

# A Nonparametric Dynamic Network via Multivariate Quantile Autoregressions

Zongwu Cai, Xiyuan Liu

Department of Economics, University of Kansas, Lawrence, KS 66045, USA

March 16, 2022

---

**Abstract:** In this article, we propose a vector autoregressive model for conditional quantiles with functional coefficients to construct a novel class of nonparametric dynamic network systems, of 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 two-stage procedure to estimate functionals in the dynamic network system by 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; 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; see, for instance, White, Kim and Manganelli (2015), Abrian and Brunnermeier (2016), Härdle, Wang and Yu (2016), Zhu, Wang, Wang and Härdle (2019) and the references therein. It is well known that when the distribution of the dependent variable has heavy-tails, heteroscedasticity, and/or outliers, the quantile regression is more reliable than mean regression models. The reader is referred to the review papers by Koenker (2005) and Koenker, Chernozhukov, He and Peng (2017) for more applications of quantile regression.

Among developments of quantile methods in the statistics literature, dynamic quantile models have attracted intensively attentions in the recent two decades. Previous researches in this area were mainly motivated by estimating Value-at-Risk (VaR), which is essentially a procedure of estimating lower-tail conditional quantile of financial return distribution. Some early works include, but not limited to, the autoregressive model for conditional quantiles (CaViaR) as in Engle and Manganelli (2004), the dynamic additive quantile model proposed in Gouriéroux and Jasiak (2008), and the conditional quantile estimation for generalized autoregressive conditional heteroscedasticity (GARCH)-type model studied by Xiao and Koenker (2009), and among others. In addition, dynamic quantile models are naturally suitable for capturing the dependence between the lower-tail conditional quantile of the distribution of financial returns and its lag or other covariates (also called tail dependence). For example, White et al. (2015) proposed an innovative method to estimate directly the sensitivity of 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 et al. (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 et al. (2019) constructed a quantile autoregressive model that embeds the observed

dependency structure in a dynamic network. The tail dependence is in particular important in reflecting the risk interdependence and contains network information in a financial system. To the best of our knowledge, much of the existing literature assumed constant tail dependence in their models or focused on the response of conditional quantile to endogenous variables or shocks. However, numerous studies have documented temporal changes of risk interdependence in financial time series and discussed their possible origins and relation to spillover effects; see, for example, Billio, Getmansky, Lo and Pelizzon (2012), Diebold and Yılmaz (2014), Härdle et al. (2016), Yang and Zhou (2017), Liu, Ji and Fan (2017), Ando and Bai (2020) and the 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. With these backgrounds, it is desirable to consider modeling the interaction between varying patterns of tail dependence and macroeconomic circumstances. These theoretical and empirical studies inspire us to build a more general framework to capture the time-varying interdependences among dynamic quantiles.

In this article, we propose a nonparametric approach involving multivariate dynamic quantile models with nonlinear structures. Different from previous studies, we capture nonlinearities in data by using a functional coefficient setting, which allows coefficients of the multivariate dynamic quantile models to vary with a smoothing variable. 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. To interpret features of varying interdependences within various conditional quantiles, we form a VAR model with functional coefficients where the quantiles of several random variables depend on lagged quantiles and other lagged covariates. For this reason, this model is termed as a functional-coefficient VAR model for dynamic quantiles (FCVAR-DQ) and is presented in (1) later. In an effort to study nonlinear relationship between the quantile of response variable and its covariates, various smoothing techniques (e.g., kernel methods, splines,

and their variants) have been used to estimate the nonparametric quantile regression for both independent and time series data, to name just a few, He and Ng (1999), Honda (2000, 2004), Wei and He (2006), Kim (2007), Cai and Xu (2008), Qu and Yoon (2015), and Li, Li and Li (2021). Among many kinds of methods, we adapt one of modeling methods to analyze dynamic quantiles, called the functional coefficient modeling approach. Compared with the existing literature, our approach is different mainly in three parts. First, we provide a kernel-based estimation framework for a new type of dynamic quantile model, which imposes relatively less restriction on model's structure. Second, our model admits nonlinearities of tail dependence, which can be ignored commonly by dynamic quantile models with fixed coefficients. Third, the proposed model allows for studying interaction between tail dependence and the variable of interest.

One of our motivations for this study comes from analyzing the dynamic mechanism of financial network in international equity markets. It is well documented in the literature that financial systems contain enormous numbers of institutions that interplay with each other. These interactions form a financial network in which a node represents each institution and a linkage between two nodes acts as an observable or unobservable interaction of some forms between two institutions. Also, it is well-established that the possibility of major financial distress is closely related to the degree of correlation among the assets of institutions and how sensitive they are to the changes in economic conditions. Based on these intuitions, provided that the node of a network is represented by the VaR of returns of institutions' assets or of market indexes, one may construct a financial network that can capture interdependences among VaRs within the financial system. Since VaRs and interdependences among them appear to be unobservable in practice, as addressed in Sewell and Chen (2015), Zhu et al. (2019), Bräuning and Koopman (2020) and Lee, Li and Wilson (2020), it is unnecessarily feasible to apply commonly known technologies that have access to the binary data with observed network structures for estimating the risk network formed by VaRs. An influential precedent of analyzing the network topology of unobservable connectedness of risk attributes to the paper by Diebold and Yilmaz (2014) by

constructing a risk network based on forecast error variance decompositions of classical VAR models and studying the volatility connectedness by methods of network analysis. Compared to the literature thus far, we consider capturing unobserved interconnectedness of tail risk among institutions in the dynamic network, which can not be achieved by models with observed network data and by measuring conditional correlation as in Diebold and Yılmaz (2014). Moreover, in order to illustrate overall patterns of time-varying network of risk across institutions, the main interest in this paper lies in modeling the relationship between the general states of economy and a financial network formed by VaRs of global major market index's return series. More specifically, we allow interdependences among VaRs of market index's return series to vary with a smoothing variable of economic status to capture the dynamic changes. Some recent studies found increasing evidences to show that the variation of risk interdependence not only reveals the behavior of spillover effects of risk but also contains the information about the stability of financial systems; see, e.g., Acemoglu, Ozdaglar and Tahbaz-Salehi (2015). Both practitioners and policymakers may be interested in knowing how a financial network changes with the macroeconomic climate or financial market circumstances, and the way to evaluate the influences of economic policies to the whole network within the financial market. Extensive reviews about financial network can be found in Diebold and Yılmaz (2014) and Härdle et al. (2016). The empirical study in this paper shows that the proposed FCVAR-DQ model should be suitable for estimating a novel class of dynamic financial network and providing some new insights. A detailed analysis of this class of nonparametric financial network is reported in Section 4.

Lastly, our contributions to the literature can be summarized as follows. First, the model setting in this paper (see (1) later) is general enough to nest many well-known dynamic quantile models in the literature; see, for example, the CaViaR model proposed by Engle and Manganelli (2004) and further studied 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 interdependence of risk and the state variable of economy or time. Third, a new and simple-to-implement estimation procedure is developed for estimating the proposed quantile model with highly nonlinear structure and latent covariates. Finally, a large sample theory for the proposed estimator is established to construct confidence intervals for functional coefficients in the empirical study.

The rest of this paper is organized as follows. In Section 2, the model setup and the two-stage estimation procedure are presented for the FCVAR-DQ model. In addition, a large sample theory for the proposed estimator is investigated in this section too, together with constructing a consistent estimator of the asymptotic covariance matrix. 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 all the technical proofs are gathered in the Appendix. 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.

## 2 FCVAR Model for Dynamic Quantiles

### 2.1 Model Setup

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

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_{it}) + \sum_{s=1}^q \gamma_{i,s,\tau}^T(Z_{it}) \mathbf{q}_{\tau,t-s} + \sum_{l=1}^p \beta_{i,l,\tau}^T(Z_{it}) \mathbb{Y}_{t-l} \quad (1)$$

for some  $p$  and  $q$ , where  $\mathbf{q}_{\tau,t} = (q_{\tau,t,1}, \dots, q_{\tau,t,\kappa})^T$  and  $\mathbb{Y}_t$  is a  $\kappa_1 \times 1$  vector of covariates, 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  $\tau$ , both  $\boldsymbol{\gamma}_{i,s,\tau}(\cdot) = (\gamma_{si1,\tau}(\cdot), \dots, \gamma_{si\kappa,\tau}(\cdot))^T$  and  $\boldsymbol{\beta}_{i,l,\tau}(\cdot) = (\beta_{li1,\tau}(\cdot), \dots, \beta_{li\kappa_1,\tau}(\cdot))^T$  are  $\kappa \times 1$  and  $\kappa_1 \times 1$  vectors of functional coefficients, respectively, and they are allowed to depend on  $\tau$  too. Here,  $Z_{it}$  is an observable scalar smoothing variable, which might be one part of  $\mathbb{Y}_{t-l}$  and/or time 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 similar to the setting of the multi-quantile CaViaR model as in White, Kim and Manganelli (2008), one may further generalize model (1) by allowing  $\tau$  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  $\mathbb{Y}_t = (|Y_{1t}|, \dots, |Y_{\kappa t}|)^T$  with  $|\cdot|$  representing absolute value. 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  $q = p = 1$ ,  $\kappa = \kappa_1$ ,  $\mathbb{Y}_t = (|Y_{1t}|, \dots, |Y_{\kappa t}|)^T$ , and  $Z_{it} = Z_t$  for all  $1 \leq i \leq \kappa$ . Then, model (1) can be rewritten as

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

where  $\boldsymbol{g}_{i,\tau}(\cdot) = (\gamma_{i0,\tau}(\cdot), \gamma_{i1,\tau}(\cdot), \dots, \gamma_{i\kappa,\tau}(\cdot), \beta_{i1,\tau}(\cdot), \dots, \beta_{i\kappa,\tau}(\cdot))^T$  is a  $(2\kappa + 1) \times 1$  vector of functional coefficients and  $\boldsymbol{X}_t = (1, q_{\tau,t-1,1}, \dots, q_{\tau,t-1,\kappa}, |Y_{1(t-1)}|, \dots, |Y_{\kappa(t-1)}|)^T$ .

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 sensitivity of VaR of returns for one portfolio at time  $t$  to 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

$$\mathbf{\Gamma}_{1,\tau}(Z_t) = (\gamma_{ij,\tau}(Z_t))_{\kappa \times \kappa}. \quad (3)$$

Then, (2) can be expressed as a matrix form, which, indeed, is a FCVAR model for  $\mathbf{q}_{\tau,t}$  with exogenous variables,

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

where  $\boldsymbol{\gamma}_{0,\tau}(Z_t)$  and  $\mathbf{\Gamma}_{\beta,1,\tau}(Z_t)$  are defined obviously. Therefore,  $\mathbf{\Gamma}_{1,\tau}(Z_t)$  in (3) can serve as a dynamic network system changing with both  $\tau$  and some information variable  $Z_t$ , and it is in a nonparametric nature, so that it is a nonparametric dynamic network. Notice that the general setting in the dynamic network system (3) 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 Abrian and Brunnermeier (2016) and Härdle et al. (2016) as special cases.

To investigate the large sample behavior of the proposed estimator (see Theorem 1 later), it is assumed throughout this article that the process  $\{(Y_{it}, x_{it}, Z_t)\}$  in model (1) is strictly stationary and  $\alpha$ -mixing (strongly mixing). Indeed, in the Appendix (see Appendix B), we provide some regularity conditions to show that under these conditions, the joint process  $\{(Y_{it}, x_{it}, Z_t, q_{\tau,t,i})\}$  generated by model (1) is strictly stationary and  $\alpha$ -mixing. Actually, sufficient conditions for the mixing property of nonlinear time series have been studied extensively in literature. By Pham (1986), a geometrically ergodic time series is an  $\alpha$ -mixing sequence. Meanwhile, it is well-known that an ergodic Markov process initiated from its invariant distribution is (strictly) stationary. Thus, geometrical ergodicity plays an important role in establishing strictly stationarity and  $\alpha$ -mixing properties. Some results in this direction include the papers by Chen and Tsay (1993) and Cai, Fan and Yao (2000), providing some sufficient conditions to ensure geometrical ergodicity for functional-coefficient autoregressive time series models without rigorously theoretical justifications. In addition, An and Chen (1997) and An and Huang (1996) surveyed various sufficient conditions for the ergodicity of nonlinear autoregressive models. Also, Cai and Masry



(2000) presented some sufficient conditions for additive nonlinear autoregressive models with exogenous regressors to be stationary and strongly mixing. 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 network. Therefore, we provide some sufficient conditions that imply these important probabilistic properties and corresponding rigorously theoretical justifications in the Appendix (see Appendix B).

**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 discussed 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  $\sigma_{it}^2$  is the conditional variance of  $Y_{it}$  and  $e_{it}$  is an independent and identically distributed (i.i.d.) 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  $\kappa$  ( $\kappa \geq 1$ ) returns extended from the setting in Taylor (1986) as follows

$$\sigma_{it} = \gamma_{i0}(Z_t) + \sum_{s=1}^q \gamma_{i,s}^T(Z_t) \Sigma_{t-s} + \sum_{l=1}^p \beta_{i,l}^T(Z_t) \mathbb{Y}_{t-l}, \quad (4)$$

where  $\Sigma_t = (\sigma_{it}, \dots, \sigma_{\kappa t})^T$  and  $\mathbb{Y}_t = (|Y_{1t}|, \dots, |Y_{\kappa t}|)^T$ , then, model (1) reduces to following dynamic quantile model:

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_t) + \sum_{s=1}^q \gamma_{i,s}^T(Z_t) \mathbf{q}_{\tau,t-s} + \sum_{l=1}^p \beta_{i,l,\tau}^T(Z_t) \mathbb{Y}_{t-l}, \quad (5)$$

where  $\gamma_{i0,\tau}(\cdot) = \gamma_{i0}(\cdot) F_e^{-1}(\tau)$ ,  $\gamma_{i,s}(\cdot) = (\gamma_{si1}(\cdot), \dots, \gamma_{si\kappa}(\cdot))^T$  and  $\beta_{i,l,\tau}(\cdot) = (\beta_{li1,\tau}(\cdot), \dots, \beta_{li\kappa,\tau}(\cdot))^T$  with  $\beta_{li\kappa,\tau}(\cdot) = \beta_{lij}(\cdot) F_e^{-1}(\tau)$ . Notice that if  $\gamma$ 's and  $\beta$ 's in (5) are constant, model (5) 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  $\mathbf{q}_{\tau,t}$  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  $\{\gamma_{i,s,\tau}(Z_t)\}_{i=1,s=1}^{\kappa,q}$  and  $\{\beta_{i,l,\tau}(Z_t)\}_{i=1,l=1}^{\kappa,p}$  are all monotone increasing functions with respect to  $\tau$ , 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.** (Selection of  $Z_t$ ). Of importance is 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 et al. (2000) or Cai et al. (2015) in practice.

## 2.2 Two-stage Estimation Procedure

Since the estimation procedures for (1) and (2) are the same, we aim at estimating functional coefficients  $\mathbf{g}_{i,\tau}(\cdot)$  in the model defined in (2) for simplicity. Because  $q_{\tau,t-1,i}$  in  $\mathbf{X}_t$  depends on unknown functional coefficients  $\mathbf{g}_{i,\tau}(\cdot)$ , model (2) is more complicated than functional coefficient models with observed data. Our procedures consist of two steps. The first is to estimate latent  $q_{\tau,t-1,i}$ , and then we perform locally weighted estimation for functional coefficients using the estimated  $q_{\tau,t-1,i}$  from the first step. In this paper, we only focus on estimating functional coefficients in (2), rather than jointly forecasting  $q_{\tau,t,i}$  or doing impulse response analysis. So, it is sufficient to estimate  $\mathbf{g}_{i,\tau}(\cdot)$  in an equation-by-equation way for different  $i$ . Thus, by abuse of

notation,  $i$  will be dropped in what follows.

Given (1) and (2), let  $\boldsymbol{\gamma}_{0,\tau}(Z_t)$  define as earlier as  $(\gamma_{10,\tau}(Z_t), \dots, \gamma_{\kappa 0,\tau}(Z_t))^T$  and denote  $\boldsymbol{\Gamma}_{s,\tau}(Z_t)$  as a matrix with entries  $\gamma_{sij,\tau}(Z_t)$  and  $\boldsymbol{\Gamma}_{\beta,l,\tau}(Z_t)$  as a matrix with entries  $\beta_{lij,\tau}(Z_t)$ , for  $s = 1, \dots, q$  and  $l = 1, \dots, p$ . Furthermore, define  $\mathcal{A}_\tau(\mathcal{L}) = \sum_{l=1}^p \boldsymbol{\Gamma}_{\beta,l,\tau}(Z_t) \mathcal{L}^l$  and  $\mathcal{B}_\tau(\mathcal{L}) = I_\kappa - \sum_{s=1}^q \boldsymbol{\Gamma}_{s,\tau}(Z_t) \mathcal{L}^s$ , where each entry is a lag polynomial and  $\mathcal{L}$  denotes the lag operator. Then, under Assumption A1 presented in Section 2.3, ensuring the invertibility of  $\mathcal{B}_\tau(\mathcal{L})$ , model (1) becomes to the following formulation

$$\mathbf{q}_{\tau,t} = \mathcal{B}_\tau(\mathcal{L})^{-1} \boldsymbol{\gamma}_{0,\tau}(Z_t) + \mathcal{B}_\tau(\mathcal{L})^{-1} \mathcal{A}_\tau(\mathcal{L}) \mathbb{Y}_t.$$

Here,  $\mathcal{B}_\tau(\mathcal{L})^{-1} \boldsymbol{\gamma}_{0,\tau}(Z_t)$  and  $\mathcal{B}_\tau(\mathcal{L})^{-1} \mathcal{A}_\tau(\mathcal{L})$  can be represented by  $C_{0,t,\tau} \boldsymbol{\gamma}_{0,\tau}(Z_t)$  and a matrix series  $\sum_{l=1}^{\infty} C_{l,t,\tau} \mathcal{L}^l$  for all  $Z_t$ , respectively. Now, let  $\alpha_{0,\tau}(\cdot)$  be the  $i$ th row of matrix  $C_{0,t,\tau} \boldsymbol{\gamma}_{0,\tau}(Z_t)$  and  $\boldsymbol{\alpha}_{l,\tau}(\cdot) = (\alpha_{l1,\tau}(\cdot), \dots, \alpha_{l\kappa,\tau}(\cdot))^T$  be the  $i$ th row of matrix  $C_{l,t,\tau}$ . Therefore, with the definitions of  $\alpha_{0,\tau}(\cdot)$  and  $\boldsymbol{\alpha}_{l,\tau}(\cdot)$ , we can first approximate the latent  $q_{\tau,t}$  by using a functional-coefficient quantile function:

$$q_{\tau,t} = \alpha_{0,\tau}(Z_t) + \sum_{l=1}^{\infty} \boldsymbol{\alpha}_{l,\tau}^T(Z_t) \mathbb{Y}_{t-l}, \quad (6)$$

where the coefficients  $\boldsymbol{\alpha}_{l,\tau}(\cdot)$  satisfies summability conditions implied by Assumption A1. Then, each entry of  $\boldsymbol{\alpha}_{l,\tau}(\cdot)$  decreases at a geometric rate; that is, there exist positive constants  $\rho < 1$  and  $c$ , such that  $\max_{1 \leq t \leq n} |\alpha_{lj,\tau}(Z_t)| \leq c \rho^l$  for  $j = 1, \dots, \kappa$ . Since  $\alpha_{lj,\tau}(\cdot)$  decreases geometrically, by choosing an appropriate  $m_n = m(n) = m$ , we study following truncated equation (7) with increasing dimension of covariates:

$$q_{\tau,t} = \alpha_{0,\tau}(Z_t) + \sum_{l=1}^{m_n} \boldsymbol{\alpha}_{l,\tau}^T(Z_t) \mathbb{Y}_{t-l} \equiv \mathbf{W}_t^T \boldsymbol{\alpha}_\tau(Z_t) = q_\tau(Z_t, \mathbf{W}_t), \quad (7)$$

where  $\mathbf{W}_t = (1, \mathbb{Y}_{t-1}^T, \dots, \mathbb{Y}_{t-m}^T)^T$  is a  $(\kappa m + 1) \times 1$  vector of covariates and  $\boldsymbol{\alpha}_\tau(\cdot) = (\alpha_{0,\tau}(\cdot), \boldsymbol{\alpha}_{1,\tau}^T(\cdot), \dots, \boldsymbol{\alpha}_{m,\tau}^T(\cdot))^T$  is a  $(\kappa m + 1) \times 1$  vector of functional coefficients. Note that (7) can be regarded as an approximation of (6) and is similar to the model in Cai and Xu (2008). Under smoothness condition of coefficient functions  $\boldsymbol{\alpha}_\tau(\cdot)$  presented later in Assumption A2 in Section 2.3, for any given grid point  $z_0 \in \mathbb{R}$ , when  $Z_t$  is in a neighborhood of  $z_0$ ,  $\boldsymbol{\alpha}_\tau(Z_t)$  can be approxi-

mated by a polynomial function as  $\boldsymbol{\alpha}_\tau(Z_t) \approx \sum_{r=0}^w \boldsymbol{\alpha}_\tau^{(r)}(z_0)(Z_t - z_0)^r/r!$ , where  $\approx$  denotes the approximation by ignoring the higher orders and  $\boldsymbol{\alpha}_\tau^{(r)}(\cdot)$  is the  $r$ th derivative of  $\boldsymbol{\alpha}_\tau(\cdot)$ . Thus,  $q_{\tau,t} \approx \sum_{r=0}^w \mathbf{W}_t^T \boldsymbol{\delta}_{r,\tau} (Z_t - z_0)^r$ , where  $\boldsymbol{\delta}_{r,\tau} = \boldsymbol{\alpha}_\tau^{(r)}(z_0)/r!$ . Hence,  $\hat{\boldsymbol{\delta}} = \operatorname{argmin}_{\boldsymbol{\delta}} Q(\boldsymbol{\delta})$ , where  $Q(\boldsymbol{\delta})$  is the locally weighted loss function for fixed  $\kappa$ , given by

$$Q(\boldsymbol{\delta}) = \sum_{t=m+1}^n \rho_\tau \left\{ Y_t - \sum_{r=0}^w \mathbf{W}_t^T \boldsymbol{\delta}_r (Z_t - z_0)^r \right\} K_{h_1}(Z_t - z_0), \quad (8)$$

$\rho_\tau(y) = y[\tau - I(y < 0)]$  is called the “check” (loss) function,  $I(A)$  is the indicator function of any set  $A$ ,  $K(\cdot)$  is a kernel function,  $K_{h_1}(u) = K(u/h_1)/h_1$ , and  $h_1 = h_1(n)$  is a sequence of positive numbers tending to zero and controls the amount of smoothing used in estimation. In practice, if we smooth locally around  $Z_t$  and consider a local linear estimation, the locally weighted loss function (8) becomes to the following

$$Q_1(\boldsymbol{\delta}) = \sum_{s=m+1 \neq t}^n \rho_\tau \left\{ Y_s - \sum_{r=0}^1 \mathbf{W}_s^T \boldsymbol{\delta}_r (Z_s - Z_t)^r \right\} K_{h_1}(Z_s - Z_t). \quad (9)$$

After yielding  $\hat{\boldsymbol{\delta}}_{0,\tau}$  at  $\tau$  by minimizing (9),  $q_{\tau,t}$  can be estimated by  $\hat{q}_{\tau,t} = \mathbf{W}_t^T \hat{\boldsymbol{\delta}}_{0,\tau}$ .

**Remark 4.** (*Truncation parameter  $m(n)$* ). *Welsh (1989) and He and Shao (2000) studied non-linear M-estimation with increasing parametric dimension and discussed the possible expansion rate for the number of parameters  $m(n)$ . As for the quantile estimation for functional coefficient models with increasing dimension of covariates, Tang, Song, Wang and Zhu (2013) considered estimation and variable selection for high-dimensional quantile varying coefficient models based on B-spline approach. They showed that the oracle property for varying coefficients can be preserved when  $m_n^2 \log(p_n m_n)/n \rightarrow 0$ , where  $p_n$  is the dimension of covariates and  $m_n$  is a parameter associated with degree of polynomial and internal knots. In this step, we are interested in studying varying interdependences among conditional quantiles, rather than determining the optimal number for  $m$ . In addition, we focus on estimating (7) using kernel-based approaches, which is necessary in order to obtain asymptotic properties for functional coefficients. Under Assumption A10 in Section 2.3, it will suffice to consider a truncation  $m$  as a sufficiently large constant multiple of  $n^{1/7}$ , which is used in our simulation study in Section 3 and the empirical analysis*

in Section 4.

**Remark 5.** It is necessary to emphasize that  $\alpha_{0,\tau}(\cdot)$  and each component of  $\boldsymbol{\alpha}_{l,\tau}(\cdot)$  in (6) depend on  $\{Z_{t-l}\}_{l \geq 0}$ . Indeed, under the assumption of stationarity and Assumption A1,  $\alpha_{0,\tau}(\cdot)$  and  $\boldsymbol{\alpha}_{l,\tau}(\cdot)$  are well-defined and can be estimated on each  $Z_t$  by local smoothing approaches, regardless of the existence of other lagged  $Z_{t-l}$  in  $\alpha_{0,\tau}(\cdot)$  and  $\boldsymbol{\alpha}_{l,\tau}(\cdot)$ . Therefore, we use notations  $\alpha_{0,\tau}(Z_t)$  and  $\sum_{l=1}^{\infty} \boldsymbol{\alpha}_{l,\tau}^T(Z_t) \mathbb{Y}_{t-l}$  instead of  $\alpha_{0,\tau}(Z_t, Z_{t-1}, \dots, Z_{t-l})$  and  $\sum_{l=1}^{\infty} \boldsymbol{\alpha}_{l,\tau}^T(Z_t, Z_{t-1}, \dots, Z_{t-l}) \mathbb{Y}_{t-l}$  in (6) for notational simplicity.

To summarize, the following two-step procedures is proposed for estimating  $\mathbf{g}_{\tau}(\cdot)$ :

**Step One:** Choose the truncation parameter  $m = cn^{1/7}$  for some  $c > 0$  and estimate  $\hat{\boldsymbol{\delta}}_{0,\tau}$  at each  $Z_t$  by minimizing (9). Then, latent  $q_{\tau,t}$  is approximated by  $\hat{q}_{\tau,t} = \mathbf{W}_t^T \hat{\boldsymbol{\delta}}_{0,\tau}$ .

**Step Two:** Having obtained  $\hat{q}_{\tau,t}$  and given

$$\hat{\mathbf{X}}_t = (1, \hat{q}_{\tau,t-1,1}, \dots, \hat{q}_{\tau,t-1,\kappa}, |Y_{1(t-1)}|, \dots, |Y_{\kappa(t-1)}|)^T,$$

$\mathbf{g}_{\tau}(\cdot)$  is estimated by a local linear estimation method; see Cai and Xu (2008) for details. In particular, minimize the following locally (linear) weighted loss function  $Q_2(\Theta)$  at any given grid point  $z_0 \in \mathbb{R}$  to obtain the local linear estimate  $\hat{\Theta}$ , where

$$Q_2(\Theta) = \sum_{t=1}^n \rho_{\tau} \left\{ Y_t - \sum_{r=0}^1 \hat{\mathbf{X}}_t^T \Theta_{r,\tau} (Z_t - z_0)^r \right\} K_{h_2}(Z_t - z_0) \quad (10)$$

with  $\Theta_{r,\tau} = \mathbf{g}_{\tau}^{(r)}(\cdot)/r!$ . Similar to (9),  $K_{h_2}(u) = K(u/h_2)/h_2$  and  $h_2$  is the bandwidth used for this step, which is different from the bandwidth  $h_1$  used in (9); see Remark 6 later in Section 2.3 for more discussions. A further improvement can be achieved by applying iteration to the foregoing two-stage procedures.

