masry, E. (2000), Nonparametric Estimation in Nonlinear ARX Time Series Models: Local Linear Fitting and Projections, *Econometric Theory*, 16(4), 465-501.

Cai, Z., and Xu, X. (2008), Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models, *Journal of the American Statistical Association*, 103(484), 1595-1608.

Carrasco, M., and Chen, X. (2002), Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, *Econometric Theory*, 18(1), 17-39.

Chan, K.S., and Tong, H. (1985), On the Use of the Deterministic Lyapunov Function for the Ergodicity of Stochastic Difference Equations, *Advanced Applied Probability*, 17(3), 667-678.

Chen, B., and Hong, Y. (2016), Detecting for Smooth Structural Changes in GARCH Models, *Econometric Theory*, 32(3), 740-791.

Chen, R., and Tsay, R.S. (1993), Functional Coefficient Autoregressive Model, *Journal of the American Statistical Association*, 88(421), 298-308.

Chernozhukov, V., Härdle, W.K., Huang, C., and Wang, W. (2021), LASSO-driven Inference in Time and Space, *Annals of Statistics*, 49(3), 1702-1735.

Durrett, R. (1996), *Probability: Theory and Examples (Second Edition)*, Belmont, CA: Duxbury Press.

Hall, P., and Heyde, C.C. (1980), *Martingale Limit Theory and Its Application*, Academic Press: New York.

Horowitz, J.L., and Lee, S. (2005), Nonparametric Estimation of an Additive Quantile Regression Model, *Journal of the American Statistical Association*, 100(472), 1238-1249.

Horowitz, J.L., and Mammen, E. (2004), Nonparametric Estimation of an Additive Model With a Link Function, *Annals of Statistics*, 32, 2412-2443.

Knight, K. (1998), Limiting Distributions for L_1 Regression Estimators under General Conditions, *Annals of Statistics*, 26, 755-770.

Koenker, R., and Xiao, Z. (2006), Quantile Autoregression, *Journal of the American Statistical Association*, 101(475), 980-990.

Koenker, R., and Zhao, Q. (1996), Conditional Quantile Estimation and Inference for ARCH Models, *Econometric Theory*, 12(5), 793-813.

Pham, D.T. (1986), The Mixing Property of Bilinear and Generalized Random Coefficient Autoregressive Models, *Stochastic Processes and Their Applications*, 23(2), 291-300.

Ruppert, D., and Carroll, R. J. (1980), Trimmed Least Squares Estimation in the Linear Model, *Journal of the American Statistical Association*, 75, 828-838.

Sherwood, B., and Wang, L. (2016), Partially Linear Additive Quantile Regression in Ultra-high Dimension, *Annals of Statistics*, 44(1), 288-317.

Tang, Y., Song, X., Wang, H.J., and Zhu, Z. (2013), Variable Selection in High-Dimensional Quantile Varying Coefficient Models, *Journal of Multivariate Analysis*, 122(1), 115-132.

Van Der Vaart, A.W., and Wellner, J.A. (1996), *Weak Convergence and Empirical Processes*, Springer, New York, NY.

Welsh, A. (1989), On M-Processes and M-Estimation, *Annals of Statistics*, 17(1), 337-361.

Wu, W.B. (2005), Nonlinear System Theory: Another Look at Dependence, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102, National Acad Sciences, pp. 14150-14154.

Xiao, Z., and Koenker, R. (2009), Conditional Quantile Estimation for Generalized Autoregressive Conditional Heteroscedasticity Models, *Journal of the American Statistical Association*, 104(488), 1696-1712.

Xu, X. (2005), Semi-parametric and Nonparametric Estimation for Dynamic Quantile Regression Models, *PhD Dissertation*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.