## 2.3 Large Sample Theory

To study the asymptotic distribution of the nonparametric quantile estimator, we impose some technical conditions in this section. It is worthwhile to emphasize that the main focus in this paper is on estimating a new type of dynamic quantile model and constructing varying inter-

dependences among conditional quantiles, rather than exploring the weakest possible conditions for asymptotic theory.

**Assumption A.**

**A1:** Suppose that  $\mathcal{A}_\tau(\mathcal{L})$  and  $\mathcal{B}_\tau(\mathcal{L})$  defined in Section 2.2 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:** Each entry in the vector  $\boldsymbol{\alpha}_\tau(\cdot)$  is  $(w+1)$ th order continuously differentiable in a neighborhood of  $z_0$  for any  $z_0$ ; Similarly, each entry in the vector  $\mathbf{g}_\tau(\cdot)$  is  $(\varsigma+1)$ th order continuously differentiable in a neighborhood of  $z_0$  for any  $z_0$ .

**A3:**  $f_z(z)$  is a continuously marginal density of  $Z$  and  $f_z(z_0) > 0$ .

**A4:** The distribution of  $Y$  given  $Z$  and  $\mathbf{W}$  has an everywhere positive conditional density  $f_{Y|Z, \mathbf{W}}(\cdot)$ , which is bounded and satisfies the Lipschitz continuity condition. Here,  $\mathbf{W}_t$  is defined in (7). 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  $\nu_0 = \int K^2(\nu) d\nu$ .

**A5:**  $\{(Y_{it}, x_{it}, Z_t)\}$  is a strictly stationary sequence with  $\alpha$ -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  $\iota > (\delta-2)/\delta$ .

**A6:** There exist (small) positive constants  $\varpi_1 > 0$  and  $\varpi_2 > 0$  such that  $P\{\max_{1 \leq t \leq n} Y_t^2 > n^{\varpi_1}\} \leq \exp(-n^{\varpi_2})$ .

**A7:** Let  $\mathbf{B}_n = \frac{1}{n} \sum_{t=m+1}^n \mathbf{W}_t \mathbf{W}_t^T$  and denote the maximum and minimum eigenvalues of  $\mathbf{B}_n$  as  $\lambda_{\max}(\mathbf{B}_n)$  and  $\lambda_{\min}(\mathbf{B}_n)$ . Then,  $\liminf_{n \rightarrow \infty} \lambda_{\min}(\mathbf{B}_n) > 0$ ,  $\limsup_{n \rightarrow \infty} \lambda_{\max}(\mathbf{B}_n) < \infty$ .

It is assumed that  $E\|\mathbf{W}_t\|^{\delta^*} \leq C m^{\delta^*/2}$  with  $\delta^* > \delta$ .

**A8:**  $\mathbf{D}(z_0) \equiv E[\mathbf{W}_t \mathbf{W}_t^T | Z_t = z_0]$  is positive-definite and continuous in a neighborhood of  $z_0$  and  $\mathbf{D}^*(z_0) \equiv E[\mathbf{W}_t \mathbf{W}_t^T f_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_t)) | Z_t = z_0]$  is positive-definite and continuous in a neighborhood of  $z_0$ .

**A9:**  $E\|\mathbb{Y}_t\|^{2\delta^*} < \infty$  with  $\delta^* > \delta$ .

**A10:** The bandwidth  $h_1$  satisfies  $h_1 \rightarrow 0$ ,  $nh_1 \rightarrow \infty$ ; The bandwidth  $h_2$  satisfies  $h_2 = O(n^{-1/5})$ ,  $h_2 \rightarrow 0$ ,  $nh_2 \rightarrow \infty$ . In addition,  $h_1 = o(h_2)$ ,  $mh_1 \rightarrow 0$ .

**A11:**  $f(\mathbf{w}, \boldsymbol{\omega} | \mathbf{Y}_0, \mathbf{Y}_\ell; \ell) \leq H < \infty$  for  $\ell \geq 1$ , where  $f(\mathbf{w}, \boldsymbol{\omega} | \mathbf{Y}_0, \mathbf{Y}_\ell; \ell)$  is the conditional density of  $(Z_0, Z_\ell)$  given  $(\mathbb{Y}_0 = \mathbf{Y}_0, \mathbb{Y}_\ell = \mathbf{Y}_\ell)$ .

**A12:**  $n^{1/2-\delta/4} h_2^{\delta/\delta^*-1/2-\delta/4} = O(1)$ .

**Remark 6.** Assumptions A1 is an invertibility condition for the functional coefficients to be well-defined, which is similar to that in Chen and Hong (2016). Assumptions A2-A4 are common in nonparametric literature. Assumption A5 is a standard assumption for  $\alpha$ -mixing. Assumption A6 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). Assumption A7 guarantees the asymptotic behavior of regression estimators with increasing dimension of covariates, which is similar to but slightly weaker than that in Welsh (1989). Assumptions A8 and A9 are commonly required for the model identification and ensure the convergence of  $\mathbf{B}_n$  to  $E[\mathbf{W}_t \mathbf{W}_t^T]$ , when  $\mathbf{W}_t$  is  $\alpha$ -mixing. The assumption  $h_1 = o(h_2)$  in Assumption A10 is about the under-smoothing at the step one, which is common for the two-stage nonparametric estimation approaches; see also Cai (2002) and Cai and Xiao (2012) for more discussions. The assumption  $mh_1 \rightarrow 0$  in A10 is necessary for the proof of stochastic equi-continuity. Assumption A11 is very standard and used for the proof under mixing conditions. Assumption A12 allows one to verify standard Lindeberg-Feller conditions for asymptotic normality of the proposed estimators in the proof of Theorem 1; see Cai and Xu (2008) for details on nonparametric quantile regressions models for  $\alpha$ -mixing time series.

Before stating the asymptotic behavior of  $\hat{\mathbf{g}}_\tau(z_0)$  in the following theorem, for notational simplicity, it needs to define some notations. Define  $\Omega^*(z_0) \equiv E[\mathbf{X}_t \mathbf{X}_t^T f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t)) | Z_t = z_0]$  with  $q_\tau(z_0, \mathbf{X}_t) = \mathbf{g}_\tau^T(z_0) \mathbf{X}_t$  and  $f_{Y|Z, \mathbf{X}}(\cdot)$ . In addition, let  $\Xi(z_0) \equiv \tau(1 - \tau)\nu_0[\Omega(z_0) - H_1(z_0) + H_2(z_0)]$ , where  $\Omega(z_0) \equiv E[\mathbf{X}_t \mathbf{X}_t^T | Z_t = z_0]$ ,  $H_1(z_0) = E[\mathbf{X}_t \mathbf{W}_t^T | Z_t = z_0](\mathbf{D}^*(z_0))^{-1} \Gamma^T(z_0) + \Gamma(z_0)(\mathbf{D}^*(z_0))^{-1} E[\mathbf{W}_t \mathbf{X}_t^T | Z_t = z_0]$ ,  $H_2(z_0) = \Gamma(z_0)(\mathbf{D}^*(z_0))^{-1} \mathbf{D}(z_0)(\mathbf{D}^*(z_0))^{-1} \Gamma^T(z_0)$ , and  $\Gamma(z_0) \equiv E\left\{f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t)) \mathbf{X}_t \mathbf{g}_\tau^T(z_0) \boldsymbol{\Pi}_t \middle| Z_t = z_0\right\}$  is a  $(2\kappa + 1) \times (\kappa m + 1)$  matrix, with  $\boldsymbol{\Pi}_t^T = (0_{1 \times (\kappa m + 1)}^T, \mathbf{W}_t, \dots, \mathbf{W}_t, 0_{\kappa \times (\kappa m + 1)}^T)$ . Now, the asymptotic normality of  $\hat{\mathbf{g}}_\tau(z_0)$  is presented

in the following theorem with its detailed proof relegated to the Appendix (see Appendix A).

**Theorem 1.** (*Asymptotic Normality*) Under Assumptions A1–A12, we have

$$\sqrt{nh_2} \left[ \hat{\mathbf{g}}_\tau(z_0) - \mathbf{g}_\tau(z_0) - \frac{h_2^2 \mu_2}{2} \mathbf{g}_\tau^{(2)}(z_0) + o_p(h_2^2) \right] \xrightarrow{d} \mathcal{N}(0, \Sigma_\tau(z_0)),$$

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

**Remark 7.** It is not surprising to see from Theorem 1 that the asymptotic bias  $h_2^2 \mu_2 \mathbf{g}_\tau^{(2)}(z_0)/2$  does not depend on  $h_1$ . Indeed, since the estimation in the step one is under-smoothed by Assumption A10, so that the part that relies on  $h_1$  in the asymptotic bias term disappears, see Lemma A.10 in the Appendix for more details. However, different from the conventional non-parametric estimation,  $\Xi(z_0)$  in the asymptotic variance term contains additional two terms  $H_1(z_0)$  and  $H_2(z_0)$ , which involve  $\mathbf{W}_t$  in the first step. This formation of asymptotic variance appears because of the fact that  $\hat{\mathbf{X}}_t$  contains  $\hat{q}_{\tau,t-1}$ , which is estimated in the step one of our two-stage approaches and therefore includes information of  $\mathbf{W}_t$ . Similar results of asymptotic variance were also obtained by Xiao and Koenker (2009), which can be seen as a nature of any two-stage approach; see, for example, Cai, Das, Xiong and Wu (2006) for more discussions.

**Remark 8.** (*Bandwidth Selection*) Finally, we would like to address how to select the bandwidth  $h_2$  at the second step. It is well known that the bandwidth plays an essential role in the trade-off between reducing bias and variance. In view of (10), 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  $h_2$  in (10), which is used in our simulation study in Section 3.

## 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  $\mathbf{D}(z_0)$ ,  $\mathbf{D}^*(z_0)$ ,  $\Gamma(z_0)$ ,  $H_1(z_0)$ ,  $H_2(z_0)$ ,  $\Omega(z_0)$  and



$\Omega^*(z_0)$  consistently. For this purpose, define  $\hat{\mathbf{D}}(z_0) = \sum_{t=1}^n \mathbf{W}_t \mathbf{W}_t^T K_{h_1}(Z_t - z_0)/n$  and  $\hat{\mathbf{D}}^*(z_0) = \sum_{t=1}^n w_{1t} \mathbf{W}_t \mathbf{W}_t^T K_{h_1}(Z_t - z_0)/n$ , where  $w_{1t} = I(\mathbf{W}_t^T \hat{\boldsymbol{\alpha}}_\tau(z_0) - \delta_{1n} < Y_t \leq \mathbf{W}_t^T \hat{\boldsymbol{\alpha}}_\tau(z_0) + \delta_{1n})/(2\delta_{1n})$  for any  $\delta_{1n} \rightarrow 0$  as  $n \rightarrow \infty$ . Similar to the proof in Cai and Xu (2008), one can show that  $\hat{\mathbf{D}}(z_0) = f_z(z_0)\mathbf{D}(z_0) + o_p(1)$  and  $\hat{\mathbf{D}}^*(z_0) = f_z(z_0)\mathbf{D}^*(z_0) + o_p(1)$ , respectively. Also, let  $\mathbf{E}_{xw}(z_0) = \sum_{t=1}^n \hat{\mathbf{X}}_t \mathbf{W}_t^T K_{h_2}(Z_t - z_0)/n$ . Clearly, the consistent estimators of  $\Gamma(z_0)$ ,  $H_1(z_0)$ ,  $H_2(z_0)$ ,  $\Omega(z_0)$  and  $\Omega^*(z_0)$  can be constructed as follows:  $\hat{\Gamma}(z_0) = \sum_{t=1}^n w_{2t} \hat{\mathbf{X}}_t \hat{\mathbf{g}}_\tau^T(z_0) \mathbf{\Pi}_t K_{h_2}(Z_t - z_0)/n$ ,  $\hat{H}_1(z_0) = \mathbf{E}_{xw}(z_0)(\hat{\mathbf{D}}^*(z_0))^{-1} \hat{\Gamma}^T(z_0) + \hat{\Gamma}(z_0)(\hat{\mathbf{D}}^*(z_0))^{-1} (\mathbf{E}_{xw}(z_0))^T$ ,  $\hat{\Omega}(z_0) = \sum_{t=1}^n \hat{\mathbf{X}}_t \hat{\mathbf{X}}_t^T K_{h_2}(Z_t - z_0)/n$ ,  $\hat{H}_2(z_0) = \hat{\Gamma}(z_0)(\hat{\mathbf{D}}^*(z_0))^{-1} \hat{\mathbf{D}}(z_0)(\hat{\mathbf{D}}^*(z_0))^{-1} \hat{\Gamma}^T(z_0)$ , and  $\hat{\Omega}^*(z_0) = \sum_{t=1}^n w_{2t} \hat{\mathbf{X}}_t \hat{\mathbf{X}}_t^T K_{h_2}(Z_t - z_0)/n$ , where  $w_{2t} = I(\hat{\mathbf{g}}_\tau^T(z_0) \hat{\mathbf{X}}_t - \delta_{2n} < Y_t \leq \hat{\mathbf{g}}_\tau^T(z_0) \hat{\mathbf{X}}_t + \delta_{2n})/(2\delta_{2n})$  for any  $\delta_{2n} \rightarrow 0$ . In the Appendix (see Section A.3 in Appendix A), it shows that indeed, 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  $\hat{\boldsymbol{\alpha}}_\tau(\cdot)$  obtained from the first step of our estimation procedures, which is guaranteed by Lemma A.2 in the Appendix. Therefore, it will show in the Appendix (see Section A.3 in Appendix A) that indeed,  $\hat{\Sigma}_\tau(z_0) = (\hat{\Omega}^*(z_0))^{-1} \hat{\Xi}(z_0) (\hat{\Omega}^*(z_0))^{-1}$  is a consistent estimate of  $\Sigma_\tau(z_0)$ , where  $\hat{\Xi}(z_0) = \tau(1 - \tau)\nu_0[\hat{\Omega}(z_0) - \hat{H}_1(z_0) + \hat{H}_2(z_0)]$  is the consistent estimate of  $\Xi(z_0)$  with  $\hat{\Omega}(z_0)$ ,  $\hat{H}_1(z_0)$  and  $\hat{H}_2(z_0)$  given above.

### 3 A Monte Carlo Simulation Study

In this section, we provide a simulation example to exam the performance of our two-stage estimations for functional coefficients. In this example, the bandwidth is selected based on a rule-of-thumb idea similar to the procedure in Cai and Xiao (2012) as follows. First, we use a data-driven bandwidth selector as suggested in Cai and Xu (2008) to obtain an initial bandwidth denoted by  $\hat{h}_0$  which should be  $O(n^{-1/5})$ . At step one, the bandwidth should be under-smoothed. Therefore, by following the idea in Cai (2002) and Cai and Xiao (2012) for two-step approaches,

we take the bandwidth as  $\hat{h}_1 = A_0 \times \hat{h}_0$  with  $A_0 = n^{-1/10}$  so that  $\hat{h}_1$  satisfies Assumption A10. At step two, we choose optimal bandwidth  $\hat{h}_2$  by the nonparametric AIC criterion as in Cai and Xu (2008). Finally, the Epanechnikov kernel  $K(x) = 0.75(1 - x^2)I(|x| \leq 1)$  is used and  $m = O(n^{1/7})$ .

In this example, for  $1 \leq i \leq 4$ , the data are generated from the following process:

$$Y_{it} = \sigma_{it}\varepsilon_{it}$$

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  $U_{it} \sim$  i.i.d. Uniform  $[0, 1]$  for  $1 \leq i \leq 4$ . In addition,  $\gamma_{ij,\epsilon_{it}}(z)$  and  $\beta_j(z)$  for  $1 \leq j \leq 4$  and  $1 \leq i \leq 4$  are defined as follows. For  $i = 1$ ,  $\gamma_{i1,\epsilon_{it}}(z) = 0.1 \{1 + \exp(-4z)\}^{-1} \epsilon_{it}$ ,  $\gamma_{i2,\chi_{it}}(z) = (0.1 \sin(-0.5\pi z) + 0.1)\chi_{it}$ ,  $\gamma_{i3,\epsilon_{it}}(z) = (0.04z^2)\epsilon_{it}$ ,  $\gamma_{i4,\chi_{it}}(z) = (-0.04z^2 + 0.15)\chi_{it}$ ,  $\beta_{i1}(z) = 0.1 \sin(0.5\pi z) + 0.1$ ,  $\beta_{i2}(z) = 0.1 \sin^2(z)$ ,  $\beta_{i3}(z) = 0.02 \exp(-z)$ , and  $\beta_{i4}(z) = 0.1 \cos^2(z)$ . For  $i = 2$ ,  $\gamma_{i1,\epsilon_{it}}(z) = (0.1 \sin(-0.5\pi z) + 0.1)\epsilon_{it}$ ,  $\gamma_{i2,\chi_{it}}(z) = 0.1 \{1 + \exp(-4z)\}^{-1} \chi_{it}$ ,  $\gamma_{i3,\epsilon_{it}}(z) = (-0.04z^2 + 0.15)\epsilon_{it}$ ,  $\gamma_{i4,\chi_{it}}(z) = (0.04z^2)\chi_{it}$ ,  $\beta_{i1}(z) = 0.1 \sin^2(z)$ ,  $\beta_{i2}(z) = 0.1 \sin(0.5\pi z) + 0.1$ ,  $\beta_{i3}(z) = 0.1 \cos^2(z)$ , and  $\beta_{i4}(z) = 0.02 \exp(-z)$ . For  $i = 3$ ,  $\gamma_{i1,\epsilon_{it}}(z) = 0.1 \{1 + 2 \exp(-2z)\}^{-1} \epsilon_{it}$ ,  $\gamma_{i2,\chi_{it}}(z) = (0.1 \sin(-0.6\pi z) + 0.1)\chi_{it}$ ,  $\gamma_{i3,\epsilon_{it}}(z) = (0.04z^2)\epsilon_{it}$ ,  $\gamma_{i4,\chi_{it}}(z) = (-0.04z^2 + 0.15)\chi_{it}$ ,  $\beta_{i1}(z) = 0.1 \sin(0.6\pi z) + 0.1$ ,  $\beta_{i2}(z) = 0.1 \sin^2(z)$ ,  $\beta_{i3}(z) = 0.02 \exp(-z)$ , and  $\beta_{i4}(z) = 0.1 \cos^2(z)$ . For  $i = 4$ ,  $\gamma_{i1,\epsilon_{it}}(z) = (0.1 \sin(-0.6\pi z) + 0.1)\epsilon_{it}$ ,  $\gamma_{i2,\chi_{it}}(z) = 0.1 \{1 + 2 \exp(-2z)\}^{-1} \chi_{it}$ ,  $\gamma_{i3,\epsilon_{it}}(z) = (-0.04z^2 + 0.15)\epsilon_{it}$ ,  $\gamma_{i4,\chi_{it}}(z) = (0.04z^2)\chi_{it}$ ,  $\beta_{i1}(z) = 0.1 \sin^2(z)$ ,  $\beta_{i2}(z) = 0.1 \sin(0.6\pi z) + 0.1$ ,  $\beta_{i3}(z) = 0.1 \cos^2(z)$ , and  $\beta_{i4}(z) = 0.02 \exp(-z)$ . Finally,  $\varepsilon_{it}$  are mutually i.i.d. from  $\mathcal{N}(0, 1)$ . Thus, for  $1 \leq i \leq 4$ , our data generating process is given by

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_t) + \sum_{j=1}^4 \gamma_{ij,\tau}(Z_t)q_{\tau,t-1,i} + \sum_{j=1}^4 \beta_{ij,\tau}(Z_t)|Y_{i(t-1)}|,$$

where  $Z_t$  is generated from Uniform  $[-2, 2]$  independently. Notice that our data generating process corresponds to the model in (1) or (2) with  $\kappa = 4$ ,  $\mathbb{Y}_t = (|Y_{1t}|, |Y_{2t}|, |Y_{3t}|, |Y_{4t}|)^T$ ,  $q = p = 1$

and  $Z_{it} = Z_t$ . 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  $1 \leq i, j \leq 4$ , 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  $\tau$ , suggesting different covariate effects at different levels of  $\tau$ .

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

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

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  $\{z_k = 0.1(k-1) - 1.75 : 1 \leq k \leq n_0 = 36\}$  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  $\tau$ '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  $\tau$ . Moreover, the performances for  $\gamma_{i0,\tau}(\cdot)$  and  $\beta_{ij,\tau}(\cdot)$  at  $\tau = 0.15$  are slightly better than those for  $\tau = 0.05$  and  $0.95$ . This observation is because of the sparsity of data in the tailed regions, which is similar to that in Cai and Xu (2008). Nevertheless, since the data that are used to estimate  $\gamma_{ij,\tau}(\cdot)$  at  $\tau = 0.05$  and  $0.95$  are conditional quantiles, the distributional information at tailed regions is preserved, which may reduce the problem of data sparsity. For this reason, the performances for  $\gamma_{ij,\tau}(\cdot)$  at  $\tau = 0.15$  are not necessarily superior to that for  $\tau = 0.05$  and  $0.95$ .

Finally, we illustrate the finite sample performance for the consistent covariance estimation 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  $1 \leq i, j \leq 4$ , defined as,

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

where  $se(\hat{\gamma}_{ij,\tau}(\cdot)) = \left[ \widehat{Var}(\hat{\gamma}_{ij,\tau}(\cdot)) / nh_2 \right]^{1/2}$ ,  $I_b\{\gamma_{ij,\tau}(\cdot) \in \hat{\gamma}_{ij,\tau}(\cdot) \pm 1.96 \times se(\hat{\gamma}_{ij,\tau}(\cdot))\}$  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  $b$ th time of replication (equals to 0, otherwise), and the number of replication times  $B$  is 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 AECRs of 95% pointwise CIs are close to the nominal level 0.95 for all settings. In general, the results of this simulated experiment demonstrate that the proposed procedure is reliable and works fairly well.

## 4 A Real 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 network in international equity markets. Different from the existing literatures, the interdependences of this class of network vary with a smoothing variable of general economy. To capture the inter-temporal transition of risk and avoid endogeneity, we consider the interaction between current and one-day lagged VaR. In particular, we define each linkage between a pair of VaRs in our network as the sensitivity of VaR of returns of one market index at time  $t$  to that of another at time  $t - 1$ . Therefore, our network can be

written as following equation system:

$$\text{VaR}_{it} = \boldsymbol{\gamma}_{i,\tau}^T(Z_{t-1})\text{VaR}_{t-1}, \quad i = 1, 2, \dots, \kappa, \quad (11)$$

where  $\text{VaR}_{t-1} = (\text{VaR}_{1(t-1)}, \dots, \text{VaR}_{\kappa(t-1)})^T$  is a vector of VaRs for all market index returns at time  $t - 1$  and  $\text{VaR}_{it}$  is the VaR of the  $i$ th market index return at time  $t$ , which is described as follows

$$\text{VaR}_{it} = -\inf\{Y \in \mathbb{R} : P(Y_{it} > Y | \mathcal{F}_{i,t-1}) \leq 1 - \tau\} = -\inf\{Y \in \mathbb{R} : F(Y | \mathcal{F}_{i,t-1}) > \tau\}$$

for  $i = 1, 2, \dots, \kappa$  at a given  $\tau \in (0, 1)$ . Here,  $\mathcal{F}_{i,t-1}$  is the information set to present all information of the  $i$ th return available at time  $t - 1$  and  $F(\cdot | \mathcal{F}_{i,t-1})$  represents the conditional distribution function of  $Y_{it}$  given  $\mathcal{F}_{i,t-1}$ . In addition,  $Z_{t-1}$  is a smoothing variable of general economy and  $\boldsymbol{\gamma}_{i,\tau}(\cdot) = (\gamma_{i1,\tau}(\cdot), \dots, \gamma_{i\kappa,\tau}(\cdot))^T$  is a  $\kappa \times 1$  vector of functional coefficients. Then, we extract the quantile estimation of functional coefficients from equation system (11) and construct the matrix  $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$  as our financial network as follows:

$$|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})| = (|\hat{\gamma}_{ij,\tau}(Z_{t-1})|)_{\kappa \times \kappa}$$

in which,  $|\hat{\gamma}_{ij,\tau}(Z_{t-1})|$  represents the absolute value of the sensitivity of VaR of return for the market index  $j$  at time  $t$  to that of return for the index  $i$  at time  $t - 1$ , under  $\tau$ -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 will be reported in Section 4.3 later. Thus, matrix  $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$  is useful to capture risk interdependence and how it changes with a smoothing variable  $Z_{t-1}$ . Notice that entries  $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$  correspond to the absolute value of the estimated values of  $\{\gamma_{ij,\tau}(\cdot)\}$  in the network model in (3). Therefore, our two-stage procedures can be applied here for direct estimation of the interdependence among VaRs of returns for the market indexes. In general, the proposed framework is particularly suitable to investigate the dynamic characteristics of risk spillover across global market indexes under the changes of economic circumstance.

## 4.2 Data

Our dataset includes the daily series between January 5, 2006 and February 10, 2021 for four major world equity market indexes: 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. We model the  $i$ th index's return series  $Y_{it} = 10 \log(\pi_{it}/\pi_{i(t-1)})$ , where  $i = 1, 2, 3, 4$  correspond to the four aforementioned market indexes in turn and  $\pi_{it}$  is  $i$ th index level on the  $t$ th day. The studies of global market indexes help to explore the dynamic of risk dependences in the global financial market, and 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 indexes are downloaded in Yahoo Finance and the estimation sample sizes  $n = 3254$ . Thus, we take  $m = n^{1/7} \approx 3$  in this empirical study. Although it is feasible to introduce more kinds of market index into the equation system (11), due to the computational burdens, we only consider risk co-dependences among four major markets' indexes.

As for the smoothing variable  $Z_t$ , we choose  $Z_t = 10 \log(D_t/D_{t-1})$ , where  $D_t$  is the U.S. dollar index on the  $t$ th 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 (11) 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 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 (11) 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  $1 \leq i \leq 4$  and  $1 \leq j \leq 4$  in (11) 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 these coefficient functions vary significantly over the interval  $[-0.075, 0.075]$ , 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  $\tau$  are dropped from  $|\hat{\gamma}_{ij,\tau}(Z_{t-1})|$  and  $|\hat{\gamma}_{ji,\tau}(Z_{t-1})|$  in the matrix  $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$ , in what follows. Then,  $|\hat{\gamma}_{ji}|$  in the matrix  $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$  represents the intensity of influence from the risk of market index  $i$  at time  $t - 1$  to 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 is the strength of the influence of all indexes' VaR at time  $t - 1$  to the VaR of market index  $j$  at time  $t$ . Analogously, the connectedness with respect to outgoing links (COL) is defined as  $\sum_{i=1}^4 |\hat{\gamma}_{ij}|$ , which is the strength of the influence of index  $j$ 's VaR at time  $t - 1$  to the VaRs of all indexes at time  $t$ . Here,  $i, j = 1, 2, 3, 4$  correspond to the four aforementioned market

indexes in turn. The CIL measures the risk spillover that was emitted from all four market indexes one day ago and is received currently by a certain market index; the COL measures the risk spillover emitted from a certain market index one day ago and is received currently by all market indexes. Intuitively, the CIL measures exposures of individual indexes to systemic shocks from the financial network, while the COL measures contributions of individual indexes 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 axes 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 indexes vary greatly over the interval  $(-0.075, 0.075)$  and exhibit almost asymmetrically U-shaped. In particular, when the U.S. dollar experiences appreciation and during the “bad times” of the market (when  $Z_{t-1}$  is large 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 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 influential to the global market and to be influenced by global market more easily, respectively. Thus, both curves in the panel of S&P 500 index increase. As for the case when U.S. dollar depreciated, profits of investment on domestic corporations in European, Japanese



and Chinese markets may increase, which lead the total amount of investment in these three markets to grow. As a result, 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.

It is interesting that in the European and Japanese markets, during the U.S. dollar appreciation ( $Z_{t-1}$  is large), 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  $Z_{t-1}$  is large, 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. This increase of excess returns may attract more carry traders to borrow the Japanese or European assets and thus, make these two markets more influential to the global market. For this reason, the COL becomes larger than CIL 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.

On the other hand, during the U.S. dollar depreciation, carry traders who borrowed the Japanese or European asset may be unable to repay due to the significant loss of returns, which cause the Japanese or European market to become more vulnerable. Consequently, the CIL in both Japan and Europe markets increases drastically relative to the COL. Yet, in the U.S. market, the price of risky assets affect the solvency of carry traders in the world, which let the U.S. market become more influential to the world. Thus, the COL rises compared to the CIL for S&P 500 index. As for the Chinese market, when U.S. dollar depreciated, 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 magnifies the influence of global risk events on the Chinese market, causing the CIL to dominate the COL.

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 indexes demonstrates an U-shaped and asymmetric pattern. It means that total risk spillover in the four major markets decrease when  $Z_{t-1}$  becomes larger within the interval  $[-0.075, -0.025]$ . As  $Z_{t-1}$  exceeds  $-0.025$ , the risk spillover intensity is magnified. In general, Figure 3 shows that the response of total risk spillover to 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.

## 5 Conclusion

In this paper, we investigate a functional coefficient VAR model for conditional quantiles, which is new to the literature. A two-stage kernel method 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 methods of estimation 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 two-stage estimation approaches, the proposed framework allows to study how specific state of economy has an influence on the network characteristics of risk spillover in a financial system.

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 indexes 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.

## References

- Acemoglu, D., Ozdaglar, A. and Tahbaz-Salehi, A. (2015). Systemic Risk and Stability in Financial Networks. *American Economic Review*, 105(2), 564-608.
- Adrian, T. and Brunnermeier, M.K. (2016). CoVaR. *American Economic Review*, 106(7), 1705-1741.
- An, H.Z. and Chen, S.G. (1997). A Note on the Ergodicity of Nonlinear Autoregressive Models. *Statistics and Probability Letters*, 34(4), 365-372.
- An, H.Z. and Huang, F.C. (1996). The Geometrical Ergodicity of Nonlinear Autoregressive Models. *Statistica Sinica*, 6(4), 943-956.
- 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.
- 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.
- Bräuning, F. and Koopman, S.J. (2020). The Dynamic Factor Network Model with an Application to International Trade. *Journal of Econometrics*, 216(2), 494-515.
- Cai, Z. (2002). Two-step Likelihood Estimation Procedure for Varying-Coefficient Models. *Journal of Multivariate Analysis*, 82(1), 189-209.
- 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., T. Juhl and B. Yang (2015). Functional Index Coefficient Models with Variable Selection. *Journal of Econometrics*, 189(1), 272-284.
- 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 Xiao, Z. (2012). Semiparametric Quantile Regression Estimation in Dynamic Models with Partially Varying Coefficients. *Journal of Econometrics*, 167(2), 413-425.
- 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.
- Chen, R. and Tsay, R.S. (1993). Functional Coefficient Autoregressive Model. *Journal of the American Statistical Association*, 88(421), 298-308.
- Dahlhaus, R. and S. Subba Rao. (2006). Statistical Inference for Time-varying ARCH Process. *Annals of Statistics*, 34(3), 1075-1114.
- Diebold, F.X. and Yılmaz, 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.
- 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.
- Gourieroux, C. and Jasiak, J. (2008). Dynamic Quantile Models. *Journal of Econometrics*, 147(1), 198-205.
- 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 Ng, P. (1999). Quantile Splines with Several Covariates. *Journal of Statistical Planning and Inference*, 75(2), 343-352.
- He, X. and Shao, Q. (2000). On Parameters of Increasing Dimensions. *Journal of Multivariate Analysis*, 73(1), 120-135.
- Honda, T. (2000). Nonparametric Estimation of a Conditional Quantile for  $\alpha$ -mixing Processes. *Annals of the Institute of Statistical Mathematics*, 52(3), 459-470.
- Honda, T. (2004). Quantile Regression in Varying Coefficient Models. *Journal of Statistical Planning and Inferences*, 121(1), 113-125.
- Kim, M. (2007). Quantile Regression With Varying Coefficients. *Annals of Statistics*, 35(1), 92-108.
- 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.
- Lee, J., Li, G. and Wilson, J.D. (2020). Varying-Coefficient Models for Dynamic Networks. *Computational Statistics and Data Analysis*, 152, 197952.
- Li, D., Li, Q. and Li, Z. (2021). Nonparametric Quantile Regression Estimation With Mixed Discrete and Continuous Data. *Journal of Business & Economic Statistics*, 39(3), 741-756.
- 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.
- Pham, D.T. (1986). The Mixing Property of Bilinear and Generalized Random Coefficient Autoregressive Models. *Stochastic Processes and Their Applications*, 23(2), 291-300.
- Qu, Z. and Yoon, J. (2015). Nonparametric Estimation and Inference on Conditional Quantile Processes. *Journal of Econometrics*, 185(1), 1-19.
- Sewell, D.K., and Chen, Y. (2015). Latent Space Models for Dynamic Networks. *Journal of the American Statistical Association*, 110(512), 1646-1657.

- 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.
- Wei, Y. and He, X. (2006). Conditional Growth Charts (with discussion). *Annals of Statistics*, 34(4), 2069-2097.
- Welsh, A. (1989). On M-Processes and M-Estimation. *Annals of Statistics*, 17(1), 337-361.
- 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.
- Yang, Z. and Zhou, Y. (2017). Quantitative Easing and Volatility Spillovers Across Countries and Asset Classes. *Management Science*, 63(2), 333-354.
- Zhu, X., Wang, W., Wang, H. and Härdle, W.K. (2019). Network Quantile Autoregression. *Journal of Econometrics*, 212(1), 345-358.

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  $1 \leq j \leq 4$ .

$\tau$	$n = 500$		$n = 1500$		$n = 4000$	
	MADE( $\gamma_{10}$ )	MADE( $\gamma_{20}$ )	MADE( $\gamma_{10}$ )	MADE( $\gamma_{20}$ )	MADE( $\gamma_{10}$ )	MADE( $\gamma_{20}$ )
0.05	0.649 (0.110)	0.679 (0.108)	0.548 (0.050)	0.548 (0.050)	0.424 (0.036)	0.384 (0.035)
0.15	0.376 (0.055)	0.376 (0.055)	0.338 (0.031)	0.290 (0.035)	0.291 (0.022)	0.225 (0.024)
0.95	0.732 (0.188)	0.638 (0.126)	0.518 (0.061)	0.580 (0.068)	0.432 (0.038)	0.412 (0.036)
	MADE( $\gamma_{30}$ )	MADE( $\gamma_{40}$ )	MADE( $\gamma_{30}$ )	MADE( $\gamma_{40}$ )	MADE( $\gamma_{30}$ )	MADE( $\gamma_{40}$ )
0.05	0.627 (0.102)	0.700 (0.126)	0.563 (0.050)	0.569 (0.048)	0.488 (0.033)	0.458 (0.031)
0.15	0.403 (0.053)	0.409 (0.049)	0.307 (0.033)	0.305 (0.032)	0.245 (0.024)	0.243 (0.023)
0.95	0.754 (0.186)	0.691 (0.157)	0.522 (0.064)	0.579 (0.071)	0.464 (0.037)	0.369 (0.037)
	MADE( $\gamma_{11}$ )	MADE( $\gamma_{12}$ )	MADE( $\gamma_{11}$ )	MADE( $\gamma_{12}$ )	MADE( $\gamma_{11}$ )	MADE( $\gamma_{12}$ )
0.05	0.148 (0.063)	0.139 (0.063)	0.111 (0.045)	0.126 (0.056)	0.093 (0.036)	0.100 (0.042)
0.15	0.116 (0.051)	0.141 (0.063)	0.081 (0.036)	0.104 (0.048)	0.069 (0.032)	0.085 (0.034)
0.95	0.182 (0.085)	0.201 (0.103)	0.141 (0.055)	0.153 (0.061)	0.108 (0.040)	0.122 (0.047)
	MADE( $\gamma_{13}$ )	MADE( $\gamma_{14}$ )	MADE( $\gamma_{13}$ )	MADE( $\gamma_{14}$ )	MADE( $\gamma_{13}$ )	MADE( $\gamma_{14}$ )
0.05	0.147 (0.063)	0.147 (0.068)	0.115 (0.050)	0.124 (0.054)	0.095 (0.035)	0.094 (0.040)
0.15	0.105 (0.051)	0.153 (0.065)	0.082 (0.036)	0.113 (0.047)	0.069 (0.026)	0.078 (0.039)
0.95	0.176 (0.081)	0.212 (0.092)	0.132 (0.054)	0.153 (0.060)	0.108 (0.036)	0.120 (0.045)
	MADE( $\gamma_{21}$ )	MADE( $\gamma_{22}$ )	MADE( $\gamma_{21}$ )	MADE( $\gamma_{22}$ )	MADE( $\gamma_{21}$ )	MADE( $\gamma_{22}$ )
0.05	0.164 (0.077)	0.120 (0.060)	0.119 (0.047)	0.111 (0.049)	0.097 (0.039)	0.087 (0.038)
0.15	0.134 (0.054)	0.125 (0.057)	0.098 (0.037)	0.101 (0.043)	0.086 (0.034)	0.092 (0.034)
0.95	0.194 (0.073)	0.183 (0.076)	0.154 (0.058)	0.140 (0.056)	0.115 (0.040)	0.108 (0.038)
	MADE( $\gamma_{23}$ )	MADE( $\gamma_{24}$ )	MADE( $\gamma_{23}$ )	MADE( $\gamma_{24}$ )	MADE( $\gamma_{23}$ )	MADE( $\gamma_{24}$ )
0.05	0.156 (0.071)	0.133 (0.066)	0.124 (0.052)	0.111 (0.047)	0.099 (0.038)	0.087 (0.035)
0.15	0.120 (0.054)	0.127 (0.053)	0.094 (0.040)	0.098 (0.041)	0.079 (0.031)	0.095 (0.035)
0.95	0.186 (0.073)	0.184 (0.075)	0.143 (0.065)	0.135 (0.054)	0.113 (0.036)	0.104 (0.038)

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

$\tau$	$n = 500$		$n = 1500$		$n = 4000$	
	MADE( $\gamma_{31}$ )	MADE( $\gamma_{32}$ )	MADE( $\gamma_{31}$ )	MADE( $\gamma_{32}$ )	MADE( $\gamma_{31}$ )	MADE( $\gamma_{32}$ )
0.05	0.146 (0.064)	0.143 (0.063)	0.107 (0.043)	0.124 (0.052)	0.099 (0.041)	0.087 (0.036)
0.15	0.115 (0.059)	0.140 (0.065)	0.082 (0.035)	0.105 (0.048)	0.069 (0.028)	0.093 (0.034)
0.95	0.178 (0.085)	0.200 (0.093)	0.135 (0.053)	0.149 (0.061)	0.108 (0.040)	0.113 (0.049)
	MADE( $\gamma_{33}$ )	MADE( $\gamma_{34}$ )	MADE( $\gamma_{33}$ )	MADE( $\gamma_{34}$ )	MADE( $\gamma_{33}$ )	MADE( $\gamma_{34}$ )
0.05	0.153 (0.062)	0.147 (0.062)	0.116 (0.051)	0.121 (0.054)	0.099 (0.037)	0.093 (0.037)
0.15	0.100 (0.047)	0.136 (0.062)	0.085 (0.036)	0.113 (0.045)	0.073 (0.028)	0.087 (0.036)
0.95	0.180 (0.084)	0.212 (0.097)	0.136 (0.049)	0.153 (0.057)	0.104 (0.041)	0.131 (0.043)
	MADE( $\gamma_{41}$ )	MADE( $\gamma_{42}$ )	MADE( $\gamma_{41}$ )	MADE( $\gamma_{42}$ )	MADE( $\gamma_{41}$ )	MADE( $\gamma_{42}$ )
0.05	0.156 (0.079)	0.123 (0.065)	0.118 (0.050)	0.116 (0.050)	0.099 (0.039)	0.091 (0.036)
0.15	0.129 (0.063)	0.115 (0.061)	0.099 (0.040)	0.098 (0.041)	0.079 (0.030)	0.081 (0.031)
0.95	0.195 (0.085)	0.180 (0.086)	0.148 (0.059)	0.141 (0.056)	0.113 (0.043)	0.105 (0.036)
	MADE( $\gamma_{43}$ )	MADE( $\gamma_{44}$ )	MADE( $\gamma_{43}$ )	MADE( $\gamma_{44}$ )	MADE( $\gamma_{43}$ )	MADE( $\gamma_{44}$ )
0.05	0.147 (0.080)	0.139 (0.066)	0.118 (0.056)	0.111 (0.047)	0.093 (0.039)	0.085 (0.034)
0.15	0.109 (0.055)	0.118 (0.054)	0.092 (0.037)	0.099 (0.039)	0.072 (0.027)	0.086 (0.030)
0.95	0.188 (0.088)	0.184 (0.076)	0.146 (0.060)	0.137 (0.052)	0.105 (0.041)	0.108 (0.037)

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

$\tau$	$n = 500$		$n = 1500$		$n = 4000$	
	MADE( $\beta_{11}$ )	MADE( $\beta_{12}$ )	MADE( $\beta_{11}$ )	MADE( $\beta_{12}$ )	MADE( $\beta_{11}$ )	MADE( $\beta_{12}$ )
0.05	0.215 (0.098)	0.214 (0.098)	0.131 (0.052)	0.137 (0.054)	0.087 (0.033)	0.088 (0.035)
0.15	0.137 (0.058)	0.145 (0.060)	0.084 (0.036)	0.089 (0.043)	0.056 (0.024)	0.060 (0.025)
0.95	0.253 (0.115)	0.286 (0.125)	0.151 (0.053)	0.159 (0.060)	0.095 (0.036)	0.103 (0.036)
	MADE( $\beta_{13}$ )	MADE( $\beta_{14}$ )	MADE( $\beta_{13}$ )	MADE( $\beta_{14}$ )	MADE( $\beta_{13}$ )	MADE( $\beta_{14}$ )
0.05	0.210 (0.092)	0.210 (0.097)	0.124 (0.052)	0.130 (0.054)	0.082 (0.031)	0.083 (0.033)
0.15	0.136 (0.062)	0.143 (0.062)	0.079 (0.034)	0.083 (0.039)	0.051 (0.020)	0.058 (0.023)
0.95	0.246 (0.114)	0.255 (0.119)	0.149 (0.055)	0.151 (0.057)	0.084 (0.034)	0.094 (0.029)
	MADE( $\beta_{21}$ )	MADE( $\beta_{22}$ )	MADE( $\beta_{21}$ )	MADE( $\beta_{22}$ )	MADE( $\beta_{21}$ )	MADE( $\beta_{22}$ )
0.05	0.213 (0.104)	0.218 (0.104)	0.135 (0.058)	0.133 (0.052)	0.084 (0.030)	0.084 (0.034)
0.15	0.132 (0.059)	0.150 (0.062)	0.088 (0.036)	0.090 (0.036)	0.061 (0.026)	0.064 (0.023)
0.95	0.249 (0.099)	0.260 (0.105)	0.150 (0.060)	0.160 (0.063)	0.091 (0.031)	0.100 (0.036)
	MADE( $\beta_{23}$ )	MADE( $\beta_{24}$ )	MADE( $\beta_{23}$ )	MADE( $\beta_{24}$ )	MADE( $\beta_{23}$ )	MADE( $\beta_{24}$ )
0.05	0.219 (0.102)	0.204 (0.104)	0.122 (0.050)	0.123 (0.052)	0.086 (0.031)	0.080 (0.031)
0.15	0.132 (0.058)	0.140 (0.061)	0.084 (0.034)	0.087 (0.034)	0.058 (0.021)	0.059 (0.022)
0.95	0.237 (0.096)	0.251 (0.107)	0.150 (0.061)	0.153 (0.065)	0.095 (0.032)	0.090 (0.029)



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

$\tau$	$n = 500$		$n = 1500$		$n = 4000$	
	MADE( $\beta_{31}$ )	MADE( $\beta_{32}$ )	MADE( $\beta_{31}$ )	MADE( $\beta_{32}$ )	MADE( $\beta_{31}$ )	MADE( $\beta_{32}$ )
0.05	0.218 (0.086)	0.219 (0.099)	0.131 (0.054)	0.132 (0.054)	0.089 (0.035)	0.091 (0.035)
0.15	0.138 (0.065)	0.144 (0.067)	0.087 (0.037)	0.091 (0.037)	0.058 (0.022)	0.061 (0.024)
0.95	0.262 (0.119)	0.260 (0.137)	0.151 (0.058)	0.161 (0.058)	0.095 (0.037)	0.106 (0.041)
	MADE( $\beta_{33}$ )	MADE( $\beta_{34}$ )	MADE( $\beta_{33}$ )	MADE( $\beta_{34}$ )	MADE( $\beta_{33}$ )	MADE( $\beta_{34}$ )
0.05	0.207 (0.092)	0.218 (0.094)	0.121 (0.052)	0.130 (0.052)	0.076 (0.032)	0.087 (0.033)
0.15	0.130 (0.068)	0.129 (0.068)	0.082 (0.034)	0.083 (0.039)	0.057 (0.021)	0.056 (0.028)
0.95	0.247 (0.119)	0.255 (0.137)	0.147 (0.059)	0.151 (0.060)	0.089 (0.033)	0.110 (0.037)
	MADE( $\beta_{41}$ )	MADE( $\beta_{42}$ )	MADE( $\beta_{41}$ )	MADE( $\beta_{42}$ )	MADE( $\beta_{41}$ )	MADE( $\beta_{42}$ )
0.05	0.219 (0.101)	0.234 (0.108)	0.132 (0.057)	0.139 (0.052)	0.091 (0.032)	0.088 (0.034)
0.15	0.132 (0.066)	0.141 (0.069)	0.087 (0.034)	0.084 (0.034)	0.057 (0.021)	0.057 (0.023)
0.95	0.253 (0.110)	0.265 (0.119)	0.157 (0.061)	0.167 (0.066)	0.089 (0.032)	0.097 (0.035)
	MADE( $\beta_{43}$ )	MADE( $\beta_{44}$ )	MADE( $\beta_{43}$ )	MADE( $\beta_{44}$ )	MADE( $\beta_{43}$ )	MADE( $\beta_{44}$ )
0.05	0.211 (0.109)	0.207 (0.100)	0.124 (0.048)	0.123 (0.056)	0.082 (0.031)	0.082 (0.032)
0.15	0.131 (0.061)	0.125 (0.066)	0.080 (0.034)	0.083 (0.032)	0.058 (0.021)	0.056 (0.022)
0.95	0.234 (0.109)	0.238 (0.115)	0.144 (0.057)	0.152 (0.071)	0.090 (0.028)	0.088 (0.029)

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

$\tau$	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	0.959	0.934	0.948	0.941	0.925	0.936	0.933	0.938
0.15	0.945	0.943	0.953	0.921	0.955	0.954	0.957	0.954
0.95	0.925	0.913	0.929	0.912	0.909	0.938	0.935	0.943
	$\hat{\gamma}_{21,\tau}$	$\hat{\gamma}_{22,\tau}$	$\hat{\gamma}_{23,\tau}$	$\hat{\gamma}_{24,\tau}$	$\hat{\beta}_{21,\tau}$	$\hat{\beta}_{22,\tau}$	$\hat{\beta}_{23,\tau}$	$\hat{\beta}_{24,\tau}$
0.05	0.916	0.935	0.930	0.937	0.931	0.932	0.929	0.934
0.15	0.923	0.953	0.934	0.952	0.958	0.952	0.956	0.953
0.95	0.930	0.938	0.934	0.936	0.947	0.938	0.942	0.935
	$\hat{\gamma}_{31,\tau}$	$\hat{\gamma}_{32,\tau}$	$\hat{\gamma}_{33,\tau}$	$\hat{\gamma}_{34,\tau}$	$\hat{\beta}_{31,\tau}$	$\hat{\beta}_{32,\tau}$	$\hat{\beta}_{33,\tau}$	$\hat{\beta}_{34,\tau}$
0.05	0.949	0.939	0.942	0.939	0.944	0.958	0.949	0.940
0.15	0.952	0.936	0.955	0.921	0.957	0.961	0.956	0.956
0.95	0.927	0.905	0.927	0.913	0.913	0.934	0.940	0.932
	$\hat{\gamma}_{41,\tau}$	$\hat{\gamma}_{42,\tau}$	$\hat{\gamma}_{43,\tau}$	$\hat{\gamma}_{44,\tau}$	$\hat{\beta}_{41,\tau}$	$\hat{\beta}_{42,\tau}$	$\hat{\beta}_{43,\tau}$	$\hat{\beta}_{44,\tau}$
0.05	0.930	0.934	0.936	0.926	0.923	0.921	0.939	0.929
0.15	0.923	0.954	0.932	0.943	0.951	0.955	0.956	0.957
0.95	0.936	0.947	0.929	0.945	0.944	0.941	0.949	0.942

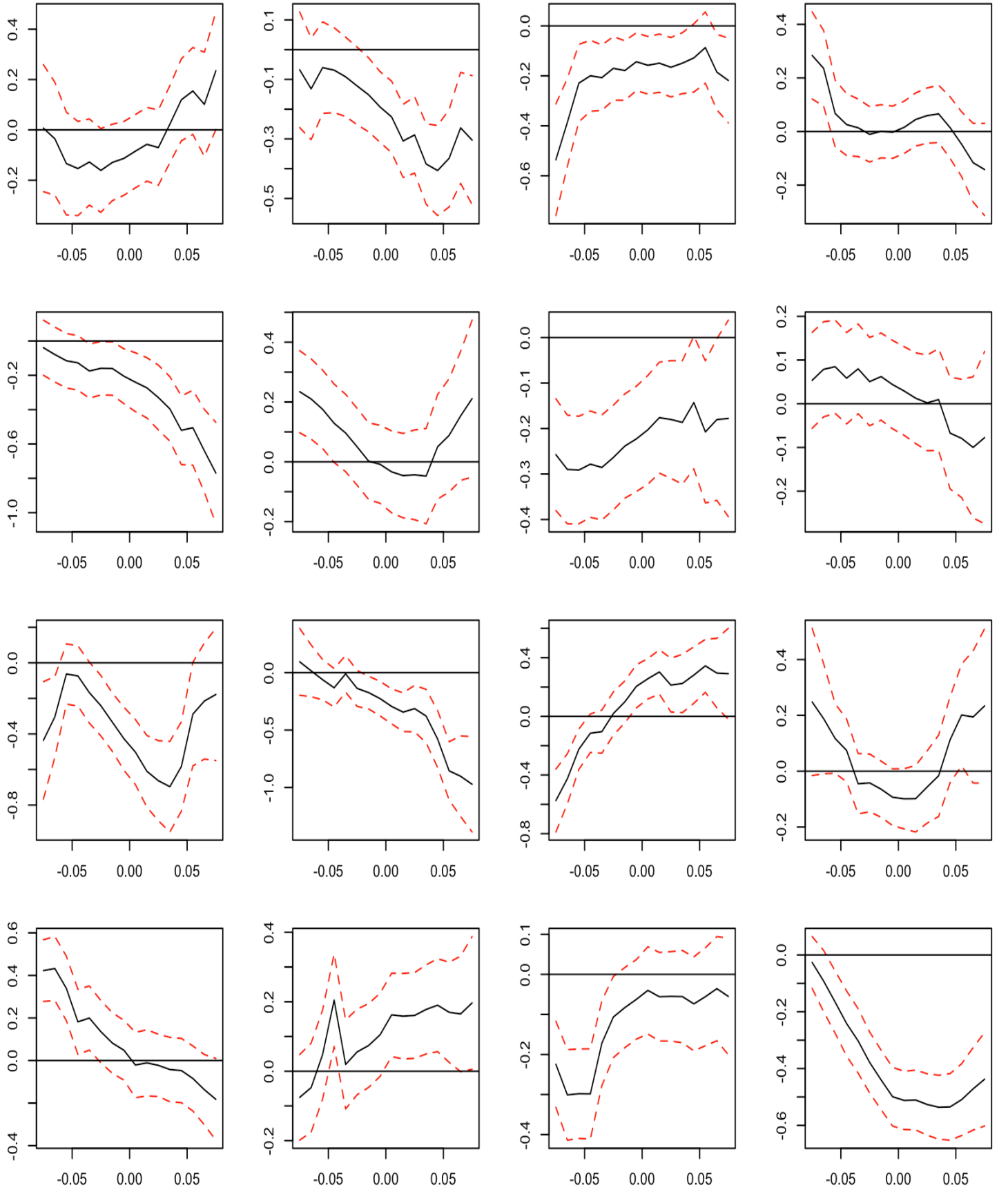


Figure 1: Plots of the estimated coefficient functions  $\gamma_{ij,\tau}(\cdot)$  for  $1 \leq i \leq 4$  and  $1 \leq j \leq 4$  in (11) in the main article under  $\tau = 0.05$  (black solid lines), in which  $ij$ -th panel represents the result for  $\gamma_{ij,\tau}(\cdot)$ , respectively. The red dashed lines in each panel indicate the 95% pointwise confidence interval for the estimate with the asymptotic bias ignored.

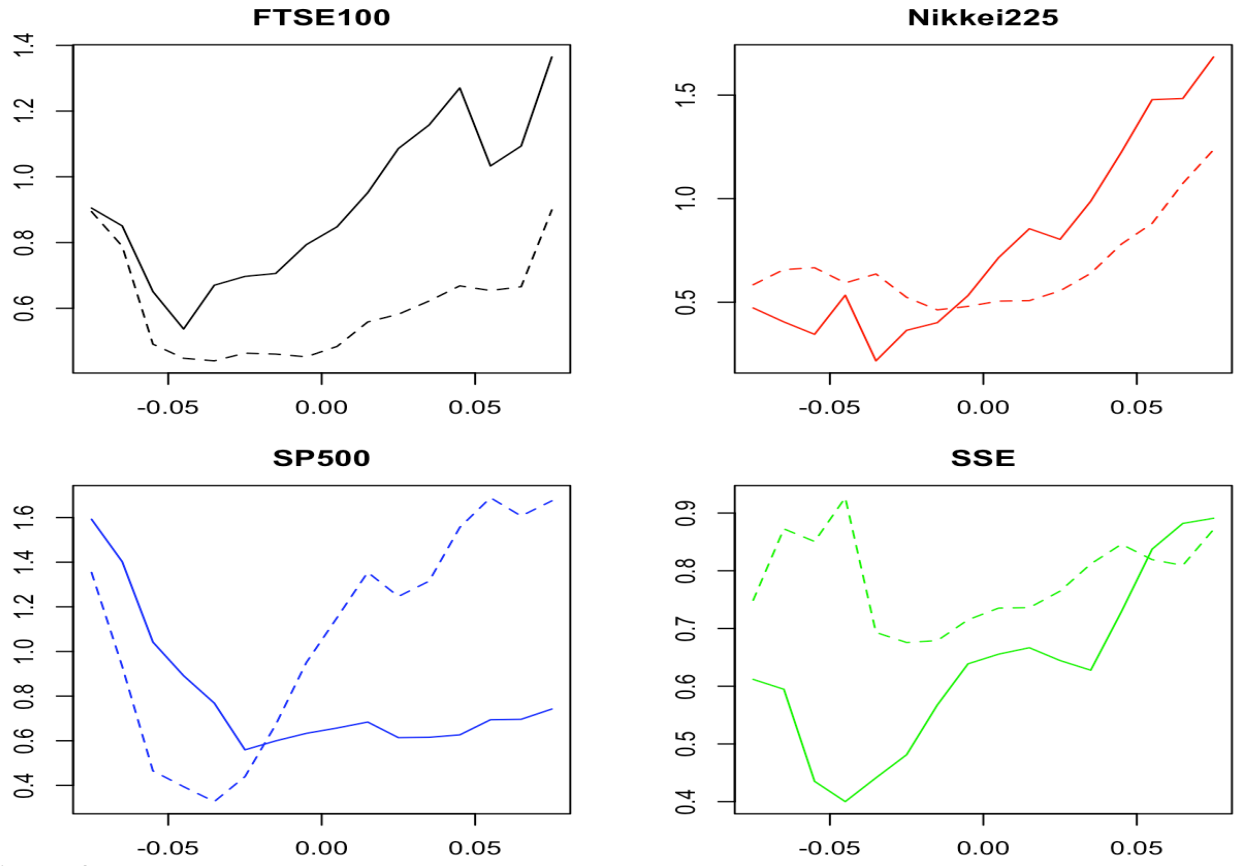


Figure 2: Connectedness with respect to outgoing links and connectedness with respect to incoming links for four market indexes with  $\tau = 0.05$ . The solid line in each panel represents values of connectedness with respect to outgoing links and the dashed line in each panel indicates values of connectedness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

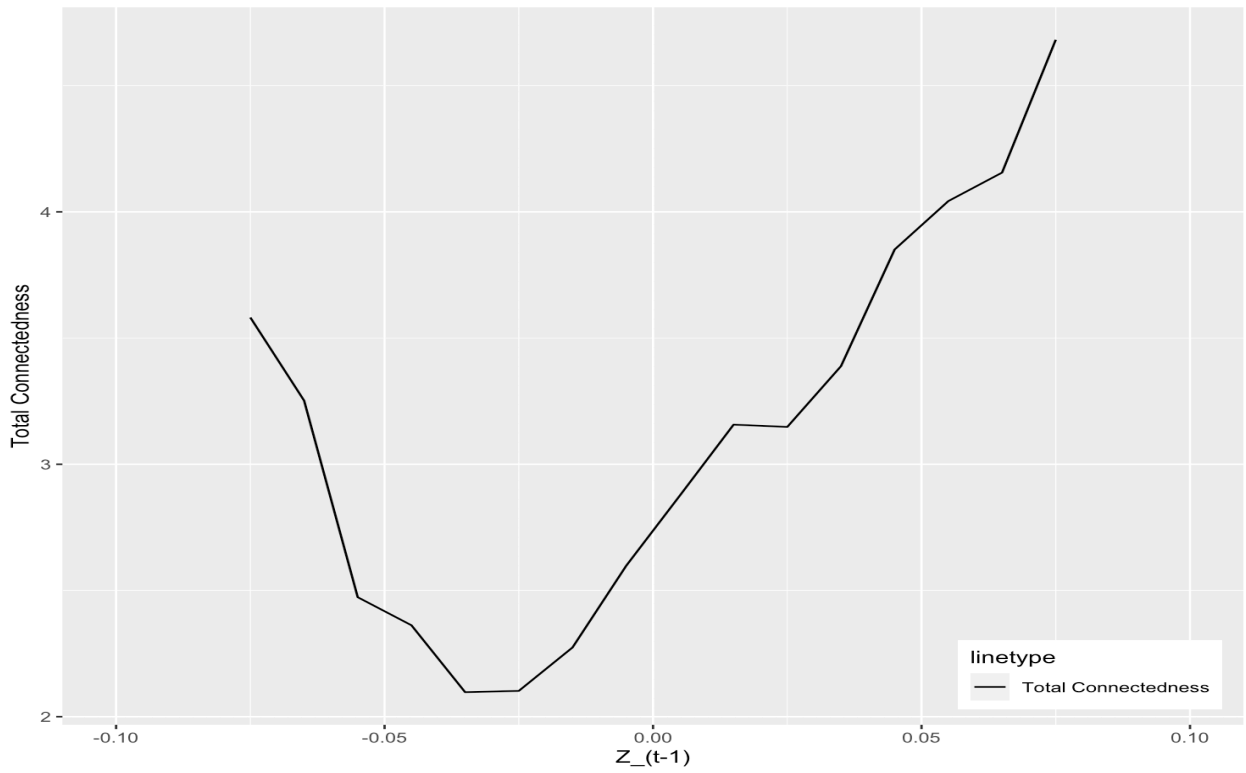


Figure 3: Total connectedness in international equity markets with  $\tau = 0.05$ .

# Appendix to “A Nonparametric Dynamic Network via Multivariate Quantile Autoregressions”

# Appendix: Mathematical Proofs

## Appendix A: Mathematical Proofs of Theorem 1 and Consistency of $\hat{\Sigma}_\tau(z_o)$

In this section, we give certain lemmas with their detailed proofs that are useful for proving the theorem 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.

### A.1 Some Lemmas

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

*Proof.* The proof follows from Ruppert and Carroll (1980).  $\square$

Now, some notations are introduced here to make a convenient presentation of our Bahadur results given in Lemma A.6 (below). In Lemmas A.2 - A.6,  $\tau$  is dropped from  $\alpha_\tau(\cdot)$  and write  $h_1$  as  $h$  for simplicity. Let  $a_n = (nh_1)^{-1/2}$ , for  $1 \leq s \neq t \leq n$  and for any fixed  $Z_t \neq Z_s$ , define  $\vartheta_0 = a_n^{-1}(\delta_0 - \alpha(Z_t))$  and  $\hat{\vartheta}_0 = a_n^{-1}(\hat{\delta}_0 - \alpha(Z_t))$ . Of course,  $\vartheta = a_n^{-1} \mathbf{H}_1 \begin{pmatrix} \delta_0 - \alpha(Z_t) \\ \delta_1 - \alpha^{(1)}(Z_t) \end{pmatrix}$ ,  $\hat{\vartheta} = a_n^{-1} \mathbf{H}_1 \begin{pmatrix} \hat{\delta}_0 - \alpha(Z_t) \\ \hat{\delta}_1 - \alpha^{(1)}(Z_t) \end{pmatrix}$ , where  $\mathbf{H}_1 = \text{diag}\{I_{\kappa m+1}, h_1 I_{\kappa m+1}\}$ . In addition, let  $\mathbf{W}_s^* = \begin{pmatrix} \mathbf{W}_s \\ z_{sh} \mathbf{W}_s \end{pmatrix}$ , where  $z_{sh} = (Z_s - Z_t)/h$ . Also, define  $Y_s^* = Y_s - \mathbf{W}_s^T [\alpha(Z_t) + \alpha^{(1)}(Z_t)(Z_s - Z_t)]$ . Therefore,

$$\hat{\vartheta} = \arg \min_{\vartheta} \sum_{s=m+1 \neq t}^n \rho_\tau(Y_s^* - a_n \vartheta^T \mathbf{W}_s^*) K(z_{sh}) \equiv \arg \min_{\vartheta} G(\vartheta).$$

The derivative of  $G(\vartheta)$  with respect to  $\vartheta$  (except at point  $Y_s^* = a_n \vartheta^T \mathbf{W}_s^*$ ) is given by

$$T_n(\vartheta) = a_n \sum_{s=m+1 \neq t}^n \psi_\tau(Y_s^* - a_n \vartheta^T \mathbf{W}_s^*) \mathbf{W}_s^* K(z_{sh}), \quad (\text{A.1})$$

where  $\psi_\tau(x) = \tau - I(x < 0)$ . Write  $\boldsymbol{\zeta} \equiv a_n \boldsymbol{\vartheta}$  and  $\hat{\boldsymbol{\zeta}} \equiv a_n \hat{\boldsymbol{\vartheta}}$ . Then, (A.1) becomes to

$$T_n(\boldsymbol{\zeta}) = a_n \sum_{s=m+1 \neq t}^n \psi_\tau(Y_s^* - \boldsymbol{\zeta}^T \mathbf{W}_s^*) \mathbf{W}_s^* K(z_{sh}). \quad (\text{A.2})$$

In particular, suppose that  $\mathcal{D}$  is any compact subset of  $\mathbb{R}$ . To show the uniform consistency of  $\hat{\boldsymbol{\alpha}}(\cdot)$  in Lemma A.2 later, for any  $z \in \mathcal{D}$ , define  $\hat{\boldsymbol{\vartheta}}(z) = a_n^{-1} \mathbf{H}_1 \begin{pmatrix} \hat{\boldsymbol{\alpha}}(z) - \boldsymbol{\alpha}(z) \\ \hat{\boldsymbol{\alpha}}^{(1)}(z) - \boldsymbol{\alpha}^{(1)}(z) \end{pmatrix}$  and

$$\hat{\boldsymbol{\zeta}}(z) = a_n \hat{\boldsymbol{\vartheta}}(z). \text{ Let } \mathbf{W}_s(z) = \begin{pmatrix} \mathbf{W}_s \\ ((Z_s - z)/h) \mathbf{W}_s \end{pmatrix} \text{ and } Y_s(z) \equiv Y_s - \mathbf{W}_s^T [\boldsymbol{\alpha}(z) + \boldsymbol{\alpha}^{(1)}(z)(Z_s - z)].$$

**Lemma A.2.** *Under Assumptions A1 – A12 in the theorem, one has  $\|\hat{\boldsymbol{\zeta}}(z)\| = O_p(\sqrt{m/nh})$  uniformly over  $z \in \mathcal{D}$ .*

*Proof.* Let  $\mathbf{v} \in \mathbb{R}^{2(\kappa m + 1)}$  be an arbitrary  $2(\kappa m + 1)$ -dimension vector that satisfy  $\|\mathbf{v}\| = 1$ , where  $\|\cdot\|$  is a Euclidean norm. By convexity of the objective function, for any small  $\varepsilon > 0$ , if we can show that there is a large constant  $C$  such that

$$P \left\{ \inf_{\|\mathbf{v}\|=1} \sum_{s=m+1}^n \mathbf{v}^T \psi_\tau(Y_s(z) - (C(m/nh)^{1/2} \mathbf{v})^T \mathbf{W}_s(z)) \mathbf{W}_s(z) K((Z_s - z)/h) > 0 \right\} > 1 - \varepsilon \quad (\text{A.3})$$

uniformly over  $z \in \mathcal{D}$ , then, the proof is finished. We first show that (A.3) holds for any fixed  $z_0 \in \mathcal{D}$ . To this end, define  $z_{s0h} = (Z_s - z_0)/h$  and let  $\xi_s(\mathbf{v}) = \psi_\tau(Y_s(z_0) - \mathbf{v}^T \mathbf{W}_s(z_0)) \mathbf{W}_s(z_0) K(z_{s0h}) - \psi_\tau(Y_s(z_0)) \mathbf{W}_s(z_0) K(z_{s0h})$ , then,

$$\begin{aligned} & \sum_{s=m+1}^n \mathbf{v}^T \psi_\tau(Y_s(z_0) - (C(m/nh)^{1/2} \mathbf{v})^T \mathbf{W}_s(z_0)) \mathbf{W}_s(z_0) K(z_{s0h}) \\ &= \sum_{s=m+1}^n \mathbf{v}^T \psi_\tau(Y_s(z_0)) \mathbf{W}_s(z_0) K(z_{s0h}) + \sum_{s=m+1}^n \mathbf{v}^T E\{\xi_s(C(m/nh)^{1/2} \mathbf{v})\} \\ &+ \sum_{s=m+1}^n \mathbf{v}^T [\xi_s(C(m/nh)^{1/2} \mathbf{v}) - E\{\xi_s(C(m/nh)^{1/2} \mathbf{v})\}] = M_1 + M_2 + M_3. \end{aligned}$$

Following the proof in Xiao and Koenker (2009), we first analyze  $M_3$ . Covering the ball  $\{\|\mathbf{v}\| \leq C(m/nh)^{1/2}\}$  with cubes  $\mathcal{C} = \{\mathcal{C}_k\}$ , where  $\mathcal{C}_k$  is a cube with center  $\mathbf{v}_k$  and side length  $C(m/(nh)^5)^{1/2}$ , so that  $N(n) = \#(\mathcal{C}) = (2(nh)^2)^m$ , and for  $\mathbf{v} \in \mathcal{C}_k$ ,  $\|\mathbf{v} - \mathbf{v}_k\| \leq C(m/(nh)^5)^{1/2}$ .

Since  $I(Y_{\mathfrak{s}}(z_0) < x)$  is nondecreasing in  $x$ , then,

$$\begin{aligned}
& \sup_{\|v\| \leq C(m/nh)^{1/2}} \left| \sum_{\mathfrak{s}=m+1}^n \mathbf{v}^T [\xi_{\mathfrak{s}}(v) - E\{\xi_{\mathfrak{s}}(v)\}] \right| \\
& \leq \max_{1 \leq k \leq N(n)} \left| \sum_{\mathfrak{s}=m+1}^n \mathbf{v}^T [\xi_{\mathfrak{s}}(v_k) - E\{\xi_{\mathfrak{s}}(v_k)\}] \right| \\
& \quad + \max_{1 \leq k \leq N(n)} \left| \sum_{\mathfrak{s}=m+1}^n |(\mathbf{v}^T \mathbf{W}_{\mathfrak{s}}(z_0)) K(z_{\mathfrak{s}0h})| \{b_{n\mathfrak{s}}(v_k) - E(b_{n\mathfrak{s}}(v_k))\} \right| \\
& \quad + \max_{1 \leq k \leq N(n)} \left| \sum_{\mathfrak{s}=m+1}^n |(\mathbf{v}^T \mathbf{W}_{\mathfrak{s}}(z_0)) K(z_{\mathfrak{s}0h})| \{E(d_{n\mathfrak{s}}(v_k))\} \right| \equiv M_{31} + M_{32} + M_{33},
\end{aligned}$$

where  $b_{n\mathfrak{s}}(v_k) = I(Y_{\mathfrak{s}}(z_0) < v_k^T \mathbf{W}_{\mathfrak{s}}(z_0)) - I(Y_{\mathfrak{s}}(z_0) < v_k^T \mathbf{W}_{\mathfrak{s}}(z_0) + C(m/(nh)^{5/2}) \|\mathbf{W}_{\mathfrak{s}}(z_0)\|)$  and  $d_{n\mathfrak{s}}(v_k) = I(Y_{\mathfrak{s}}(z_0) < v_k^T \mathbf{W}_{\mathfrak{s}}(z_0) + C(m/(nh)^{5/2}) \|\mathbf{W}_{\mathfrak{s}}(z_0)\|) - I(Y_{\mathfrak{s}}(z_0) < v_k^T \mathbf{W}_{\mathfrak{s}}(z_0) - C(m/(nh)^{5/2}) \|\mathbf{W}_{\mathfrak{s}}(z_0)\|)$ . The analyses of  $M_{32}$  and  $M_{33}$  are similar to those in Welsh (1989) and Xiao and Koenker (2009), so that our focus here is only on  $M_{31}$ . Notice, for any  $\mathfrak{b} > 0$ ,  $|\psi_{\tau}(Y_{\mathfrak{s}}(z_0)) - \psi_{\tau}(Y_{\mathfrak{s}}(z_0) - v_k^T \mathbf{W}_{\mathfrak{s}}(z_0))|^{\mathfrak{b}} = I(d_{3\mathfrak{s}} < Y_{\mathfrak{s}} \leq d_{4\mathfrak{s}})$ , where  $d_{3\mathfrak{s}} = \min(c_{2\mathfrak{s}}, c_{2\mathfrak{s}} + c_{3\mathfrak{s}})$  and  $d_{4\mathfrak{s}} = \max(c_{2\mathfrak{s}}, c_{2\mathfrak{s}} + c_{3\mathfrak{s}})$  with  $c_{2\mathfrak{s}} = [\boldsymbol{\alpha}(z_0) + \boldsymbol{\alpha}^{(1)}(z_0)(Z_{\mathfrak{s}} - z_0)]^T \mathbf{W}_{\mathfrak{s}}$  and  $c_{3\mathfrak{s}} = v_k^T \mathbf{W}_{\mathfrak{s}}(z_0)$ . Therefore, by Assumption A4, there exists a  $C > 0$  such that  $E\{|\psi_{\tau}(Y_{\mathfrak{s}}(z_0)) - \psi_{\tau}(Y_{\mathfrak{s}}(z_0) - v_k^T \mathbf{W}_{\mathfrak{s}}(z_0))|^{\mathfrak{b}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}} \} = F_{Y|Z, \mathbf{W}}(d_{4\mathfrak{s}}) - F_{Y|Z, \mathbf{W}}(d_{3\mathfrak{s}}) \leq C|v_k^T \mathbf{W}_{\mathfrak{s}}(z_0)| \leq C(m/nh)^{1/2} \|\mathbf{W}_{\mathfrak{s}}(z_0)\|$ , which implies that

$$\begin{aligned}
& E|\mathbf{v}^T \xi_{\mathfrak{s}}(v_k)|^{\delta} = E[|\psi_{\tau}(Y_{\mathfrak{s}}(z_0)) - \psi_{\tau}(Y_{\mathfrak{s}}(z_0) - v_k^T \mathbf{W}_{\mathfrak{s}}(z_0))|^{\delta} |\mathbf{v}^T \mathbf{W}_{\mathfrak{s}}(z_0)|^{\delta} K^{\delta}(z_{\mathfrak{s}0h})] \\
& \leq C(m/nh)^{1/2} E[\|\mathbf{W}_{\mathfrak{s}}(z_0)\| \|\mathbf{W}_{\mathfrak{s}}(z_0)\|^{\delta} K^{\delta}(z_{\mathfrak{s}0h})] \leq C((m/nh)^{1/2} m^{(1+\delta)/2} h)
\end{aligned} \tag{A.4}$$

by Assumption A7. Thus, we have

$$W_n^2 = \sum_{\mathfrak{s}=m+1}^n E[\mathbf{v}^T \{\xi_{\mathfrak{s}}(v_k) - E(\xi_{\mathfrak{s}}(v_k))\}]^2 \leq \sum_{\mathfrak{s}=m+1}^n E[\mathbf{v}^T \xi_{\mathfrak{s}}(v_k)]^2 = O((mnh)^{1/2} m^{3/2})$$

and

$$S_n^2 = \sum_{\mathfrak{s}=m+1}^n [\mathbf{v}^T \{\xi_{\mathfrak{s}}(v_k) - E(\xi_{\mathfrak{s}}(v_k))\}]^2 = O_p((mnh)^{1/2} m^{3/2}).$$

Also, notice that  $\eta_{\mathfrak{s}}(v_k) = \{\xi_{\mathfrak{s}}(v_k) - E(\xi_{\mathfrak{s}}(v_k))\}$  is a martingale difference sequence. Therefore,

let  $L = (mnh)^{1/2}$ . Thus, we have

$$\begin{aligned}
& P \left[ \max_{1 \leq k \leq N(n)} \left| \frac{1}{\sqrt{nh}} \sum_{s=m+1}^n \{\mathbf{v}^T [\xi_s(v_k) - E(\xi_s(v_k))]\} \right| > \epsilon \right] \\
& \leq N(n) \max_k P \left[ \left| \frac{1}{\sqrt{nh}} \sum_{s=m+1}^n \{\mathbf{v}^T [\xi_s(v_k) - E(\xi_s(v_k))]\} \right| > \epsilon \right] \\
& \leq N(n) \max_k P \left[ \left| \sum_{s=m+1}^n \mathbf{v}^T \eta_s(v_k) \right| > \sqrt{nh}\epsilon, W_n^2 + S_n^2 \leq L \right] \\
& \quad + N(n) \max_k P \left[ \left| \sum_{s=m+1}^n \mathbf{v}^T \eta_s(v_k) \right| > \sqrt{nh}\epsilon, W_n^2 + S_n^2 > L \right] \equiv J_1 + J_2. \tag{A.5}
\end{aligned}$$

For  $J_1$ , by exponential inequality for martingale difference sequences (see, e.g., Bercu and Touati, 2008), we have

$$N(n) \max_k P \left[ \left| \sum_{s=m+1}^n \mathbf{v}^T \eta_s(v_k) \right| > \sqrt{nh}\epsilon, W_n^2 + S_n^2 \leq L \right] \leq 2N(n) \exp \left( - \frac{(nh)\epsilon^2}{2L} \right).$$

For  $J_2$ , because  $P[W_n^2 + S_n^2 > L] \leq P[W_n^2 > L] + P[S_n^2 > L]$  and each term can be bounded exponentially under Assumptions A1, A5 and A6. Thus,  $M_3 = o_p((mnh)^{1/2})$ . As for  $M_2$ , notice that

$$\begin{aligned}
M_2 & \equiv \sum_{s=m+1}^n \mathbf{v}^T E\{\xi_s(C(m/nh)^{1/2}\mathbf{v})\} \\
& = \sum_{s=m+1}^n \mathbf{v}^T E\{\psi_\tau(Y_s(z_0) - (C(m/nh)^{1/2}\mathbf{v})^T \mathbf{W}_s(z_0)) \mathbf{W}_s(z_0) K(z_{s0h}) - \psi_\tau(Y_s(z_0)) \mathbf{W}_s(z_0) K(z_{s0h})\} \\
& = \sum_{s=m+1}^n \mathbf{v}^T E\{[F_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) + h z_{s0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s | Z_s, \mathbf{W}_s) \\
& \quad - F_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) + h z_{s0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s \\
& \quad + C(m/nh)^{1/2} \mathbf{v}^T \mathbf{W}_s(z_0) | Z_s, \mathbf{W}_s)] \mathbf{W}_s(z_0) K(z_{s0h})\} \\
& = -C(m/nh)^{1/2} \sum_{s=m+1}^n \mathbf{v}^T E\{f_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) + h z_{s0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s \\
& \quad + \eta C(m/nh)^{1/2} \mathbf{v}^T \mathbf{W}_s(z_0) | Z_s, \mathbf{W}_s) \mathbf{W}_s(z_0) \mathbf{W}_s^T(z_0) K(z_{s0h})\} \mathbf{v},
\end{aligned}$$



where  $\mathbf{W}_s(z_0)\mathbf{W}_s^T(z_0) = \begin{pmatrix} 1 & z_{s0h} \\ z_{s0h} & z_{s0h}^2 \end{pmatrix} \otimes \mathbf{W}_s\mathbf{W}_s^T$ . Similar to the idea in Xu (2005),

$$\begin{aligned} & f_{Y|Z,\mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) + h z_{s0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s + \eta C(m/nh)^{1/2} \mathbf{v}^T \mathbf{W}_s(z_0) | Z_s, \mathbf{W}_s) \\ &= f_{Y|Z,\mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) | Z_s, \mathbf{W}_s) + Ch z_{s0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s + o_p(h), \end{aligned}$$

which implies that

$$\begin{aligned} & \sum_{s=m+1}^n \mathbf{v}^T E\{\xi_s(C(m/nh)^{1/2} \mathbf{v})\} \\ & \approx -C(m/nh)^{1/2} \sum_{s=m+1}^n \mathbf{v}^T E[f_{Y|Z,\mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) | Z_s, \mathbf{W}_s) \mathbf{W}_s(z_0) \mathbf{W}_s^T(z_0) K(z_{s0h})] \mathbf{v} \\ & \quad - C(m/nh)^{1/2} h^2 \sum_{s=m+1}^n \mathbf{v}^T \{E[\|\boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s | \mathbf{W}_s(z_0) \mathbf{W}_s^T(z_0) K_h(z_{s0h})]\} \mathbf{v} = M_{21} + M_{22}. \end{aligned}$$

Again, by Assumption A7,

$$\frac{1}{n} \sum_{s=m+1}^n [\boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s]^2 = \|\boldsymbol{\alpha}^{(1)}(z_0)\|^2 \frac{\boldsymbol{\alpha}^{(1)}(z_0)^T (\frac{1}{n} \sum_{s=m+1}^n \mathbf{W}_s \mathbf{W}_s^T) \boldsymbol{\alpha}^{(1)}(z_0)}{\|\boldsymbol{\alpha}^{(1)}(z_0)\|^2} \leq Cm.$$

Hence,

$$E\left\{\frac{1}{n} \sum_{s=m+1}^n |\boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s|\right\} \leq n^{-1/2} E\left\{\left(\frac{1}{n} \sum_{s=m+1}^n [\boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s]^2\right)^{1/2}\right\} \leq C(m/n)^{1/2},$$

which implies that  $E[\|\boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_s\|] \leq C(m/n)^{1/2}$  and then,  $M_{22} = o((mnh)^{1/2})$ . Thus,

$$\begin{aligned} M_2 & \approx -C(m/nh)^{1/2} \sum_{s=m+1}^n \mathbf{v}^T E[f_{Y|Z,\mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) | Z_s, \mathbf{W}_s) \mathbf{W}_s(z_0) \mathbf{W}_s^T(z_0) K(z_{s0h})] \mathbf{v} \\ & = -\mathbf{v}^T \begin{pmatrix} L_0 & L_1 \\ L_1 & L_2 \end{pmatrix} \mathbf{v}, \end{aligned}$$

where, for  $d = 0, 1$  and  $2$ ,

$$\begin{aligned} L_d &= -C(mn)^{1/2} h^{-1/2} E[f_{Y|Z,\mathbf{W}}(q_\tau(z_0, \mathbf{W}_s) | Z_s, \mathbf{W}_s) z_{s0h}^d \mathbf{W}_s \mathbf{W}_s^T K(z_{s0h})] \\ &= -C(mn)^{1/2} h^{-1/2} E[\mathbf{D}^*(z_0) z_{s0h}^d K(\frac{Z_s - z_0}{h})] \\ &= -C(mnh)^{1/2} \int \mathbf{D}^*(z_0 + hz) z^d K(z) f_z(z_0 + hz) dz \approx -C(mnh)^{1/2} \mu_d f_z(z_0) \mathbf{D}^*(z_0). \end{aligned}$$

In addition, for  $M_1$ , since

$$\begin{aligned}
& E[\psi_\tau(Y_{\mathfrak{s}}(z_0))\mathbf{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h})] = E[\tau - I(Y_{\mathfrak{s}}(z_0) < 0)]\mathbf{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h}) \\
& = E[\tau - F_{Y|Z, \mathbf{W}}(\boldsymbol{\alpha}(z_0)^T \mathbf{W}_{\mathfrak{s}} + h z_{\mathfrak{s}0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_{\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})]\mathbf{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h}) \\
& = E[F_{Y|Z, \mathbf{W}}(q_\tau(Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) - F_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_{\mathfrak{s}}) \\
& \quad + h z_{\mathfrak{s}0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_{\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})]\mathbf{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h}) \\
& = E[f_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_{\mathfrak{s}} \\
& \quad + \eta \Lambda(h, z_0, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \Lambda(h, z_0, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \mathbf{W}_{\mathfrak{s}}(z_0) K(z_{\mathfrak{s}0h})],
\end{aligned}$$

where  $\Lambda(h, z_0, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) = q_\tau(Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) - q_\tau(z_0, \mathbf{W}_{\mathfrak{s}}) - h z_{\mathfrak{s}0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_{\mathfrak{s}}$ , an application of Taylor expansion of  $q_\tau(Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})$  at  $(z_0, \mathbf{W}_{\mathfrak{s}})$  leads to

$$\Lambda(h, z_0, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) = \frac{\boldsymbol{\alpha}^{(2)}(z_0 + \wp h z_{\mathfrak{s}0h})^T}{2} h^2 z_{\mathfrak{s}0h}^2 \mathbf{W}_{\mathfrak{s}}.$$

Therefore, by Assumptions A7 and A10, one has

$$\begin{aligned}
& E[\psi_\tau(Y_{\mathfrak{s}}(z_0))\mathbf{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h})] \\
& = \frac{h^2}{2} E[f_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_{\mathfrak{s}} + \eta \Lambda(h, z_0, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})] \\
& \quad \times \mathbf{W}_{\mathfrak{s}}(z_0) \mathbf{W}_{\mathfrak{s}}^T \boldsymbol{\alpha}^{(2)}(z_0 + \wp h z_{\mathfrak{s}0h}) z_{\mathfrak{s}0h}^2 K(z_{\mathfrak{s}0h}) \\
& = \frac{h^2}{2} E[f_{Y|Z, \mathbf{W}}(q_\tau(z_0, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}0h} \boldsymbol{\alpha}^{(1)}(z_0)^T \mathbf{W}_{\mathfrak{s}} + \eta \Lambda(h, z_0, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})] \\
& \quad \times \begin{pmatrix} 1 \\ z_{\mathfrak{s}0h} \end{pmatrix} \mathbf{D}(Z_{\mathfrak{s}}) \boldsymbol{\alpha}^{(2)}(z_0 + \wp h z_{\mathfrak{s}0h}) z_{\mathfrak{s}0h}^2 K(z_{\mathfrak{s}0h}) \\
& = \frac{h^3}{2} f_z(z_0) \left\{ \begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} \otimes \mathbf{D}^*(z_0) \right\} \boldsymbol{\alpha}^{(2)}(z_0) + o(h^3).
\end{aligned}$$

Thus,  $E[\mathbf{v}^T \psi_\tau(Y_{\mathfrak{s}}(z_0))\mathbf{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h})] = O(m^{1/2}h^3)$ . Then, by Markov's inequality, stationarity and Assumption A10,  $M_1 = \sum_{s=m+1}^n \mathbf{v}^T \psi_\tau(Y_{\mathfrak{s}}(z_0))\mathbf{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h}) = o_p(\sqrt{mnh})$ . Thus,

$$\left\{ \inf_{\|\mathbf{v}\|=1} \sum_{\mathbf{s}=m+1}^n \mathbf{v}^T \psi_\tau(Y_{\mathbf{s}}(z_0) - (C(m/nh)^{1/2} \mathbf{v})^T \mathbf{W}_{\mathbf{s}}(z_0)) \mathbf{W}_{\mathbf{s}}(z_0) K(z_{\mathbf{s}0h}) > 0 \right\} \\ \supseteq \left\{ \frac{C}{2} f_z(z_0) \mathbf{D}^*(z_0) \lambda_{\min} \left[ \mathbf{v}^T \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \mathbf{v} \right] > 0 \right\}$$

with probability going to 1 for a sufficient large  $C$  and as  $n \rightarrow \infty$ . Thus, we complete the first part of the proof.

Next, we show that (A.3) holds uniformly over  $z \in \mathcal{D}$ . To proceed, define  $\mathcal{B} \equiv \{\mathbf{v} : \|\mathbf{v}\| \leq C(m/nh)^{1/2}\}$  and  $K_{z,h} \equiv K((Z_{\mathbf{s}} - z)/h)$ . Then, we want to show that

$$P \left\{ \inf_{z \in \mathcal{D}} \inf_{\mathbf{v} \in \mathcal{B}} \sum_{\mathbf{s}=m+1}^n \mathbf{v}^T \psi_\tau(Y_{\mathbf{s}}(z) - \mathbf{v}^T \mathbf{W}_{\mathbf{s}}(z)) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} > 0 \right\} > 1 - \varepsilon.$$

Since  $\mathcal{D}$  is compact, it can be covered by a finite number  $T(n)$  of cubes  $\mathcal{D}_j = \mathcal{D}_{n,j}$  with side length  $l_n = O(T^{-1}(n)) = O(m^{1/2}(nh)^{-1/4})$  and center  $z_j$ . Clearly,  $l_n = o(1)$  due to Assumption A10. Now, write

$$\begin{aligned} & \sup_{z \in \mathcal{D}} \sup_{\mathbf{v} \in \mathcal{B}} \sum_{\mathbf{s}=m+1}^n \mathbf{v}^T [\psi_\tau(Y_{\mathbf{s}}(z) - \mathbf{v}^T \mathbf{W}_{\mathbf{s}}(z))] \mathbf{W}_{\mathbf{s}}(z) K_{z,h} \\ & \leq \sup_{z \in \mathcal{D}} \sup_{\mathbf{v} \in \mathcal{B}} \sum_{\mathbf{s}=m+1}^n E \{ \mathbf{v}^T \psi_\tau(Y_{\mathbf{s}}(z) - \mathbf{v}^T \mathbf{W}_{\mathbf{s}}(z)) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} \} \\ & \quad + \sup_{z \in \mathcal{D}} \sup_{\mathbf{v} \in \mathcal{B}} \sum_{\mathbf{s}=m+1}^n \left[ \mathbf{v}^T \psi_\tau(Y_{\mathbf{s}}(z) - \mathbf{v}^T \mathbf{W}_{\mathbf{s}}(z)) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} \right. \\ & \quad \left. - E \{ \mathbf{v}^T \psi_\tau(Y_{\mathbf{s}}(z) - \mathbf{v}^T \mathbf{W}_{\mathbf{s}}(z)) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} \} \right] \equiv K^{(1)} + K^{(2)}. \end{aligned}$$

We first consider  $K^{(2)}$ . Let  $\psi_{\tau,\mathbf{s}}(z, \mathbf{v}) = \psi_\tau(Y_{\mathbf{s}}(z) - \mathbf{v}^T \mathbf{W}_{\mathbf{s}}(z))$  for simplicity. Indeed,

$$\begin{aligned} K^{(2)} & \equiv \sup_{z \in \mathcal{D}} \sup_{\mathbf{v} \in \mathcal{B}} \sum_{\mathbf{s}=m+1}^n \left[ \mathbf{v}^T \psi_{\tau,\mathbf{s}}(z, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} - E \{ \mathbf{v}^T \psi_{\tau,\mathbf{s}}(z, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} \} \right] \\ & \leq \max_{1 \leq j \leq T(n)} \sup_{\mathbf{v} \in \mathcal{B}} \left| \sum_{\mathbf{s}=m+1}^n \mathbf{v}^T \left[ \psi_{\tau,\mathbf{s}}(z_j, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z_j) K_{z_j,h} - E \{ \psi_{\tau,\mathbf{s}}(z_j, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z_j) K_{z_j,h} \} \right] \right| \\ & \quad + \max_{1 \leq j \leq T(n)} \sup_{z \in \mathcal{D}_j} \sup_{\mathbf{v} \in \mathcal{B}} \left| \sum_{\mathbf{s}=m+1}^n \left[ \mathbf{v}^T [\psi_{\tau,\mathbf{s}}(z, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} - \psi_{\tau,\mathbf{s}}(z_j, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z_j) K_{z_j,h}] \right. \right. \\ & \quad \left. \left. - E \{ \mathbf{v}^T [\psi_{\tau,\mathbf{s}}(z, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z) K_{z,h} - \psi_{\tau,\mathbf{s}}(z_j, \mathbf{v}) \mathbf{W}_{\mathbf{s}}(z_j) K_{z_j,h}] \} \right] \right| \equiv H^{(1)} + H^{(2)}. \end{aligned}$$

We only focus on  $H^{(2)}$ , since the rate of  $H^{(1)}$  can be controlled in the same way as in (A.5), when  $z$  is fixed. Then,

$$\begin{aligned}
H^{(2)} &= \max_{1 \leq j \leq T(n)} \sup_{z \in \mathcal{D}_j} \sup_{v \in \mathcal{B}} \sum_{s=m+1}^n \left\{ \left| \mathbf{v}^T [\psi_{\tau,s}(z, v) \mathbf{W}_s(z) - \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j)] K_{z_j, h} \right. \right. \\
&\quad \left. \left. - E\{ \mathbf{v}^T [\psi_{\tau,s}(z, v) \mathbf{W}_s(z) - \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j)] K_{z_j, h} \} \right| \right\} \\
&\quad + \max_{1 \leq j \leq T(n)} \sup_{z \in \mathcal{D}_j} \sup_{v \in \mathcal{B}} \sum_{s=m+1}^n \left\{ \left| \mathbf{v}^T [\psi_{\tau,s}(z, v) \mathbf{W}_s(z) - \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j)] \right. \right. \\
&\quad \times [K_{z, h} - K_{z_j, h}] - E\{ \mathbf{v}^T [\psi_{\tau,s}(z, v) \mathbf{W}_s(z) - \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j)] \\
&\quad \times [K_{z, h} - K_{z_j, h}] \} \left. \right| \left. \right\} \\
&\quad + \max_{1 \leq j \leq T(n)} \sup_{z \in \mathcal{D}_j} \sup_{v \in \mathcal{B}} \sum_{s=m+1}^n \left\{ \left| \mathbf{v}^T \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j) [K_{z, h} - K_{z_j, h}] \right. \right. \\
&\quad \left. \left. - E\{ \mathbf{v}^T \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j) [K_{z, h} - K_{z_j, h}] \} \right| \right\} \equiv H^{(21)} + H^{(22)} + H^{(23)}.
\end{aligned}$$

For  $H^{(21)}$ , similar to the derivation of (A.4), one can show by Lipschitz continuity that for any  $\flat > 0$ , there exists a  $C > 0$  such that

$$E\{ |\psi_{\tau,s}(z, v) - \psi_{\tau,s}(z_j, v)|^\flat | Z_s, \mathbf{W}_s \} \leq C m^{1/2} l_n$$

uniformly over  $v \in \mathcal{B}$ , which implies that

$$\begin{aligned}
&E\{ |\mathbf{v}^T [\psi_{\tau,s}(z, v) \mathbf{W}_s(z) - \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j)] K_{z_j, h}|^\delta \} \\
&= E\{ |\mathbf{v}^T [\psi_{\tau,s}(z, v) - \psi_{\tau,s}(z_j, v)] \mathbf{W}_s(z) K_{z_j, h}|^\delta \} \\
&\quad + E\{ |\mathbf{v}^T [\psi_{\tau,s}(z_j, v)] (\mathbf{W}_s(z) - \mathbf{W}_s(z_j)) K_{z_j, h}|^\delta \} \\
&\leq E\{ |\psi_{\tau,s}(z, v) - \psi_{\tau,s}(z_j, v)|^\delta |\mathbf{v}^T \mathbf{W}_s(z)|^\delta K_{z_j, h}^\delta \} \\
&\quad + E\{ |\psi_{\tau,s}(z_j, v)|^\delta |\mathbf{v}^T (\mathbf{W}_s(z) - \mathbf{W}_s(z_j))|^\delta K_{z_j, h}^\delta \} \leq C l_n^\delta m^{\delta/2} h
\end{aligned}$$

by the boundedness of  $\psi_{\tau,s}(z_j, v)$  uniformly over  $v \in \mathcal{B}$ . Define  $\Delta\psi_{\tau,s}(z, z_j) = \psi_{\tau,s}(z, v) \mathbf{W}_s(z) - \psi_{\tau,s}(z_j, v) \mathbf{W}_s(z_j)$ . Thus, we have

$$G_n^2 = \sum_{s=m+1}^n E \left\{ \left| \mathbf{v}^T \Delta\psi_{\tau,s}(z, z_j) K_{z_j, h} - E\{ \mathbf{v}^T \Delta\psi_{\tau,s}(z, z_j) K_{z_j, h} \} \right| \right\}^2$$

$$\leq \sum_{s=m+1}^n E\{\mathbf{v}^T[\psi_{\tau,s}(z, v)\mathbf{W}_s(z) - \psi_{\tau,s}(z_j, v)\mathbf{W}_s(z_j)]K_{z_j,h}\}^2 \leq Cl_n^2 mnh = O((mnh)^{1/2}m^{3/2})$$

and

$$H_n^2 = \sum_{s=m+1}^n \left\{ \left| \mathbf{v}^T \Delta\psi_{\tau,s}(z, z_j)K_{z_j,h} - E\{\mathbf{v}^T \Delta\psi_{\tau,s}(z, z_j)K_{z_j,h}\} \right| \right\}^2 = O_p((mnh)^{1/2}m^{3/2}).$$

Now, let  $\chi_s(z_j) = \Delta\psi_{\tau,s}(z, z_j)K_{z_j,h} - E\{\Delta\psi_{\tau,s}(z, z_j)K_{z_j,h}\}$ . Thus, the fact that  $\Delta\psi_{\tau,s}(z, z_j)K_{z_j,h} - E\{\Delta\psi_{\tau,s}(z, z_j)K_{z_j,h}\}$  is a martingale difference sequence implies that

$$\begin{aligned} & P \left[ \max_{1 \leq j \leq T(n)} \left| \frac{1}{\sqrt{nh}} \sum_{s=m+1}^n \{\mathbf{v}^T [\Delta\psi_{\tau,s}(z, z_j)K_{z_j,h} - E\{\Delta\psi_{\tau,s}(z, z_j)K_{z_j,h}\}]\} \right| > \epsilon \right] \\ & \leq T(n) \max_j P \left[ \left| \frac{1}{\sqrt{nh}} \sum_{s=m+1}^n \{\mathbf{v}^T [\Delta\psi_{\tau,s}(z, z_j)K_{z_j,h} - E\{\Delta\psi_{\tau,s}(z, z_j)K_{z_j,h}\}]\} \right| > \epsilon \right] \\ & \leq T(n) \max_j P \left[ \left| \sum_{s=m+1}^n \mathbf{v}^T \chi_s(z_j) \right| > \sqrt{nh}\epsilon, G_n^2 + H_n^2 \leq (mnh)^{1/2} \right] \\ & \quad + T(n) \max_j P \left[ \left| \sum_{s=m+1}^n \mathbf{v}^T \chi_s(z_j) \right| > \sqrt{nh}\epsilon, G_n^2 + H_n^2 > (mnh)^{1/2} \right] \equiv I^{(1)} + I^{(2)}. \end{aligned}$$

Similar to the derivation in (A.5), under Assumptions A1, A5 and A6, one can show that  $I^{(1)}$  and  $I^{(2)}$  can be bounded exponentially. Hence,  $H^{(21)} = o_p((mnh)^{1/2})$ . We can also show that  $H^{(22)} = o_p((mnh)^{1/2})$  and  $H^{(23)} = o_p((mnh)^{1/2})$  in similar ways. Thus,  $K^{(2)} = o_p((mnh)^{1/2})$ . As for  $K^{(1)}$ , notice that

$$\begin{aligned} K^{(1)} & \equiv \sup_{z \in \mathcal{D}} \sup_{v \in \mathcal{B}} \sum_{s=m+1}^n E\{\mathbf{v}^T \psi_{\tau}(Y_s(z) - v^T \mathbf{W}_s(z)) \mathbf{W}_s(z) K_{z,h}\} \\ & \leq \sup_{z \in \mathcal{D}} \sup_{v \in \mathcal{B}} \sum_{s=m+1}^n E\{\mathbf{v}^T [\psi_{\tau}(Y_s(z) - v^T \mathbf{W}_s(z)) - \psi_{\tau}(Y_s(z))] \mathbf{W}_s(z) K_{z,h}\} \\ & \quad + \sup_{z \in \mathcal{D}} \sup_{v \in \mathcal{B}} \sum_{s=m+1}^n E\{\mathbf{v}^T \psi_{\tau}(Y_s(z)) \mathbf{W}_s(z) K_{z,h}\} \equiv K^{(11)} + K^{(12)}. \end{aligned}$$

In a similar way of calculating  $M_2$ , it can be shown by Assumption A10 that  $K^{(11)} = O((mnh)^{1/2})$  and  $K^{(12)} = O(m^{1/2}nh^3) = o((mnh)^{1/2})$  uniformly  $z \in \mathcal{D}$  and  $v \in \mathcal{B}$ . Therefore, the proof of Lemma A.2 is finished.  $\square$

In the next two lemmas, we focus on  $T_n(\zeta)$  in (A.2) to show stochastic equi-continuity for  $T_n(\zeta) - T_n(0) - E[T_n(\zeta) - T_n(0)]$ , so that we can derive the local Bahadur representation for  $\sqrt{nh}\hat{\zeta}$ . In particular, define  $D_m = \{\zeta : \|\zeta\| \leq C(m/nh)^{1/2}\}$  for each fixed  $0 < C < \infty$ .

**Lemma A.3.** *Under Assumptions A1 – A12, for any  $a \in \mathbb{R}^{2(\kappa m+1)}$  satisfying  $\|a\| = O(1)$ , one has*

$$\sup_{\zeta \in D_m} |a^T \{T_n(\zeta) - T_n(0) - E[T_n(\zeta) - T_n(0)]\}| = o_p(1).$$

*Proof.* For any  $\zeta \in D_m$ , let  $Y_{n\mathfrak{s}}^* = Y_{\mathfrak{s}}^* - \zeta^T \mathbf{W}_{\mathfrak{s}}^*$  and  $M_{n\mathfrak{s}}(\zeta) = [\psi_{\tau}(Y_{n\mathfrak{s}}^*) - \psi_{\tau}(Y_{\mathfrak{s}}^*)] \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})$ . Then,

$$T_n(\zeta) - T_n(0) = a_n \sum_{\mathfrak{s}=m+1 \neq t}^n [\psi_{\tau}(Y_{\mathfrak{s}}^* - \zeta^T \mathbf{W}_{\mathfrak{s}}^*) - \psi_{\tau}(Y_{\mathfrak{s}}^*)] \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h}) = a_n \sum_{\mathfrak{s}=m+1 \neq t}^n M_{n\mathfrak{s}}(\zeta)$$

and  $M_{n\mathfrak{s}}(\zeta) = [\psi_{\tau}(Y_{n\mathfrak{s}}^*) - \psi_{\tau}(Y_{\mathfrak{s}}^*)] \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h}) = \left( M_{n\mathfrak{s}}^{(1)}(\zeta), M_{n\mathfrak{s}}^{(2)}(\zeta) \right)^T$  with  $M_{n\mathfrak{s}}^{(1)}(\zeta) = [\psi_{\tau}(Y_{n\mathfrak{s}}^*) - \psi_{\tau}(Y_{\mathfrak{s}}^*)] \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})$  and  $M_{n\mathfrak{s}}^{(2)}(\zeta) = [\psi_{\tau}(Y_{n\mathfrak{s}}^*) - \psi_{\tau}(Y_{\mathfrak{s}}^*)] \mathbf{W}_{\mathfrak{s}}^* z_{\mathfrak{s}h} K(z_{\mathfrak{s}h})$ . Thus,

$$\begin{aligned} & \sup_{\zeta \in D_m} |a^T \{T_n(\zeta) - T_n(0) - E[T_n(\zeta) - T_n(0)]\}| \\ & \leq a_n \sup_{\zeta \in D_m} \left| \sum_{\mathfrak{s}=m+1 \neq t}^n a_1^T (M_{n\mathfrak{s}}^{(1)}(\zeta) - E M_{n\mathfrak{s}}^{(1)}(\zeta)) \right| + a_n \sup_{\zeta \in D_m} \left| \sum_{\mathfrak{s}=m+1 \neq t}^n a_2^T (M_{n\mathfrak{s}}^{(2)}(\zeta) - E M_{n\mathfrak{s}}^{(2)}(\zeta)) \right| \\ & \equiv a_n \sup_{\zeta \in D_m} \left| \sum_{\mathfrak{s}=m+1 \neq t}^n \{M_{n\mathfrak{s}}^{(1a_1)}(\zeta) - E(M_{n\mathfrak{s}}^{(1a_1)}(\zeta))\} \right| + a_n \sup_{\zeta \in D_m} \left| \sum_{\mathfrak{s}=m+1 \neq t}^n \{M_{n\mathfrak{s}}^{(2a_2)}(\zeta) - E(M_{n\mathfrak{s}}^{(2a_2)}(\zeta))\} \right| \\ & \equiv M_n^{(1)}(\zeta) + M_n^{(2)}(\zeta), \end{aligned}$$

where  $a_1 \in \mathbb{R}^{\kappa m+1}$  and  $a_2 \in \mathbb{R}^{\kappa m+1}$  are partitions of  $a$ . For  $M_n^{(1)}(\zeta)$ , it is easy to see that  $M_n^{(1)}(\zeta) \equiv a_n \sup_{\zeta \in D_m} \left| \sum_{\mathfrak{s}=m+1 \neq t}^n \{M_{n\mathfrak{s}}^{(1a_1)}(\zeta) - E(M_{n\mathfrak{s}}^{(1a_1)}(\zeta))\} \right|$ , where  $M_{n\mathfrak{s}}^{(1a_1)}(\zeta) = a_1^T M_{n\mathfrak{s}}^{(1)}(\zeta)$ . Similar to the proof of Lemma A.2, for any  $b > 0$ ,  $|\psi_{\tau}(Y_{n\mathfrak{s}}^*) - \psi_{\tau}(Y_{\mathfrak{s}}^*)|^b = I(a_{3\mathfrak{s}} < Y_t \leq a_{4\mathfrak{s}})$ , where  $a_{3\mathfrak{s}} = \min(b_{2\mathfrak{s}}, b_{2\mathfrak{s}} + b_{3\mathfrak{s}})$  and  $a_{4\mathfrak{s}} = \max(b_{2\mathfrak{s}}, b_{2\mathfrak{s}} + b_{3\mathfrak{s}})$  with  $b_{2\mathfrak{s}} = [\boldsymbol{\alpha}(Z_t) + \boldsymbol{\alpha}^{(1)}(Z_t)(Z_{\mathfrak{s}} - Z_t)]^T \mathbf{W}_{\mathfrak{s}}$  and  $b_{3\mathfrak{s}} = \zeta^T \mathbf{W}_{\mathfrak{s}}^*$ . Therefore, by Assumption A4, there exists a  $C > 0$  such that

$$E\{|\psi_{\tau}(Y_{n\mathfrak{s}}^*) - \psi_{\tau}(Y_{\mathfrak{s}}^*)|^b | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}\} = F_{Y|Z, \mathbf{W}}(a_{4\mathfrak{s}}) - F_{Y|Z, \mathbf{W}}(a_{3\mathfrak{s}}) \leq C|\zeta^T \mathbf{W}_{\mathfrak{s}}^*|,$$

which implies by Assumption A7 that

$$\begin{aligned}
E|M_{n1}^{(1a_1)}(\zeta)|^\delta &= E[|\psi_\tau(Y_{n1}^*) - \psi_\tau(Y_1^*)|^\delta |a_1^T \mathbf{W}_1|^\delta K^\delta(z_{1h})] \\
&\leq CE[|\zeta^T \mathbf{W}_1^*|^\delta \|\mathbf{W}_1\|^\delta K^\delta(z_{1h})] \leq C(a_n m^{1/2} m^{(1+\delta)/2} h). \tag{A.6}
\end{aligned}$$

Similar to the proof of Lemma A.2, covering the ball  $D_m$  with cubes  $\mathcal{C} = \{\mathcal{C}_k\}$ , where  $\mathcal{C}_k$  is a cube with center  $\zeta_k$  and side length  $C(m/nh)^{1/2}$ , so that  $N(n) = \#(\mathcal{C}) = (2(nh)^2)^m$ , and for  $\zeta \in \mathcal{C}_k$ ,  $\|\zeta - \zeta_k\| \leq C(m/(nh)^{5/2})$ . Since  $I(Y_s^* < x)$  is nondecreasing in  $x$ , then,

$$\begin{aligned}
M_n^{(1)}(\zeta) &\equiv a_n \sup_{\zeta \in D_m} \left| \sum_{s=m+1 \neq t}^n \{M_{ns}^{(1a_1)}(\zeta) - E(M_{ns}^{(1a_1)}(\zeta))\} \right| \\
&\leq \max_{1 \leq k \leq N(n)} a_n \left| \sum_{s=m+1 \neq t}^n \{M_{ns}^{(1a_1)}(\zeta_k) - E(M_{ns}^{(1a_1)}(\zeta_k))\} \right| \\
&\quad + \max_{1 \leq k \leq N(n)} \left| \sum_{s=m+1 \neq t}^n |(a_1^T \mathbf{W}_s) K(z_{sh})| \{b_{ns}^{(1a_1)}(\zeta_k) - E(b_{ns}^{(1a_1)}(\zeta_k))\} \right| \\
&\quad + \max_{1 \leq k \leq N(n)} \left| \sum_{s=m+1 \neq t}^n |(a_1^T \mathbf{W}_s) K(z_{sh})| \{E(d_{ns}^{(1a_1)}(\zeta_k))\} \right| \equiv K_1 + K_2 + K_3,
\end{aligned}$$

where  $b_{ns}^{(1a_1)}(\zeta_k) = I(Y_s^* < \zeta_k^T \mathbf{W}_s) - I(Y_s^* < \zeta_k^T \mathbf{W}_s + C(m/(nh)^{5/2}) \|\mathbf{W}_s\|)$  and

$$d_{nt}^{(1a_1)}(\zeta_k) = I(Y_s^* < \zeta_k^T \mathbf{W}_s + C(m/(nh)^{5/2}) \|\mathbf{W}_s\|) - I(Y_s^* < \zeta_k^T \mathbf{W}_s - C(m/(nh)^{5/2}) \|\mathbf{W}_s\|).$$

Now, our focus is only on  $K_1$ . By noting that  $N(n) = (2(nh)^2)^m$  and  $\|\zeta_k\| \leq C(m/nh)^{1/2}$  and  $\kappa$  is fixed, it follows by (A.6) that

$$Q_n^2 = \sum_{s=m+1 \neq t}^n E\{M_{ns}^{(1a_1)}(\zeta_k) - E(M_{ns}^{(1a_1)}(\zeta_k))\}^2 \leq \sum_{s=m+1 \neq t}^n E[M_{ns}^{(1a_1)}(\zeta_k)]^2 = O((mnh)^{1/2} m^{3/2})$$

and

$$R_n^2 = \sum_{s=m+1 \neq t}^n \{M_{ns}^{(1a_1)}(\zeta_k) - E(M_{ns}^{(1a_1)}(\zeta_k))\}^2 = O_p((mnh)^{1/2} m^{3/2}).$$

Also, notice that  $\varphi_s(\zeta_k) = \{M_{ns}^{(1)}(\zeta_k) - E(M_{ns}^{(1)}(\zeta_k))\}$  is a martingale difference sequence. Therefore, let  $L = (mnh)^{1/2}$ , we have

$$\begin{aligned}
& P \left[ \max_{1 \leq k \leq N(n)} \left| a_n \sum_{s=m+1 \neq t}^n \{M_{ns}^{(1a_1)}(\zeta_k) - E(M_{ns}^{(1a_1)}(\zeta_k))\} \right| > \epsilon \right] \\
& \leq N(n) \max_k P \left[ \left| \frac{1}{\sqrt{nh}} \sum_{s=m+1 \neq t}^n \{M_{ns}^{(1a_1)}(\zeta_k) - E(M_{ns}^{(1a_1)}(\zeta_k))\} \right| > \epsilon \right] \\
& \leq N(n) \max_k P \left[ \left| \sum_{s=m+1 \neq t}^n a_1^T \varphi_s(\zeta_k) \right| > \sqrt{nh}\epsilon, Q_n^2 + R_n^2 \leq L \right] \\
& \quad + N(n) \max_k P \left[ \left| \sum_{s=m+1 \neq t}^n a_1^T \varphi_s(\zeta_k) \right| > \sqrt{nh}\epsilon, Q_n^2 + R_n^2 > L \right] \equiv K_{11} + K_{12}.
\end{aligned}$$

For  $K_{11}$ , by exponential inequality for martingale difference sequences (see, e.g., Bercu and Touati, 2008), we have

$$N(n) \max_k P \left[ \left| \sum_{s=m+1 \neq t}^n a_1^T \varphi_s(\zeta_k) \right| > \sqrt{nh}\epsilon, Q_n^2 + R_n^2 \leq L \right] \leq 2N(n) \exp \left( -\frac{(nh)\epsilon^2}{2L} \right).$$

For  $K_{12}$ , because

$$P [Q_n^2 + R_n^2 > L] \leq P [Q_n^2 > L] + P [R_n^2 > L]$$

and each term can be bounded exponentially under Assumptions A1, A5 and A6. Thus,  $M_n^{(1)}(\zeta) = o_p(1)$ . Similarly, it can be shown that  $M_n^{(2)}(\zeta) = o_p(1)$ . These complete the proof of the lemma.  $\square$

**Lemma A.4.** *Under Assumptions A1 – A12, for any  $a \in \mathbb{R}^{2(\kappa m+1)}$  satisfying  $\|a\| = O(1)$ , one has*

$$\sup_{\zeta \in D_m} \|a^T \{E[T_n(\zeta) - T_n(0)] + f_z(Z_t) \mathbf{D}_1^*(Z_t) \sqrt{nh} \zeta\}\| = o(1),$$

where  $\mathbf{D}_1^*(Z_t) = \text{diag}\{\mathbf{D}^*(Z_t), \mu_2 \mathbf{D}^*(Z_t)\}$ .

*Proof.* First, notice that

$$\begin{aligned}
& a_n \sum_{s=m+1 \neq t}^n E[(\psi_\tau(Y_s^* - \zeta^T \mathbf{W}_s^*) - \psi_\tau(Y_s^*)) \mathbf{W}_s^* K(z_{sh})] \\
& = a_n \sum_{s=m+1 \neq t}^n E[I(Y_s^* < 0) - I(Y_s^* < \zeta^T \mathbf{W}_s^*)] \mathbf{W}_s^* K(z_{sh})
\end{aligned}$$



$$\begin{aligned}
&= a_n \sum_{\mathfrak{s}=m+1 \neq t}^n E[F_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \\
&\quad - F_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} + b_{3\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})] \\
&= -\frac{1}{nh} \sum_{\mathfrak{s}=m+1 \neq t}^n E[f_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} \\
&\quad + \varpi b_{3\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \mathbf{W}_{\mathfrak{s}}^* \mathbf{W}_{\mathfrak{s}}^{*T} \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})],
\end{aligned}$$

where  $\mathbf{W}_{\mathfrak{s}}^* \mathbf{W}_{\mathfrak{s}}^{*T} = \begin{pmatrix} 1 & z_{\mathfrak{s}h} \\ z_{\mathfrak{s}h} & z_{\mathfrak{s}h}^2 \end{pmatrix} \otimes \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T$ . Therefore, similar to the proof of Lemma A.2,

$$\begin{aligned}
&f_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} + \varpi b_{3\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \\
&= f_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) + C h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} + o_p(h).
\end{aligned}$$

Hence, it follows that

$$a_n \sum_{\mathfrak{s}=m+1 \neq t}^n E[(\psi_\tau(Y_{\mathfrak{s}}^* - \boldsymbol{\zeta}^T \mathbf{W}_{\mathfrak{s}}^*) - \psi_\tau(Y_{\mathfrak{s}}^*)) \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})] = \begin{pmatrix} A_0 & A_1 \\ A_1 & A_2 \end{pmatrix} + o(1),$$

where for  $d = 0, 1$  and  $2$ ,

$$\begin{aligned}
A_d &= -\frac{1}{nh} \sum_{\mathfrak{s}=m+1 \neq t}^n E[f_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) + g(Z_t, h, Z, \mathbf{W}, \varpi) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \\
&\quad \times z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})] \\
&= -\frac{1}{h} E[f_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})] \\
&\quad - \frac{1}{h} E\{g(Z_t, h, Z, \mathbf{W}, \varpi) z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})\} \\
&= -\frac{1}{h} E[f_{Y|Z, \mathbf{W}}(q_\tau(Z_t, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})] \\
&\quad - CE\{|\boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} | z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})\} + o(1) \\
&= -\frac{1}{h} E[\mathbf{D}^*(Z_{\mathfrak{s}}) \sqrt{nh} \boldsymbol{\zeta} z_{\mathfrak{s}h}^d K(\frac{Z_{\mathfrak{s}} - Z_t}{h})] \\
&\quad - CE\{|\boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} | z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})\} + o(1)
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{h} \int \mathbf{D}^*(z) \sqrt{nh} \boldsymbol{\zeta} \left( \frac{z - Z_t}{h} \right)^d K \left( \frac{z - Z_t}{h} \right) f_z(z) dz \\
&\quad - CE\{|\boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}}| z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})\} + o(1) \\
&= - \int \mathbf{D}^*(Z_t + hz) \sqrt{nh} \boldsymbol{\zeta} z^d K(z) f_z(Z_t + hz) dz \\
&\quad - CE\{|\boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}}| z_{\mathfrak{s}h}^d \mathbf{W}_{\mathfrak{s}} \mathbf{W}_{\mathfrak{s}}^T \sqrt{nh} \boldsymbol{\zeta} K(z_{\mathfrak{s}h})\} + o(1),
\end{aligned}$$

with  $g(Z_t, h, Z, \mathbf{W}, \varpi) = h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} + \varpi b_{3\mathfrak{s}}$ . Note that

$$- \int \mathbf{D}^*(Z_t + hz) \sqrt{nh} \boldsymbol{\zeta} z^d K(z) f_z(Z_t + hz) dz + o(1) = -\mu_d f_z(Z_t) \mathbf{D}^*(Z_t) \sqrt{nh} \boldsymbol{\zeta} + o(1).$$

Also, by Assumption A7, one has  $E[|\boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}}|] \leq C(m/n)^{1/2}$ . Then, by choosing sufficiently large  $C > 0$  and by Assumption A10,  $\|E[T_n(\boldsymbol{\zeta}) - T_n(0)] + f_z(Z_t) \mathbf{D}_1^*(Z_t) \sqrt{nh} \boldsymbol{\zeta}\| \leq C m n^{-1/2} m h = o(1)$ . Thus,  $|a^T \{E[T_n(\boldsymbol{\zeta}) - T_n(0)] + f_z(Z_t) \mathbf{D}_1^*(Z_t) \sqrt{nh} \boldsymbol{\zeta}\}| \leq C \|E[T_n(\boldsymbol{\zeta}) - T_n(0)] + f_z(Z_t) \mathbf{D}_1^*(Z_t) \sqrt{nh} \boldsymbol{\zeta}\| = o(1)$ . Combining the above analysis with the methods of constructing cubes in the proof of Lemma A.3, the lemma is proved.  $\square$

**Lemma A.5.** *Let  $S_{\mathfrak{s}} = \psi_{\tau}(Y_{\mathfrak{s}}^*) \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})$ . Under Assumptions A1 – A12, for  $1 \leq \mathfrak{s} \neq t \leq n$  and for any fixed  $Z_t \neq Z_s$ , one has*

$$E[S_{\mathfrak{s}}] = \frac{h^3 f_z(Z_t)}{2} \begin{pmatrix} \mu_2 \mathbf{D}^*(Z_t) \boldsymbol{\alpha}^{(2)}(Z_t) \\ 0 \end{pmatrix} + o(h^3),$$

and

$$\text{Var}[S_{\mathfrak{s}}] = h\tau(1 - \tau) f_z(Z_t) \mathbf{D}_1(Z_t) + o(h),$$

where  $\mathbf{D}_1(Z_t) = \text{diag}\{\nu_0 \mathbf{D}(Z_t), \nu_2 \mathbf{D}(Z_t)\}$ . Further,

$$\text{Var}[T_n(0)] \rightarrow \tau(1 - \tau) f_z(Z_t) \mathbf{D}_1(Z_t).$$

Therefore,  $\|T_n(0)\| = O_p(1)$ .

*Proof.* This proof follows from the proof of Lemma 3.5 in Xu (2005). Firstly, we calculate  $E[S_{\mathfrak{s}}]$ .

Indeed,

$$\begin{aligned}
E[S_{\mathfrak{s}}] &= E[\psi_{\tau}(Y_{\mathfrak{s}}^*) \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})] = E[\tau - I(Y_{\mathfrak{s}}^* < 0)] \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h}) \\
&= E[\tau - F_{Y|Z, \mathbf{W}}(\boldsymbol{\alpha}(Z_t)^T \mathbf{W}_{\mathfrak{s}} + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})] \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h}) \\
&= E[F_{Y|Z, \mathbf{W}}(q_{\tau}(Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) - F_{Y|Z, \mathbf{W}}(q_{\tau}(Z_t, \mathbf{W}_{\mathfrak{s}}) \\
&\quad + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})] \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})\} \\
&= E\{f_{Y|Z, \mathbf{W}}(q_{\tau}(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} \\
&\quad + \xi \Lambda(h, Z_t, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \Lambda(h, Z_t, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \mathbf{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h})\},
\end{aligned}$$

where  $\Lambda(h, Z_t, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) = q_{\tau}(Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) - q_{\tau}(Z_t, \mathbf{W}_{\mathfrak{s}}) - h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}}$ . An application of the Taylor expansion of  $q_{\tau}(Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})$  at  $(Z_t, \mathbf{W}_{\mathfrak{s}})$  leads to

$$\Lambda(h, Z_t, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) = \frac{\boldsymbol{\alpha}^{(2)}(Z_t + \varsigma h z_{\mathfrak{s}h})^T}{2} h^2 z_{\mathfrak{s}h}^2 \mathbf{W}_{\mathfrak{s}}.$$

Therefore, similar to the proof in Lemma A.2,

$$\begin{aligned}
E[S_{\mathfrak{s}}] &= \frac{h^2}{2} E[f_{Y|Z, \mathbf{W}}(q_{\tau}(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} + \xi \Lambda(h, Z_t, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \\
&\quad \times \mathbf{W}_{\mathfrak{s}}^* \mathbf{W}_{\mathfrak{s}}^T \boldsymbol{\alpha}^{(2)}(Z_t + \varsigma h z_{\mathfrak{s}h}) z_{\mathfrak{s}h}^2 K(z_{\mathfrak{s}h})] \\
&= \frac{h^2}{2} E\left\{ f_{Y|Z, \mathbf{W}}(q_{\tau}(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}h} \boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} + \xi \Lambda(h, Z_t, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) | Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \right. \\
&\quad \times \left. \begin{pmatrix} 1 \\ z_{\mathfrak{s}h} \end{pmatrix} \mathbf{D}(Z_{\mathfrak{s}}) \boldsymbol{\alpha}^{(2)}(Z_t + \varsigma h z_{\mathfrak{s}h}) z_{\mathfrak{s}h}^2 K(z_{\mathfrak{s}h}) \right\} \\
&= \frac{h^3}{2} f_z(Z_t) \left\{ \begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} \otimes \mathbf{D}^*(Z_t) \right\} \boldsymbol{\alpha}^{(2)}(Z_t) + o(h^3). \tag{A.7}
\end{aligned}$$

As for  $E[S_{\mathfrak{s}} S_{\mathfrak{s}}^T]$ , one has

$$\begin{aligned}
E[S_{\mathfrak{s}} S_{\mathfrak{s}}^T] &= E[\psi_{\tau}^2(Y_{\mathfrak{s}}^*) \mathbf{W}_{\mathfrak{s}}^* \mathbf{W}_{\mathfrak{s}}^{*T} K^2(z_{\mathfrak{s}h})] \\
&= E\{[\tau^2 - (2\tau - 1)I(Y_{\mathfrak{s}}^* < 0)] \mathbf{W}_{\mathfrak{s}}^* \mathbf{W}_{\mathfrak{s}}^{*T} K^2(z_{\mathfrak{s}h})\} \\
&= (2\tau - 1)E\{[\tau - I(Y_{\mathfrak{s}}^* < 0)] \mathbf{W}_{\mathfrak{s}}^* \mathbf{W}_{\mathfrak{s}}^{*T} K^2(z_{\mathfrak{s}h})\} + \tau(1 - \tau)E[\mathbf{W}_{\mathfrak{s}}^* \mathbf{W}_{\mathfrak{s}}^{*T} K^2(z_{\mathfrak{s}h})] \equiv R^{(1)} + R^{(2)}.
\end{aligned}$$

Similar to the above derivation, it is not difficult to show that

$$\begin{aligned}
R^{(1)} &\equiv (2\tau - 1)E\{[\tau - I(Y_{\mathfrak{s}}^* < 0)]\mathbf{W}_{\mathfrak{s}}^*\mathbf{W}_{\mathfrak{s}}^{*T}K^2(z_{sh})\} \\
&= (2\tau - 1)E\{[F_{Y|Z,\mathbf{W}}(q_{\tau}(Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})|Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}}) \\
&\quad - F_{Y|Z,\mathbf{W}}(q_{\tau}(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{sh}\boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}}|Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})]\mathbf{W}_{\mathfrak{s}}^*\mathbf{W}_{\mathfrak{s}}^{*T}K^2(z_{sh})\} \\
&= (2\tau - 1)E[f_{Y|Z,\mathbf{W}}(q_{\tau}(Z_t, \mathbf{W}_{\mathfrak{s}}) + h z_{sh}\boldsymbol{\alpha}^{(1)}(Z_t)^T \mathbf{W}_{\mathfrak{s}} \\
&\quad + \xi\Lambda(h, Z_t, Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})|Z_{\mathfrak{s}}, \mathbf{W}_{\mathfrak{s}})\frac{\boldsymbol{\alpha}^{(2)}(Z_t + \zeta h z_{sh})^T \mathbf{W}_{\mathfrak{s}}}{2}h^2 z_{sh}^2 \\
&\quad \times \mathbf{W}_{\mathfrak{s}}^*\mathbf{W}_{\mathfrak{s}}^{*T}K^2(z_{sh})] = o(h^2)
\end{aligned}$$

and

$$R^{(2)} \equiv \tau(1 - \tau)E[\mathbf{W}_{\mathfrak{s}}^*\mathbf{W}_{\mathfrak{s}}^{*T}K^2(z_{sh})] = h\tau(1 - \tau)f_z(Z_t) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \mathbf{D}(Z_t)(1 + o(1)). \quad (\text{A.8})$$

Next, it is shown that the last part of lemma holds true. To this end, it is easy to check that

$$\begin{aligned}
\text{Var}[T_n(0)] &\leq \frac{1}{h}[ \text{Var}(S_1) + 2 \sum_{\ell=1}^{n-1} (1 - \frac{\ell}{n}) \text{Cov}(S_1, S_{\ell+1}) ] \\
&\leq \frac{1}{h} \text{Var}(S_1) + \frac{2}{h} \sum_{\ell=1}^{d_n-1} |\text{Cov}(S_1, S_{\ell+1})| + \frac{2}{h} \sum_{\ell=d_n}^{\infty} |\text{Cov}(S_1, S_{\ell+1})| \equiv J_4 + J_5 + J_6
\end{aligned}$$

By (A.7) and (A.8),

$$J_4 \rightarrow \tau(1 - \tau)f_z(Z_t) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \mathbf{D}(Z_t).$$

Now, it remains to show that  $|J_5| = o(1)$  and  $|J_6| = o(1)$ . First, we consider  $J_6$ . To this end, using Davydov's inequality (see, e.g., Corollary A.2 of Hall and Heyde (1980)) and the boundedness of  $\psi_{\tau}(\cdot)$ , one has

$$|\text{Cov}(S_1, S_{\ell+1})| \leq C\alpha^{1-2/\delta}(\ell)[E|S_1|^{\delta}]^{2/\delta} \leq Cmh^{2/\delta}\alpha^{1-2/\delta}(\ell),$$

which gives

$$J_6 \leq Cmh^{2/\delta-1} \sum_{\ell=d_n}^{\infty} \alpha^{1-2/\delta}(\ell) \leq Cmh^{2/\delta-1}d_n^{-\mathfrak{d}} \sum_{\ell=d_n}^{\infty} \ell^{\mathfrak{d}} \alpha^{1-2/\delta}(\ell) = o(mh^{2/\delta-1}d_n^{-\mathfrak{d}}) = o(1),$$

by choosing  $d_n$  to satisfy  $d_n^{\mathfrak{d}}m^{-1}h^{1-2/\delta} = c$ . As for  $J_5$ , following the proof of Lemma 3.5 in Xu (2005), one has  $|J_5| = o(1)$ . These prove Lemma A.5.  $\square$

**Lemma A.6.** (*Bahadur representation*) Under Assumptions A1 – A12, for any fixed  $Z_t \neq Z_s$ , one has,

$$\hat{\boldsymbol{\vartheta}} \equiv \sqrt{nh_1} \hat{\boldsymbol{\zeta}} = \frac{1}{\sqrt{nh_1} f_z(Z_t)} (\mathbf{D}_1^*(Z_t)^{-1}) \sum_{s=m+1 \neq t}^n \psi_\tau(Y_s^*) \mathbf{W}_s^* K(z_{sh_1}) + o_p(1),$$

$$\text{where } \mathbf{D}_1^*(Z_t) = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \otimes \mathbf{D}^*(Z_t).$$

*Proof.* We first derive the local Bahadur representation for  $\hat{\boldsymbol{\vartheta}}$ . Indeed, by Lemma A.2,  $\|\hat{\boldsymbol{\zeta}}\| = O_p((m/nh)^{1/2})$ . On the other hand, by Lemmas A.3, A.4 and A.5,  $T_n(\boldsymbol{\zeta})$  satisfies  $\|T_n(0)\| = O_p(1)$  and  $\sup_{\|\boldsymbol{\zeta}\| \leq C(m/nh)^{1/2}} |a^T \{T_n(\boldsymbol{\zeta}) + D\sqrt{nh}\boldsymbol{\zeta} - T_n(0)\}| = o_p(1)$  with  $D = f_z(Z_t) \mathbf{D}_1^*(Z_t)$ . In addition, it follows from Assumption A10 and Lemma A.1 that  $\|T_n(\hat{\boldsymbol{\zeta}})\| = o_p(1)$ . Then, replacing  $a$  by  $D^{-1}a$ , the lemma is proved.  $\square$

**Lemma A.7.** Define  $K_{n\mathfrak{L}} = \{(\Delta, \boldsymbol{\vartheta}) : \|\boldsymbol{\vartheta}\| \leq \mathfrak{L}, \|\Delta\| \leq M\}$  for some  $0 < M < \infty$  and  $0 < \mathfrak{L} < \infty$ , let  $V_n(\boldsymbol{\vartheta})$  and  $V_n(\Delta, \boldsymbol{\vartheta})$  be vectors that satisfy (i)  $-\Delta^T V_n(\lambda \Delta, \boldsymbol{\vartheta}) \geq -\Delta^T V_n(\Delta, \boldsymbol{\vartheta})$  for  $\lambda \geq 1$  and  $\|\boldsymbol{\vartheta}\| \leq \mathfrak{L}$ , and (ii)

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

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

*Proof.* The proof follows from Koenker and Zhao (1996) and Conditions (i) and (ii) that

$$V_n(\Delta_n, \boldsymbol{\vartheta}_n) + V_n(\boldsymbol{\vartheta}_n) + D\Delta_n - A_n = o_p(1). \text{ This completes the proof. } \square$$

To show Lemmas A.8 and A.9 later,  $\tau$  is dropped from  $\mathbf{g}_\tau(z_0)$  and  $h_2$  is written as  $h$  for simplicity. For the notational convenience again, define  $b_n = (nh_2)^{-1/2}$ , let  $\boldsymbol{\theta}_0 = b_n^{-1}(\Theta_0 - \mathbf{g}(z_0))$  and  $\boldsymbol{\theta}_1 = hb_n^{-1}(\Theta_1 - \mathbf{g}^{(1)}(z_0))$ . Then,  $\boldsymbol{\theta} = b_n^{-1} \mathbf{H}_2 \begin{pmatrix} \Theta_0 - \mathbf{g}(z_0) \\ \Theta_1 - \mathbf{g}^{(1)}(z_0) \end{pmatrix}$ , where  $\mathbf{H}_2 = \text{diag}\{I_{2\kappa+1}, h_2 I_{2\kappa+1}\}$ . For convenience of analysis, we rewrite  $\hat{\mathbf{X}}_t \equiv \mathbf{X}_t(\hat{\boldsymbol{\vartheta}}_0) \equiv \mathbf{X}_t(\boldsymbol{\alpha}(Z_t) + (nh_1)^{-1/2} \hat{\boldsymbol{\vartheta}}_0)$  because it contains  $\hat{q}_{\tau,t} = \mathbf{W}_t^T \hat{\boldsymbol{\delta}}_0$ . Similarly,  $\mathbf{X}_t(\boldsymbol{\vartheta}_0) \equiv \mathbf{X}_t(\boldsymbol{\alpha}(Z_t) + (nh_1)^{-1/2} \boldsymbol{\vartheta}_0)$ ,  $\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) \equiv \mathbf{X}_t^*(\boldsymbol{\alpha}(Z_t) +$

$(nh_1)^{-1/2}\boldsymbol{\vartheta}_0$  and  $\hat{\mathbf{X}}_t^* \equiv \mathbf{X}_t^*(\hat{\boldsymbol{\vartheta}}_0) \equiv \mathbf{X}_t^*(\boldsymbol{\alpha}(Z_t) + (nh_1)^{-1/2}\hat{\boldsymbol{\vartheta}}_0)$ , where  $\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) = \begin{pmatrix} \mathbf{X}_t(\boldsymbol{\vartheta}_0) \\ z_{th}\mathbf{X}_t(\boldsymbol{\vartheta}_0) \end{pmatrix}$

and  $\mathbf{X}_t^*(\hat{\boldsymbol{\vartheta}}_0) = \begin{pmatrix} \mathbf{X}_t(\hat{\boldsymbol{\vartheta}}_0) \\ z_{th}\mathbf{X}_t(\hat{\boldsymbol{\vartheta}}_0) \end{pmatrix}$  and  $z_{th} = (Z_t - z_0)/h$ . Of course,  $\mathbf{X}_t^*(0) \equiv \mathbf{X}_t^* = \begin{pmatrix} \mathbf{X}_t \\ z_{th}\mathbf{X}_t \end{pmatrix}$ .

Hence,  $\partial\mathbf{X}_t(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\vartheta}_0 = a_n\boldsymbol{\Pi}_t$ , where  $\boldsymbol{\Pi}_t^T = (0_{1 \times (\kappa m + 1)}^T, \mathbf{W}_t, \dots, \mathbf{W}_t, 0_{\kappa \times (\kappa m + 1)}^T)$  has the same definition as that in the main article. Next, denote  $v_t^*(\boldsymbol{\vartheta}_0) = Y_t - \mathbf{X}_t^T(\boldsymbol{\vartheta}_0)[\mathbf{g}(z_0) + \mathbf{g}^{(1)}(z_0)(Z_t - z_0)]$ ,  $v_t^*(0) = Y_t - \mathbf{X}_t^T[\mathbf{g}(z_0) + \mathbf{g}^{(1)}(z_0)(Z_t - z_0)]$  and  $v_{nt}^* = v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) = v_t^*(\boldsymbol{\vartheta}_0) - b_n\boldsymbol{\theta}^T\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)$ . In addition, define  $\Gamma^*(Z_t) = E[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t))\mathbf{X}_t^*\mathbf{g}_\tau(z_0)^T\boldsymbol{\Pi}_t|Z_t]$  and  $\Gamma(Z_t) = E[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t))\mathbf{X}_t\mathbf{g}_\tau(z_0)^T\boldsymbol{\Pi}_t|Z_t]$ . Again, let  $A_m = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq M\}$  and  $B_m = \{\boldsymbol{\vartheta}_0 : \|\boldsymbol{\vartheta}_0\| \leq \mathfrak{L}\}$  for some  $0 < M < \infty$  and for some  $0 < \mathfrak{L} < \infty$ . Therefore,

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{t=1}^n \rho_\tau(v_t^*(\hat{\boldsymbol{\vartheta}}_0) - b_n\boldsymbol{\theta}^T\mathbf{X}_t^*(\hat{\boldsymbol{\vartheta}}_0))K(z_{th}) \equiv \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}).$$

Now, define vector functions of  $\boldsymbol{\theta}$  and  $\boldsymbol{\vartheta}_0$

$$V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) = b_n \sum_{t=1}^n \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0) - b_n\boldsymbol{\theta}^T\mathbf{X}_t^*(\boldsymbol{\vartheta}_0))\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th}),$$

and

$$V_n(\boldsymbol{\vartheta}_0) = b_n \sum_{t=1}^n \Gamma^*(Z_t)[a_n\boldsymbol{\vartheta}_0]K(z_{th}),$$

where  $\psi_\tau(x) = \tau - I(x < 0)$ . In the next three lemmas, we show that  $V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0)$  and  $V_n(\boldsymbol{\vartheta}_0)$  satisfy Lemma A.7, so that we can derive the local Bahadur representation for  $\hat{\boldsymbol{\theta}}$ .

**Lemma A.8.** *Under the assumptions in Theorem 1, one has*

$$\sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)]\| = o_p(1).$$

*Proof.* For any  $\boldsymbol{\theta} \in A_m$  and for any  $\boldsymbol{\vartheta}_0 \in B_m$ , we have

$$\begin{aligned} & V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0) \\ &= b_n \sum_{t=1}^n [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0) - b_n\boldsymbol{\theta}^T\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))]\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th}) \\ & \quad + b_n \sum_{t=1}^n [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))](\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*)K(z_{th}) \end{aligned}$$

$$\begin{aligned}
& + b_n \sum_{t=1}^n [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))] \mathbf{X}_t^* K(z_{th}) + b_n \sum_{t=1}^n \Gamma^*(Z_t) [a_n \boldsymbol{\vartheta}_0] K(z_{th}) \\
& = b_n \sum_{t=1}^n V_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) + b_n \sum_{t=1}^n U_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) + b_n \sum_{t=1}^n W_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) + b_n \sum_{t=1}^n R_{nt}(\boldsymbol{\vartheta}_0),
\end{aligned}$$

where  $V_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) = \left( V_{nt}^{(1)T}, V_{nt}^{(2)T} \right)^T$ ,  $U_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) K(z_{th}) = \left( U_{nt}^{(1)T}, U_{nt}^{(2)T} \right)^T$ ,  $W_{nt}(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))] \mathbf{X}_t^* \times K(z_{th}) = \left( W_{nt}^{(1)T}, W_{nt}^{(2)T} \right)^T$ , and  $R_{nt}(\boldsymbol{\vartheta}_0) = a_n \Gamma^*(Z_t) \boldsymbol{\vartheta}_0 K(z_{th}) = \left( R_{nt}^{(1)T}, R_{nt}^{(2)T} \right)^T$  with  $V_{nt}^{(1)} = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] \mathbf{X}_t(\boldsymbol{\vartheta}_0) K(z_{th})$ ,  $V_{nt}^{(2)} = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] \mathbf{X}_t(\boldsymbol{\vartheta}_0) z_{th} K(z_{th})$ ,  $U_{nt}^{(1)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] (\mathbf{X}_t(\boldsymbol{\vartheta}_0) - \mathbf{X}_t) K(z_{th})$ , and  $U_{nt}^{(2)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] (\mathbf{X}_t(\boldsymbol{\vartheta}_0) - \mathbf{X}_t) z_{th} K(z_{th})$ . In addition,  $W_{nt}^{(1)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))] \mathbf{X}_t K(z_{th})$ ,  $W_{nt}^{(2)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))] \mathbf{X}_t z_{th} K(z_{th})$ ,  $R_{nt}^{(1)} = a_n \Gamma(Z_t) \boldsymbol{\vartheta}_0 K(z_{th})$  and  $R_{nt}^{(2)} = a_n \Gamma(Z_t) \boldsymbol{\vartheta}_0 z_{th} K(z_{th})$ . Thus,

$$\begin{aligned}
& \|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)]\| \\
& = \left\| b_n \begin{pmatrix} \sum_{t=1}^n (V_{nt}^{(1)} - EV_{nt}^{(1)}) \\ \sum_{t=1}^n (V_{nt}^{(2)} - EV_{nt}^{(2)}) \end{pmatrix} \right\| + \left\| b_n \begin{pmatrix} \sum_{t=1}^n (U_{nt}^{(1)} - EU_{nt}^{(1)}) \\ \sum_{t=1}^n (U_{nt}^{(2)} - EU_{nt}^{(2)}) \end{pmatrix} \right\| \\
& \quad + \left\| b_n \begin{pmatrix} \sum_{t=1}^n (W_{nt}^{(1)} - EW_{nt}^{(1)}) \\ \sum_{t=1}^n (W_{nt}^{(2)} - EW_{nt}^{(2)}) \end{pmatrix} \right\| + \left\| b_n \begin{pmatrix} \sum_{t=1}^n (R_{nt}^{(1)} - ER_{nt}^{(1)}) \\ \sum_{t=1}^n (R_{nt}^{(2)} - ER_{nt}^{(2)}) \end{pmatrix} \right\| \\
& \leq b_n \left\| \sum_{t=1}^n (V_{nt}^{(1)} - EV_{nt}^{(1)}) \right\| + b_n \left\| \sum_{t=1}^n (V_{nt}^{(2)} - EV_{nt}^{(2)}) \right\| \\
& \quad + b_n \left\| \sum_{t=1}^n (U_{nt}^{(1)} - EU_{nt}^{(1)}) \right\| + b_n \left\| \sum_{t=1}^n (U_{nt}^{(2)} - EU_{nt}^{(2)}) \right\| \\
& \quad + b_n \left\| \sum_{t=1}^n (W_{nt}^{(1)} - EW_{nt}^{(1)}) \right\| + b_n \left\| \sum_{t=1}^n (W_{nt}^{(2)} - EW_{nt}^{(2)}) \right\| \\
& \quad + b_n \left\| \sum_{t=1}^n (R_{nt}^{(1)} - ER_{nt}^{(1)}) \right\| + b_n \left\| \sum_{t=1}^n (R_{nt}^{(2)} - ER_{nt}^{(2)}) \right\| \\
& \equiv V_n^{(1)} + V_n^{(2)} + U_n^{(1)} + U_n^{(2)} + W_n^{(1)} + W_n^{(2)} + R_n^{(1)} + R_n^{(2)}.
\end{aligned}$$

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

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

where  $V_{nt}^{(1i)} = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))]X_{it}(\boldsymbol{\vartheta}_0)K(z_{th})$ , 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,

$$\begin{aligned}
E(V_n^{(1i)})^2 &= \frac{1}{nh} E \left\{ \sum_{t=1}^n (V_{nt}^{(1i)} - EV_{nt}^{(1i)}) \right\}^2 \\
&= \frac{1}{nh} \left[ \sum_{t=1}^n \text{Var}(V_{nt}^{(1i)}) + 2 \sum_{\ell=1}^{n-1} \left(1 - \frac{\ell}{n}\right) \text{Cov}(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)}) \right] \\
&\leq \frac{1}{h} \text{Var}(V_{n1}^{(1i)}) + \frac{2}{h} \sum_{\ell=1}^{d_n-1} |\text{Cov}(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)})| + \frac{2}{h} \sum_{\ell=d_n}^{\infty} |\text{Cov}(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)})| \\
&\equiv J_7 + J_8 + J_9
\end{aligned}$$

with  $d_n \rightarrow \infty$  specified later. First, we consider the last term,  $J_9$ , in the above equation. To this end, using Davydov's inequality (see, e.g., Corollary A.2 of Hall and Heyde (1980)), one has

$$|\text{Cov}(V_{n1}^{(1i)}, V_{n(\ell+1)}^{(1i)})| \leq C\alpha^{1-2/\delta}(\ell)[E|V_{n1}^{(1i)}|^\delta]^{2/\delta}. \quad (\text{A.9})$$

Notice that for any  $k > 0$ ,  $|\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))|^k = I(r_{3t} < Y_t \leq r_{4t})$ , where  $r_{3t} = \min(p_{2t}, p_{2t} + p_{3t})$  and  $r_{4t} = \max(p_{2t}, p_{2t} + p_{3t})$  with  $p_{2t} = [\mathbf{g}_\tau(z_0) + \mathbf{g}_\tau^{(1)}(z_0)(Z_t - z_0)]^T \mathbf{X}_t(\boldsymbol{\vartheta}_0)$  and  $p_{3t} = \frac{1}{\sqrt{nh}} \boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)$ . Therefore, by Assumption A4, there exists a  $C > 0$  such that

$$E\{|\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))|^k | Z_t, \mathbf{X}_t\} = F_{Y|Z, \mathbf{X}}(r_{4t}) - F_{Y|Z, \mathbf{X}}(r_{3t}) \leq Cb_n |\boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)|,$$

which implies by Assumption A9 that

$$\begin{aligned}
E|V_{n1}^{(1i)}|^\delta &= E[|\psi_\tau(v_{n1}^*) - \psi_\tau(v_1^*(\boldsymbol{\vartheta}_0))|^\delta | X_{i1}(\boldsymbol{\vartheta}_0)|^\delta K^\delta(z_{1h})] \\
&\leq Cb_n E[|\boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)| | X_{i1}(\boldsymbol{\vartheta}_0)|^\delta K^\delta(z_{1h})].
\end{aligned}$$

Notice that since  $\|\boldsymbol{\vartheta}_0\| \leq \mathfrak{L}$ , by mean value theorem and triangle inequality, one can choose a sufficiently large  $C > 0$ , such that  $\|\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)\| \leq C\|\mathbf{X}_t^*\|$ . Then,

$$\begin{aligned}
E|V_{n1}^{(1i)}|^\delta &= E[|\psi_\tau(v_{n1}^*) - \psi_\tau(v_1^*(\boldsymbol{\vartheta}_0))|^\delta | X_{i1}(\boldsymbol{\vartheta}_0)|^\delta K^\delta(z_{1h})] \\
&\leq Cb_n E[|\boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)| | X_{i1}(\boldsymbol{\vartheta}_0)|^\delta K^\delta(z_{1h})] \leq Cb_n E[|\boldsymbol{\theta}^T \mathbf{X}_t^*| | X_{i1}|^\delta K^\delta(z_{1h})] \leq Cb_n h.
\end{aligned}$$

This, in conjunction with (A.9), gives that



$$J_9 \leq C b_n^{2/\delta} h^{2/\delta-1} \sum_{\ell=d_n}^{\infty} \alpha^{1-2/\delta}(\ell) \leq C b_n^{2/\delta} h^{2/\delta-1} d_n^{-w} \sum_{\ell=d_n}^{\infty} \ell^w \alpha^{1-2/\delta}(\ell) = o(b_n^{2/\delta} h^{2/\delta-1} d_n^{-w}) = o(1).$$

As for  $J_8$ , again by choosing sufficiently large  $C > 0$ , we use Assumptions A4 and A11 to obtain

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

It follows that  $J_8 = o(1)$  by  $d_n h \rightarrow 0$ . Analogously,

$$J_7 = h^{-1} Var(V_{n1}^{(1i)}) \leq h^{-1} E(V_{n1}^{(1i)})^2 = O(b_n).$$

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

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

where  $U_{nt}^{(1i)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))](X_{ti}(\boldsymbol{\vartheta}_0) - X_{ti})K(z_{th})$  and  $U_n^{(1i)} = b_n \sum_{t=1}^n (U_{nt}^{(1i)} - EU_{nt}^{(1i)})$ . By mean value theorem, there exists  $\boldsymbol{\vartheta}'_0 \in (0, \boldsymbol{\vartheta}_0)$ , such that

$$\begin{aligned} E|U_{n1}^{(1i)}|^\delta &= E[|\psi_\tau(v_1^*(\boldsymbol{\vartheta}_0))|^\delta |X_{1i}(\boldsymbol{\vartheta}_0) - X_{1i}|^\delta K^\delta(z_{1h})] \\ &\leq CE[|X_{1i}(\boldsymbol{\vartheta}_0) - X_{1i}|^\delta K^\delta(z_{1h})] \leq CE \left[ \left| \left( \frac{\partial X_{1i}(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_0} \right) \Big|_{\boldsymbol{\vartheta}_0 = \boldsymbol{\vartheta}'_0} \boldsymbol{\vartheta}_0 \right|^\delta K^\delta(z_{1h}) \right] \leq C a_n^\delta h \end{aligned}$$

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  $k > 0$ ,

$$|\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))|^k = I(c_{3t} < Y_t \leq c_{4t}), \text{ where } c_{3t} = \min(d_{2t}, d_{3t}) \text{ and } c_{4t} = \max(d_{2t}, d_{3t})$$

with  $d_{2t} = [\mathbf{g}_\tau(z_0) + \mathbf{g}_\tau^{(1)}(z_0)(Z_t - z_0)]^T \mathbf{X}_t(\boldsymbol{\vartheta}_0)$  and  $d_{3t} = [\mathbf{g}_\tau(z_0) + \mathbf{g}_\tau^{(1)}(z_0)(Z_t - z_0)]^T \mathbf{X}_t$ . Therefore,

by Assumption A4, there exists a  $C > 0$  such that

$$E\{|\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))|^k | Z_t, \mathbf{X}_t\} = F_{Y|Z, \mathbf{X}}(c_{4t}) - F_{Y|Z, \mathbf{X}}(c_{3t}) \leq C \left| \left( \frac{\partial X_{1i}(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_0} \right) \Big|_{\boldsymbol{\vartheta}_0 = \boldsymbol{\vartheta}'_0} \boldsymbol{\vartheta}_0 \right|,$$

which implies by Assumption A9 that

$$E|W_{n1}^{(1i)}|^\delta = E[|\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))|^\delta |X_{i1}|^\delta K^\delta(z_{1h})] \leq Ca_n^\delta h.$$

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  $\boldsymbol{\vartheta}_0 \in B_m$ ,

$$\|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)]\| = o_p(1). \quad (\text{A.10})$$

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

$$\sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)]\| = o_p(1).$$

Now, we decompose  $A_m$  and  $B_m$  into cubes, respectively, based on the grid  $(j_1 \hbar M, \dots, j_{2(2\kappa+1)} \hbar M)$  and  $(i_1 \mathbb{k} \mathfrak{L}, \dots, i_{2(2\kappa+1)} \mathbb{k} \mathfrak{L})$ , where  $j_k = 0, \pm 1, \dots, \pm[1/\hbar] + 1$ ,  $i_k = 0, \pm 1, \dots, \pm[1/\mathbb{k}] + 1$ ,  $[\cdot]$  denotes taking integer part of  $\cdot$ , and  $\hbar$  and  $\mathbb{k}$  are fixed positive small numbers. Denote  $D(\boldsymbol{\theta})$  and  $D(\boldsymbol{\vartheta}_0)$  the lower vertex of cubes that contain  $\boldsymbol{\theta}$  and  $\boldsymbol{\vartheta}_0$ , respectively. Then,

$$\begin{aligned} & \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)]\| \\ & \leq \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(D(\boldsymbol{\theta}), 0) - V_n(0, 0) - E[V_n(D(\boldsymbol{\theta}), 0) - V_n(0, 0)]\| \\ & \quad + \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), 0) - E[V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), 0)]\| \\ & \quad + \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0) - E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)]\| \\ & \quad + \sup_{\boldsymbol{\vartheta}_0 \in B_m} \|V_n(\boldsymbol{\vartheta}_0) - E[V_n(\boldsymbol{\vartheta}_0)]\| \\ & \equiv H_1 + H_2 + H_3 + H_4. \end{aligned}$$

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  $\mathbf{X}_t \equiv \mathbf{X}_t(0)$ , it follows easily from

(A.10) that

$$H_1 \equiv \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(D(\boldsymbol{\theta}), 0) - V_n(0, 0) - E[V_n(D(\boldsymbol{\theta}), 0) - V_n(0, 0)]\| = o_p(1).$$

As for the first term of  $H_3$ , notice that

$$\begin{aligned}
& \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)\| \\
&= b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [\psi_\tau(v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) \right\| \\
&\leq b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0) < 0) - I(v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0)) < 0)] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) \right\| \\
&\quad + b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I(v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0)) < 0) - I(v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) < 0)] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) \right\| \\
&\leq 2b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) \right\| \\
&\leq 2b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0))) K(z_{th}) \right\| \\
&\quad + 2b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] \mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0)) K(z_{th}) \right\| \\
&\leq 2b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] (\mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0)) + \mathfrak{L}) K(z_{th}) \right\| \tag{A.11} \\
&\leq 2Cb_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] \right. \\
&\quad \times \mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0)) K(z_{th}) \left. \right\| \\
&\quad + 2Cb_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] \mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0)) K(z_{th}) \right\| \\
&\leq 2Cb_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] \right. \\
&\quad \times \mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0)) K(z_{th}) \left. \right\| + (2C/h) \max\{\hbar, \mathbb{k}\} \|E[\mathbf{X}_t^* K(z_{th})]\| \\
&\leq 2Cb_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\hbar, \mathbb{k}\}}{\sqrt{nh}}\}}] \right. \\
&\quad \times \mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0)) K(z_{th}) \left. \right\| + 2C \max\{\hbar, \mathbb{k}\},
\end{aligned}$$

where the fourth inequality follows from the Lipschitz continuity. Since the number of the elements in  $\{D(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| \leq M\}$  and  $\{D(\boldsymbol{\vartheta}_0) : \|\boldsymbol{\vartheta}_0\| \leq \mathfrak{L}\}$  are finite, one can easily show that

$$2Cb_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\bar{h}, \bar{k}\}}{\sqrt{nh}}\}}] - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\bar{h}, \bar{k}\}}{\sqrt{nh}}\}}] \right\} \\ \times \mathbf{X}_t^*(D(\boldsymbol{\vartheta}_0))K(z_{th})\| = o_p(1)$$

by following the same steps as in (A.10). Let  $\max\{\bar{h}, \bar{k}\} \rightarrow 0$ . 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 (A.11),

$$\sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)]\| \\ = b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \right\| \\ \leq 2nb_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|E[I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C \max\{\bar{h}, \bar{k}\}}{\sqrt{nh}}\}}] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\| \leq C \max\{\bar{h}, \bar{k}\}.$$

When  $\max\{\bar{h}, \bar{k}\} \rightarrow 0$ , one has

$$\sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)]\| = o(1).$$

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

$$\sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), 0)\| \\ = b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)) \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0)) \mathbf{X}_t^*] K(z_{th}) \right\| \\ \leq b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0)))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) \right\| \\ + b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) \right\| \\ + b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) K(z_{th}) \right\| \equiv H_{21} + H_{22} + H_{23}.$$

It is easy to see that by following the same deduction as in (A.11), one can derive  $H_{21} = o_p(1)$

and  $H_{22} = o_p(1)$ . Also, notice that for  $H_{23}$ , by mean value theorem,

$$H_{23} \equiv b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) K(z_{th}) \right\| \\ \leq Ca_n b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n [\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))] K(z_{th}) \right\|,$$

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  $H_2$ ,

$$\begin{aligned}
& \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|E\{V_n(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0) - V_n(D(\boldsymbol{\theta}), 0)\}\| \\
&= b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n E[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)) \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0)) \mathbf{X}_t^*] K(z_{th}) \right\| \\
&\leq b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n E[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) \right\| \\
&\quad + b_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| \sum_{t=1}^n E[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) K(z_{th}) \right\| \equiv H'_{21} + H'_{22}.
\end{aligned}$$

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

$$\begin{aligned}
H'_{22} &\equiv \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(D(\boldsymbol{\theta}), 0))] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) K(z_{th})\} \right\| \\
&= \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| b_n \sum_{t=1}^n E\{[\tau - F_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t \right. \\
&\quad \left. + b_n D(\boldsymbol{\theta})^T \mathbf{X}_t^* | Z_t, \mathbf{X}_t)] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) K(z_{th})\} \right\| \\
&= \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| b_n \sum_{t=1}^n E\{[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t \right. \\
&\quad \left. + \Im \Pi(h, z_0, Z_t, \mathbf{X}_t) | Z_t, \mathbf{X}_t)] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) \right\} \\
&\quad \times \Pi(h, z_0, Z_t, \mathbf{X}_t) K(z_{th}) \left\|,
\end{aligned}$$

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

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

Therefore, it results in that by mean value theorem, there exists  $\boldsymbol{\vartheta}'_0 \in (0, \boldsymbol{\vartheta}_0)$ , such that

$$\begin{aligned}
& \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| b_n \sum_{t=1}^n E\{[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t \right. \\
&\quad \left. + \Im \Pi(h, z_0, Z_t, \mathbf{X}_t) | Z_t, \mathbf{X}_t)] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) \right\} \Pi(h, z_0, Z_t, \mathbf{X}_t) K(z_{th}) \left\|
\end{aligned}$$

$$\begin{aligned}
&\leq \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \left\| b_n \sum_{t=1}^n E\{[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t \right. \\
&\quad \left. + \Im \Pi(h, z_0, Z_t, \mathbf{X}_t) | Z_t, \mathbf{X}_t] \left( \frac{\partial \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_0} \Big|_{\boldsymbol{\vartheta}_0 = \boldsymbol{\vartheta}'_0} \right) \boldsymbol{\vartheta}_0 \right\} \\
&\quad \times \Pi(h, z_0, Z_t, \mathbf{X}_t) K(z_{th}) \| = o(1).
\end{aligned}$$

In the same way as in analyzing (A.11), it can be easily shown that  $H'_{21} = o_p(1)$ . So,  $H_2 = o_p(1)$ .

The proof of Lemma A.8 is completed.  $\square$

**Lemma A.9.** *Under the assumptions in Theorem 1, one has*

$$\sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \|E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)] + f_z(z_0) \Omega_1^*(z_0) \boldsymbol{\theta}\| = o(1),$$

where  $\Omega_1^*(z_0) = \text{diag}\{\Omega^*(z_0), \mu_2 \Omega^*(z_0)\}$ .

*Proof.* Notice that

$$E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)] = E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(\boldsymbol{\theta}, 0) + V_n(\boldsymbol{\vartheta}_0)] + E[V_n(\boldsymbol{\theta}, 0) - V_n(0, 0)] \equiv R_1 + R_2.$$

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,

$$\begin{aligned}
R_1 &\equiv b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0)) \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \psi_\tau(v_{nt}^*(\boldsymbol{\theta}, 0)) \mathbf{X}_t^*] K(z_{th})\} + E[V_n(\boldsymbol{\vartheta}_0)] \\
&= b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(0, \boldsymbol{\vartheta}_0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th})\} \\
&\quad + b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(0, 0)) - \psi_\tau(v_{nt}^*(\boldsymbol{\theta}, 0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th})\} \\
&\quad + b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(0, \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(0, 0))] \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th})\} \\
&\quad + b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta}, 0))] (\mathbf{X}_t^*(\boldsymbol{\vartheta}_0) - \mathbf{X}_t^*) K(z_{th})\} + b_n \sum_{t=1}^n E\{\Gamma^*(Z_t) a_n \boldsymbol{\vartheta}_0 K(z_{th})\} \\
&\equiv R_{11} + R_{12} + R_{13} + R_{14} + R_{15}.
\end{aligned}$$

Here,  $R_{14}$  can be vanished in the same way as that in proving Lemma A.8. We first consider  $R_{11}$  as follows

$$\begin{aligned}
R_{11} &\equiv b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(0, \boldsymbol{\vartheta}_0))]\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \\
&= b_n \sum_{t=1}^n E\{[F_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t(\boldsymbol{\vartheta}_0)) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t(\boldsymbol{\vartheta}_0)|Z_t, \mathbf{X}_t) \\
&\quad - F_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t(\boldsymbol{\vartheta}_0)) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t(\boldsymbol{\vartheta}_0) \\
&\quad + b_n \boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)|Z_t, \mathbf{X}_t)]\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \\
&= -\frac{1}{h} E\{[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t(\boldsymbol{\vartheta}_0)) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t(\boldsymbol{\vartheta}_0) \\
&\quad + \partial b_n \boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)|Z_t, \mathbf{X}_t)]\boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0) \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \\
&= -\frac{1}{h} E\{[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t(\boldsymbol{\vartheta}_0))|Z_t, \mathbf{X}_t)]\boldsymbol{\theta}^T \mathbf{X}_t^* \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} + o(1).
\end{aligned}$$

In the same way, one can easily show by Assumption A4 that

$$\begin{aligned}
R_{11} + R_{12} &= \frac{1}{h} E\{[f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t)|Z_t, \mathbf{X}_t) - f_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t(\boldsymbol{\vartheta}_0))|Z_t, \mathbf{X}_t)] \\
&\quad \times \boldsymbol{\theta}^T \mathbf{X}_t^* \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} + o(1) \\
&\leq C \frac{1}{h} E\{\mathbf{g}_\tau(z_0)^T (\mathbf{X}_t - \mathbf{X}_t(\boldsymbol{\vartheta}_0)) \boldsymbol{\theta}^T \mathbf{X}_t^* \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} + o(1) = o(1).
\end{aligned}$$

As for  $R_{13}$  and  $R_{15}$ , by applying mean value theorem, there exists  $\boldsymbol{\vartheta}'_0 \in (0, \boldsymbol{\vartheta}_0)$  such that

$$\begin{aligned}
R_{13} &\equiv b_n \sum_{t=1}^n E\{[\psi_\tau(v_{nt}^*(0, \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(0, 0))]\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \\
&= b_n \sum_{t=1}^n E\{[F_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t|Z_t, \mathbf{X}_t) \\
&\quad - F_{Y|Z, \mathbf{X}}(q_\tau(z_0, \mathbf{X}_t(\boldsymbol{\vartheta}_0)) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)^T \mathbf{X}_t(\boldsymbol{\vartheta}_0)|Z_t, \mathbf{X}_t)]\mathbf{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \\
&= -b_n \sum_{t=1}^n E\{[f_{Y|Z, \mathbf{X}}(\tilde{\mathbf{X}}_t^T(\mathbf{g}_\tau(z_0) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0))|Z_t, \mathbf{X}_t)] \\
&\quad \times \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)(\mathbf{X}_t(\boldsymbol{\vartheta}_0) - \mathbf{X}_t)^T [\mathbf{g}_\tau(z_0) + h z_{th} \mathbf{g}_\tau^{(1)}(z_0)]K(z_{th})\} \\
&= -b_n \sum_{t=1}^n E\{\Gamma^*(Z_t) a_n \boldsymbol{\vartheta}_0 K(z_{th})\} + o(h)
\end{aligned}$$

by some simple calculations, where  $\tilde{\mathbf{X}}_t \equiv \mathbf{X}_t + C a_n \boldsymbol{\vartheta}_0$ . This implies that  $R_{13} + R_{15} = o(1)$ .

Thus, one has

$$\|E[V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) - V_n(0, 0) + V_n(\boldsymbol{\vartheta}_0)] + f_z(z_0)\Omega_1^*(z_0)\boldsymbol{\theta}\| = o(1). \quad (\text{A.12})$$

Similar to the proof of Lemma A.3 in Xu (2005), one can prove that (A.12) holds uniformly in  $A_m$  and  $B_m$  with the details omitted. These complete the proof of Lemma A.9.  $\square$

**Lemma A.10.** *Let  $B_t = [\psi_\tau(v_t^*(0))\mathbf{X}_t^* - \psi_\tau(Y_t^*)\Gamma^*(Z_t)(\mathbf{D}^*(Z_t))^{-1}\mathbf{W}_t]K(z_{th_2})$ . Then, under the assumptions in Theorem 1, one has*

$$E[B_1] = \frac{h_2^3 f_z(z_0)}{2} \begin{pmatrix} \mu_2 \Omega^*(z_0) \mathbf{g}_\tau^{(2)}(z_0) \\ 0 \end{pmatrix} + o(h_2^3),$$

and

$$\text{Var}[B_1] = h_2 \tau(1 - \tau) f_z(z_0) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \left\{ \Omega(z_0) - H_1(z_0) + H_2(z_0) \right\} + o(h_2),$$

where  $H_1(z_0) = E[\mathbf{X}_1 \mathbf{W}_1^T | Z_1 = z_0](\mathbf{D}^*(z_0))^{-1} \Gamma^T(z_0) + \Gamma(z_0)(\mathbf{D}^*(z_0))^{-1} E[\mathbf{W}_1 \mathbf{X}_1^T | Z_1 = z_0]$  and

$H_2(z_0) = \Gamma(z_0)(\mathbf{D}^*(z_0))^{-1} \mathbf{D}(z_0)(\mathbf{D}^*(z_0))^{-1} \Gamma^T(z_0)$ . Then,

$$\text{Var}\left\{ \frac{1}{\sqrt{nh_2}} \sum_{t=1}^n B_t \right\} = \tau(1 - \tau) f_z(z_0) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \left\{ \Omega(z_0) - H_1(z_0) + H_2(z_0) \right\} + o(1).$$

*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

$$\begin{aligned} E[B_1] &= E\{[\psi_\tau(v_1^*(0))\mathbf{X}_1^* - \psi_\tau(Y_1^*)\Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1]K(z_{1h_2})\} \\ &= E\{\psi_\tau(v_1^*(0))\mathbf{X}_1^*K(z_{1h_2})\} - E\{\psi_\tau(Y_1^*)\Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1K(z_{1h_2})\} \equiv Q_1 + Q_2. \end{aligned}$$

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

$$Q_1 = \frac{h_2^3}{2} f_z(z_0) \left\{ \begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} \otimes \Omega^*(z_0) \right\} \mathbf{g}_\tau^{(2)}(z_0) + o(h_2^3) \quad (\text{A.13})$$

with the detail omitted. For  $Q_2$ , similar to the derivation in (A.7) and by Assumption A10,

$$Q_2 \equiv -E\{\psi_\tau(Y_1^*)\Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1K(z_{1h_2})\} = O(h_1^2 h_2) = o(h_2^3).$$

As for  $E[B_1 B_1^T]$ , we have



$$\begin{aligned}
E[B_1 B_1^T] &= E \left( \left\{ \psi_\tau^2(v_1^*(0)) \mathbf{X}_1^* \mathbf{X}_1^{*T} - [\psi_\tau(v_1^*(0)) \psi_\tau(Y_1^*) \mathbf{X}_1^* \mathbf{W}_1^T (\mathbf{D}^*(Z_1))^{-1} \Gamma^{*T}(Z_1) \right. \right. \\
&\quad + \psi_\tau(v_1^*(0)) \psi_\tau(Y_1^*) \Gamma^*(Z_1) (\mathbf{D}^*(Z_1))^{-1} \mathbf{W}_1 \mathbf{X}_1^{*T}] \\
&\quad \left. \left. + \psi_\tau^2(Y_1^*) \Gamma^*(Z_1) (\mathbf{D}^*(Z_1))^{-1} \mathbf{W}_1 \mathbf{W}_1^T (\mathbf{D}^*(Z_1))^{-1} \Gamma^{*T}(Z_1) \right\} K^2(z_{1h_2}) \right) \\
&= E \{ \psi_\tau^2(v_1^*(0)) \mathbf{X}_1^* \mathbf{X}_1^{*T} K^2(z_{1h_2}) \} \\
&\quad - E \{ [\psi_\tau(v_1^*(0)) \psi_\tau(Y_1^*) \mathbf{X}_1^* \mathbf{W}_1^T (\mathbf{D}^*(Z_1))^{-1} \Gamma^{*T}(Z_1) \\
&\quad + \psi_\tau(v_1^*(0)) \psi_\tau(Y_1^*) \Gamma^*(Z_1) (\mathbf{D}^*(Z_1))^{-1} \mathbf{W}_1 \mathbf{X}_1^{*T}] K^2(z_{1h_2}) \} \\
&\quad + E \{ \psi_\tau^2(Y_1^*) \Gamma^*(Z_1) (\mathbf{D}^*(Z_1))^{-1} \mathbf{W}_1 \mathbf{W}_1^T (\mathbf{D}^*(Z_1))^{-1} \Gamma^{*T}(Z_1) K^2(z_{1h_2}) \} \\
&\equiv P^{(1)} + P^{(2)} + P^{(3)}.
\end{aligned}$$

For  $P^{(1)}$ , similar to the derivation in Lemma A.5, one has

$$P^{(1)} \equiv \tau(1-\tau) E \{ \mathbf{X}_1^* \mathbf{X}_1^{*T} K^2(z_{1h_2}) \} + o(h_2^2) = h_2 \tau(1-\tau) f_z(z_0) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \Omega(z_0)(1+o(1)) + o(h_2^2). \quad (\text{A.14})$$

Similarly,

$$\begin{aligned}
P^{(3)} &\equiv E[\psi_\tau^2(Y_1^*) \Gamma^*(Z_1) (\mathbf{D}^*(Z_1))^{-1} \mathbf{W}_1 \mathbf{W}_1^T (\mathbf{D}^*(Z_1))^{-1} \Gamma^{*T}(Z_1) K^2(z_{1h_2})] \\
&= \tau(1-\tau) E \{ \Gamma^*(Z_1) (\mathbf{D}^*(Z_1))^{-1} \mathbf{W}_1 \mathbf{W}_1^T (\mathbf{D}^*(Z_1))^{-1} \Gamma^{*T}(Z_1) K^2(z_{1h_2}) \} + o(h_2^2) \\
&= \tau(1-\tau) E \{ \Gamma^*(Z_1) (\mathbf{D}^*(Z_1))^{-1} E[\mathbf{W}_1 \mathbf{W}_1^T | Z_1] (\mathbf{D}^*(Z_1))^{-1} \Gamma^{*T}(Z_1) K^2(z_{1h_2}) \} + o(h_2^2) \\
&= h_2 \tau(1-\tau) f_z(z_0) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \left\{ \Gamma(z_0) (\mathbf{D}^*(z_0))^{-1} \mathbf{D}(z_0) (\mathbf{D}^*(z_0))^{-1} \Gamma^T(z_0) \right\} (1+o(1)) + o(h_2^2) \\
&= h_2 \tau(1-\tau) f_z(z_0) \left\{ \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes H_2(z_0) \right\} (1+o(1)) + o(h_2^2). \quad (\text{A.15})
\end{aligned}$$

As for  $P^{(2)}$ , by Assumption A10, one has

$$\begin{aligned}
P^{(2)} &\equiv -E\{\psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)[\mathbf{X}_1^*\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1) \\
&\quad + \Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^{*T}]K^2(z_{1h_2})\} \\
&= -E\{[\tau - I_{\{v_1^*(0)<0\}}][\tau - I_{\{Y_1^*<0\}}][\mathbf{X}_1^*\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1) \\
&\quad + \Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^{*T}]K^2(z_{1h_2})\} \\
&= -E\{[\tau^2 - \tau(I_{\{Y_1^*<0\}} + I_{\{v_1^*(0)<0\}}) + I_{\{Y_1^*<0\}}][\mathbf{X}_1^*\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1) \\
&\quad + \Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^{*T}]K^2(z_{1h_2})\} \\
&= -E\{[(\tau - 1)(\tau - I_{\{Y_1^*<0\}}) + \tau(\tau - I_{\{v_1^*(0)<0\}})][\mathbf{X}_1^*\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1) \\
&\quad + \Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^{*T}]K^2(z_{1h_2})\} \\
&\quad - \tau(1 - \tau)E\{[\mathbf{X}_1^*\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1) + \Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^{*T}]K^2(z_{1h_2})\} \\
&\equiv P^{(21)} + P^{(22)}.
\end{aligned}$$

It can be shown that  $P^{(21)} = o(h_2^2)$ , using the same idea in proving Lemma A.5. We now focus on evaluating  $P^{(22)}$ . A simple algebra gives that

$$\begin{aligned}
P^{(22)} &\equiv -\tau(1 - \tau)E\{[\mathbf{X}_1^*\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1) + \Gamma^*(Z_1)(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^{*T}]K^2(z_{1h_2})\} \\
&= -\tau(1 - \tau)E\left\{\begin{pmatrix} \mathbf{X}_1\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1} \\ z_{1h_2}\mathbf{X}_1\mathbf{W}_1^T(\mathbf{D}^*(Z_1))^{-1} \end{pmatrix} \begin{pmatrix} \Gamma^T(Z_1) & z_{1h_2}\Gamma^T(Z_1) \end{pmatrix} K^2(z_{1h_2})\right\} \\
&\quad - \tau(1 - \tau)E\left\{\begin{pmatrix} \Gamma(Z_1) \\ z_{1h_2}\Gamma(Z_1) \end{pmatrix} \begin{pmatrix} (\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^T & z_{1h_2}(\mathbf{D}^*(Z_1))^{-1}\mathbf{W}_1\mathbf{X}_1^T \end{pmatrix} K^2(z_{1h_2})\right\} \\
&= -\tau(1 - \tau)E\left\{\begin{pmatrix} 1 & z_{1h_2} \\ z_{1h_2} & z_{1h_2}^2 \end{pmatrix} \otimes E[\mathbf{X}_1\mathbf{W}_1^T|Z_1](\mathbf{D}^*(Z_1))^{-1}\Gamma^T(Z_1)K^2(z_{1h_2})\right\} \\
&\quad - \tau(1 - \tau)E\left\{\begin{pmatrix} 1 & z_{1h_2} \\ z_{1h_2} & z_{1h_2}^2 \end{pmatrix} \otimes \Gamma(Z_1)(\mathbf{D}^*(Z_1))^{-1}E[\mathbf{W}_1\mathbf{X}_1^T|Z_1]K^2(z_{1h_2})\right\}
\end{aligned}$$

$$\begin{aligned}
&= -h_2\tau(1-\tau)f_z(z_0) \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes \left\{ E[\mathbf{X}_1 \mathbf{W}_1^T | Z_1 = z_0] (\mathbf{D}^*(z_0))^{-1} \Gamma^T(z_0) \right. \\
&\quad \left. + \Gamma(z_0) (\mathbf{D}^*(z_0))^{-1} E[\mathbf{W}_1 \mathbf{X}_1^T | Z_1 = z_0] \right\} (1 + o(1)) \\
&= -h_2\tau(1-\tau)f_z(z_0) \left\{ \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes H_1(z_0) \right\} (1 + o(1)).
\end{aligned}$$

Therefore,

$$P^{(2)} = -h_2\tau(1-\tau)f_z(z_0) \left\{ \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \otimes H_1(z_0) \right\} (1 + o(1)) + o(h_2^2). \quad (\text{A.16})$$

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

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

By (A.13), (A.14), (A.15), (A.16) and Assumption A10,

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

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

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

which gives

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

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 A.10.  $\square$

## A.2 Proof of Theorem 1:

*Proof.* Following Cai and Xu (2008),  $\|V_n(0, 0)\| = O_p(1)$ . Thus, by Lemmas A.8, A.9 and A.10,  $V_n(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0)$  satisfies Condition (ii) in Lemma A.7; that is,  $\|A_n\| = O_p(1)$  and  $\sup_{\|\Delta\| \leq M, \|\boldsymbol{\vartheta}_0\| \leq \mathfrak{C}} \|V_n(\Delta, \boldsymbol{\vartheta}_0) + V_n(\boldsymbol{\vartheta}_0) + D\Delta - A_n\| = o_p(1)$  with  $D = f_z(z_0)\Omega_1^*(z_0)$  and  $A_n = V_n(0, 0)$ . Next, we want to show that  $\|V_n(\hat{\boldsymbol{\vartheta}}_0)\| = O_p(1)$ . Indeed, by Lemma A.6,

$$\begin{aligned}
& E[V_n(\hat{\boldsymbol{\vartheta}}_0)] \\
&= b_n \sum_{t=1}^n E \left\{ \left[ \Gamma^*(Z_t)(\mathbf{D}^*(Z_t)^{-1}) \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n \psi_\tau(Y_s^*) \mathbf{W}_s K(z_{sh_1}) \right] K(z_{th_2}) \right\} \\
&= b_n \sum_{t=1}^n E \left\{ \left[ \Gamma^*(Z_t)(\mathbf{D}^*(Z_t)^{-1}) \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n \{ \psi_\tau(Y_t^*) \mathbf{W}_t + \psi_\tau(Y_s^*) \mathbf{W}_s \right. \right. \\
&\quad \left. \left. - \psi_\tau(Y_t^*) \mathbf{W}_t \} K(z_{sh_1}) \right] K(z_{th_2}) \right\} \\
&= b_n \sum_{t=1}^n E \left\{ \left[ \psi_\tau(Y_t^*) \Gamma^*(Z_t)(\mathbf{D}^*(Z_t)^{-1}) \mathbf{W}_t \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n K(z_{sh_1}) \right] K(z_{th_2}) \right\} \\
&\quad + b_n \sum_{t=1}^n E \left\{ \left[ \Gamma^*(Z_t)(\mathbf{D}^*(Z_t)^{-1}) \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n \{ \psi_\tau(Y_s^*) \mathbf{W}_s - \psi_\tau(Y_t^*) \mathbf{W}_t \} \right. \right. \\
&\quad \left. \left. \times K(z_{sh_1}) \right] K(z_{th_2}) \right\} \equiv T^{(1)} + T^{(2)}.
\end{aligned}$$

For  $T^{(1)}$ , using the technique in deriving (A.7), one has

$$\begin{aligned}
T^{(1)} &\equiv b_n \sum_{t=1}^n E \left\{ \left[ \psi_\tau(Y_t^*) \Gamma^*(Z_t)(\mathbf{D}^*(Z_t)^{-1}) \mathbf{W}_t \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n K(z_{sh_1}) \right] K(z_{th_2}) \right\} \\
&= b_n \sum_{t=1}^n E \{ [\psi_\tau(Y_t^*) \Gamma^*(Z_t)(\mathbf{D}^*(Z_t)^{-1}) \mathbf{W}_t] K(z_{th_2}) \} + o(1) \\
&= O((nh_2)^{1/2} h_1^2) + o(1) = o(1),
\end{aligned}$$

by the fact that  $f_z^{-1}(Z_t)(nh_1)^{-1} \sum_{s=m+1 \neq t}^n K(z_{sh_1}) = 1 + o(1)$  and by Assumption A10. As for  $T^{(2)}$ , it is not hard to show that  $T^{(2)} = o(1)$ . Thus,  $E[V_n(\hat{\boldsymbol{\vartheta}}_0)] = o(1)$ . In addition, similar to the proof of Lemma A.8, one can obtain that  $\text{Var}[V_n(\hat{\boldsymbol{\vartheta}}_0)] = o(1)$ . Therefore,  $\|V_n(\hat{\boldsymbol{\vartheta}}_0)\| = O_p(1)$ . To show  $\|V_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}_0)\| = o_p(1)$ , it follows from Lemma A.1 and mean value theorem that

$$\begin{aligned} \|V_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}_0)\| &= b_n \left\| \sum_{t=1}^n [\psi_\tau(v_t^*(\hat{\boldsymbol{\vartheta}}_0) - b_n \hat{\boldsymbol{\theta}}^T \mathbf{X}_t^*(\hat{\boldsymbol{\vartheta}}_0))] \mathbf{X}_t^*(\hat{\boldsymbol{\vartheta}}_0) K(z_{th_2}) \right\| \leq b_n \max_{1 \leq t \leq n} \|\mathbf{X}_t^*(\hat{\boldsymbol{\vartheta}}_0) K(z_{th_2})\| \\ &\leq b_n \max_{1 \leq t \leq n} \|\mathbf{X}_t^* K(z_{th_2})\| + C b_n \max_{1 \leq t \leq n} \left\| \left( \frac{\partial \mathbf{X}_t^*(\hat{\boldsymbol{\vartheta}}_0)}{\partial \hat{\boldsymbol{\vartheta}}_0} \Big|_{\hat{\boldsymbol{\vartheta}}_0 = \hat{\boldsymbol{\vartheta}}_0'} \right) K(z_{th_2}) \right\| = o(1), \end{aligned}$$

where  $\hat{\boldsymbol{\theta}}$  is the minimizer of  $J(\boldsymbol{\theta})$ . Finally, because  $\psi_\tau(x)$  is an increasing function of  $x$ ; then  $-\boldsymbol{\theta}^T V_n(\lambda \boldsymbol{\theta}, \boldsymbol{\vartheta}_0) = a_n \sum_{t=1}^n \psi_\tau[v_t^*(\boldsymbol{\vartheta}_0) + \lambda a_n (-\boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0))] (-\boldsymbol{\theta}^T \mathbf{X}_t^*(\boldsymbol{\vartheta}_0)) K(z_{th_2})$  is an increasing function of  $\lambda$ . Thus, Condition (i) in Lemma A.7 is satisfied. Then, it follows from Lemma A.6, Lemmas A.8 and A.9 that

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n [\psi_\tau(v_t^*(0)) \mathbf{X}_t^* - a_n \Gamma^*(Z_t) \hat{\boldsymbol{\vartheta}}_0] K(z_{th_2}) + o_p(1) \\ &= \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \left[ \psi_\tau(v_t^*(0)) \mathbf{X}_t^* - \Gamma^*(Z_t) (\mathbf{D}^*(Z_t))^{-1} \right. \\ &\quad \times \left. \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n \psi_\tau(Y_s^*) \mathbf{W}_s K(z_{sh_1}) \right] K(z_{th_2}) + o_p(1) \\ &= \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \left[ \psi_\tau(v_t^*(0)) \mathbf{X}_t^* - \Gamma^*(Z_t) (\mathbf{D}^*(Z_t))^{-1} \right. \\ &\quad \times \left. \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n \{ \psi_\tau(Y_t^*) \mathbf{W}_t + \psi_\tau(Y_s^*) \mathbf{W}_s - \psi_\tau(Y_t^*) \mathbf{W}_t \} K(z_{sh_1}) \right] K(z_{th_2}) + o_p(1) \\ &= \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \left[ \psi_\tau(v_t^*(0)) \mathbf{X}_t^* - \psi_\tau(Y_t^*) \Gamma^*(Z_t) (\mathbf{D}^*(Z_t))^{-1} \mathbf{W}_t \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n K(z_{sh_1}) \right] K(z_{th_2}) \\ &\quad - \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \left[ \Gamma^*(Z_t) (\mathbf{D}^*(Z_t))^{-1} \right. \\ &\quad \times \left. \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{s=m+1 \neq t}^n \{ \psi_\tau(Y_s^*) \mathbf{W}_s - \psi_\tau(Y_t^*) \mathbf{W}_t \} K(z_{sh_1}) \right] K(z_{th_2}) + o_p(1). \end{aligned}$$

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

$$\hat{\boldsymbol{\theta}} = \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \left[ \psi_\tau(v_t^*(0)) \mathbf{X}_t^* - \psi_\tau(Y_t^*) \Gamma^*(Z_t) (\mathbf{D}^*(Z_t))^{-1} \mathbf{W}_t \right] K(z_{th_2}) + o_p(1),$$

by the fact that  $f_z^{-1}(Z_t) (nh_1)^{-1} \sum_{s=m+1 \neq t}^n K(z_{sh_1}) = 1 + o(1)$ . Therefore, following the proof of Theorem 1 in Cai and Xu (2008), the theorem is proved.  $\square$

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

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

$$\begin{aligned}
\hat{\Gamma}(z_0) &= \frac{1}{n} \sum_{t=1}^n w_{2t} \hat{\mathbf{X}}_t \hat{\mathbf{g}}_\tau^T(z_0) \mathbf{\Pi}_t K_{h_2}(Z_t - z_0) \\
&= \frac{1}{n} \sum_{t=1}^n w_{2t} (\hat{\mathbf{X}}_t - \mathbf{X}_t) (\hat{\mathbf{g}}_\tau(z_0) - \mathbf{g}_\tau(z_0))^T \mathbf{\Pi}_t K_{h_2}(Z_t - z_0) \\
&\quad + \frac{1}{n} \sum_{t=1}^n w_{2t} \mathbf{X}_t (\hat{\mathbf{g}}_\tau(z_0) - \mathbf{g}_\tau(z_0))^T \mathbf{\Pi}_t K_{h_2}(Z_t - z_0) \\
&\quad + \frac{1}{n} \sum_{t=1}^n w_{2t} (\hat{\mathbf{X}}_t - \mathbf{X}_t) \mathbf{g}_\tau^T(z_0) \mathbf{\Pi}_t K_{h_2}(Z_t - z_0) + \frac{1}{n} \sum_{t=1}^n w_{2t} \mathbf{X}_t \mathbf{g}_\tau^T(z_0) \mathbf{\Pi}_t K_{h_2}(Z_t - z_0) \\
&\equiv S^{(1)} + S^{(2)} + S^{(3)} + S^{(4)}.
\end{aligned}$$

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

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

On the other hand, by applying mean value theorem, there exists  $\hat{\boldsymbol{\vartheta}}'_0 \in (0, \hat{\boldsymbol{\vartheta}}_0)$  such that

$$\hat{\mathbf{X}}_t \equiv \mathbf{X}_t(\hat{\boldsymbol{\vartheta}}_0) = \mathbf{X}_t + \left( \frac{\partial \mathbf{X}_t(\hat{\boldsymbol{\vartheta}}_0)}{\partial \hat{\boldsymbol{\vartheta}}_0} \bigg|_{\hat{\boldsymbol{\vartheta}}_0 = \hat{\boldsymbol{\vartheta}}'_0} \right) \hat{\boldsymbol{\vartheta}}_0 = \mathbf{X}_t + (nh_1)^{-1/2} \mathbf{\Pi}_t \hat{\boldsymbol{\vartheta}}_0.$$

Therefore, by Lemma A.2,

$$E[S^{(3)}] = (nh_1)^{-1/2} E[f_{Y|Z, \mathbf{X}}(\mathbf{g}_\tau^T(z_0) \mathbf{X}_t) \mathbf{\Pi}_t \hat{\boldsymbol{\vartheta}}_0 \mathbf{g}_\tau^T(z_0) \mathbf{\Pi}_t K_{h_2}(Z_t - z_0)] + o(1) = O(m^{3/2}/nh_1) = o(1).$$

Similar to the proof of  $Var[T_n(0)]$  in Lemma A.5 and by Lemma A.2, 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,

$$\begin{aligned}
E[S^{(4)}] &= E[f_{Y|Z, \mathbf{X}}(\mathbf{g}_\tau^T(z_0) \mathbf{X}_t) \mathbf{X}_t \mathbf{g}_\tau^T(z_0) \mathbf{\Pi}_t K_{h_2}(Z_t - z_0)] + o(1) \\
&= \int f_{Y|Z, \mathbf{X}}(\mathbf{g}_\tau^T(z_0) \mathbf{X}_t) \mathbf{X}_t \mathbf{g}_\tau^T(z_0) \mathbf{\Pi}_t K(z) f_z(z_0 + h_2 z) dz + o(1) \rightarrow f_z(z_0) \Gamma(z_0).
\end{aligned}$$

Again, similar to the proof of  $Var[T_n(0)]$  in Lemma A.5, 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  $\hat{\Omega}(z_0)$ ,  $\hat{\Omega}^*(z_0)$ ,  $\hat{H}_1(z_0)$  and  $\hat{H}_2(z_0)$  can be derived in similar ways.  $\square$

## Appendix B: Mathematical Proof for Stationarity and $\alpha$ -Mixing

In this section, we show that the model (1) in the main article can generate a strictly stationary and  $\alpha$ -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  $1 \leq i \leq \kappa$ , let  $\mathcal{F}_{i,a}^b$  be the  $\sigma$ -algebra generated by  $\{(Y_{it}, Z_{it})\}_{t=a}^b$ . Then, a stationary process  $\{(Y_{it}, Z_{it})\}_{t=-\infty}^{\infty}$  is said to be  $\alpha$ -mixing (strongly mixing) if the mixing coefficient  $\alpha(t)$  defined by

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

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

To study the probabilistic properties of model (1) in the main article,  $\mathbb{Y}_t$  and  $\mathbf{q}_{\tau,t}$  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  $U_{it}$  ( $1 \leq i \leq \kappa$ ,  $1 \leq t \leq n$ ) as an 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

$$Y_{it} = \gamma_{i0}(U_{it}, Z_t) + \sum_{s=1}^q \gamma_{i,s}^T(U_{it}, Z_t) \mathbf{q}_{\tau,t-s} + \sum_{l=1}^p \beta_{i,l}^T(U_{it}, Z_t) \mathbb{Y}_{t-l}, \quad (\text{B.1})$$

and

$$\mathbf{q}_{\tau,t,i} = \gamma_{i0,\tau}(Z_t) + \sum_{s=1}^q \gamma_{i,s,\tau}^T(Z_t) \mathbf{q}_{\tau,t-s} + \sum_{l=1}^p \beta_{i,l,\tau}^T(Z_t) \mathbb{Y}_{t-l} \quad (\text{B.2})$$

for some  $p$  and  $q$ , where  $Y_{it}$ ,  $\mathbf{q}_{\tau,t}$  and  $\mathbb{Y}_t$  in (B.1) and (B.2) have the same definition as that in (1) and equation (B.2) is the same as (1) with  $Z_t = Z_{it}$ . In addition,  $\gamma_{i0}(\cdot, \cdot)$  in (B.1) is a scalar and measurable function of  $U_{it}$  and  $Z_t$  (from  $\mathbb{R}^2$  to  $\mathbb{R}$ ), both  $\gamma_{i,s}(\cdot, \cdot) = (\gamma_{si1}(\cdot, \cdot), \dots, \gamma_{si\kappa}(\cdot, \cdot))^T$  and  $\beta_{i,l}(\cdot, \cdot) = (\beta_{li1}(\cdot, \cdot), \dots, \beta_{li\kappa}(\cdot, \cdot))^T$  in (B.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 (B.1) is monotonically increasing in  $U_{it}$ , the conditional quantile function of  $Y_{it}$  given  $(Z_t, \{\mathbf{q}_{\tau, t-s}\}_{s=1}^q, \{\mathbb{Y}_{t-l}\}_{l=1}^p)$  becomes (B.2). Note that (B.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 (B.1) and (B.2) into an autoregression process of order 1 as follows

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

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

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

Here, for  $s = 1, \dots, q$  and  $l = 1, \dots, p$ ,  $\boldsymbol{\Gamma}_{\beta, U_t}(Z_t) = (\boldsymbol{\Gamma}_{\beta, 1, U_t}(Z_t), \dots, \boldsymbol{\Gamma}_{\beta, p, U_t}(Z_t))$ , where  $\boldsymbol{\Gamma}_{\beta, l, U_t}(Z_t) = (\beta_{lij}(U_{it}, Z_t))_{1 \leq i \leq \kappa, 1 \leq j \leq \kappa}$  is a  $\kappa \times \kappa$  matrix. In addition,  $\boldsymbol{\Gamma}_{U_t}(Z_t) = (\boldsymbol{\Gamma}_{1, U_t}(Z_t), \dots, \boldsymbol{\Gamma}_{q, U_t}(Z_t))$ , where  $\boldsymbol{\Gamma}_{s, U_t}(Z_t) = (\gamma_{sij}(U_{it}, Z_t))_{1 \leq i \leq \kappa, 1 \leq j \leq \kappa}$  is a  $\kappa \times \kappa$  matrix. Similarly,  $\boldsymbol{\Gamma}_{\beta, \tau}(Z_t) = (\boldsymbol{\Gamma}_{\beta, 1, \tau}(Z_t), \dots, \boldsymbol{\Gamma}_{\beta, p, \tau}(Z_t))$ , where  $\boldsymbol{\Gamma}_{\beta, l, \tau}(Z_t) = (\beta_{lij, \tau}(Z_t))_{1 \leq i \leq \kappa, 1 \leq j \leq \kappa}$  is a  $\kappa \times \kappa$  matrix. Also,  $\boldsymbol{\Gamma}_{\tau}(Z_t) = (\boldsymbol{\Gamma}_{1, \tau}(Z_t), \dots, \boldsymbol{\Gamma}_{q, \tau}(Z_t))$ , where  $\boldsymbol{\Gamma}_{s, \tau}(Z_t) = (\gamma_{sij, \tau}(Z_t))_{1 \leq i \leq \kappa, 1 \leq j \leq \kappa}$  is a  $\kappa \times \kappa$  matrix. Furthermore,  $\boldsymbol{\mu}(Z_t) = (E_U^T(\gamma_0(U_{it}, Z_t)), 0, \dots, 0, \gamma_{0, \tau}^T(Z_t), 0, \dots, 0)^T$ , where  $E_U(\gamma_0(U_{it}, Z_t)) = (E_U(\gamma_{10}(U_{1t}, Z_t)), \dots, E_U(\gamma_{\kappa 0}(U_{\kappa t}, Z_t)))^T$  and  $\gamma_{0, \tau}(Z_t) = (\gamma_{10, \tau}(Z_t), \dots, \gamma_{\kappa 0, \tau}(Z_t))^T$ . 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,  $\mathbf{D}_{U_t}(Z_t) = (\check{\gamma}_{10}(U_{1t}, Z_t), \dots, \check{\gamma}_{\kappa 0}(U_{\kappa t}, Z_t), 0_{1 \times \kappa(p+q-1)})^T$ , where  $\check{\gamma}_{i0}(U_{it}, Z_t) = \gamma_{i0}(U_{it}, Z_t) - E_U(\gamma_{i0}(U_{it}, Z_t))$ .

**Remark B.1.** Notice that when setting  $Z_t$  as a smoothing variable, the equations corresponding to  $(\kappa p + 1)$ -th,  $\dots$ ,  $(\kappa p + \kappa)$ -th rows of (B.3) are exactly the (B.2) and the model (1) in the main article, while the  $i$ th row of (B.3) with  $i = 1, \dots, \kappa$  is equation (B.1). Given these relations, one can conclude that  $\mathbb{Y}_t$  and  $\mathbf{q}_{\tau, t}$  jointly follow a VAR process of order 1 in (B.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  $\lambda_{\max}(\mathbf{A}_{U_t})$  as the largest eigenvalue in absolute value of following matrix  $\mathbf{A}_{U_t}$ :

$$\mathbf{A}_{U_t} = \begin{pmatrix} \Gamma_{\beta,1,U_t} & \Gamma_{\beta,2,U_t} & \cdots & \Gamma_{\beta,p-1,U_t} & \Gamma_{\beta,p,U_t} & \Gamma_{1,U_t} & \Gamma_{2,U_t} & \cdots & \Gamma_{q-1,U_t} & \Gamma_{q,U_t} \\ I_{\kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} \\ 0_{\kappa \times \kappa} & I_{\kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & I_{\kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} \\ \Gamma_{\beta,1} & \Gamma_{\beta,2} & \cdots & \Gamma_{\beta,p-1} & \Gamma_{\beta,p} & \Gamma_1 & \Gamma_2 & \cdots & \Gamma_{q-1} & \Gamma_q \\ 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & I_{\kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} \\ 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & I_{\kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & 0_{\kappa \times \kappa} & \cdots & I_{\kappa} & 0_{\kappa \times \kappa} \end{pmatrix},$$

where

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

$$\Gamma_{\beta,l} = \begin{pmatrix} \beta_{l11,\tau} & \beta_{l12,\tau} & \cdots & \beta_{l1\kappa,\tau} \\ \beta_{l21,\tau} & \beta_{l22,\tau} & \cdots & \beta_{l2\kappa,\tau} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{l\kappa 1,\tau} & \beta_{l\kappa 2,\tau} & \cdots & \beta_{l\kappa \kappa,\tau} \end{pmatrix}, \quad \text{and} \quad \Gamma_s = \begin{pmatrix} \gamma_{s11,\tau} & \gamma_{s12,\tau} & \cdots & \gamma_{s1\kappa,\tau} \\ \gamma_{s21,\tau} & \gamma_{s22,\tau} & \cdots & \gamma_{s2\kappa,\tau} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{s\kappa 1,\tau} & \gamma_{s\kappa 2,\tau} & \cdots & \gamma_{s\kappa \kappa,\tau} \end{pmatrix},$$

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

## Assumption B.

**B1:** Let  $\{\mathbb{X}_t\}$  be a  $\phi$ -irreducible and aperiodic Markov chain. For  $i = 1, \dots, \kappa$ ,  $j = 1, \dots, \kappa$ ,  $l = 1, \dots, p$  and  $s = 1, \dots, q$ , each entry of  $\mathbf{\Gamma}_{s,U_t}(Z_t)$  and  $\mathbf{\Gamma}_{\beta,l,U_t}(Z_t)$  in (B.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 (B.2) is bounded such that  $|\gamma_{sij,\tau}(\cdot)| \leq \gamma_{sij,\tau}$  and  $|\beta_{lij,\tau}(\cdot)| \leq \beta_{lij,\tau}$ . Furthermore,  $E\{[\lambda_{\max}(\mathbf{A}_{U_t})]^2\} < 1$ .

**B2:** For  $i = 1, \dots, \kappa$ ,  $\tilde{\gamma}_{i0}(U_{it}, Z_t)$  in  $\mathbf{D}_{U_t}(Z_t)$  is bounded such that  $|\tilde{\gamma}_{i0}(U_{it}, \cdot)| \leq \tilde{\gamma}_{i0}(U_{it})$ , where  $\{\tilde{\gamma}_{i0}(U_{it})\}$  are i.i.d. random variables with mean 0 and finite variance. In addition, denote  $\mathbf{D}_{U_t} = (\tilde{\gamma}_{10}(U_{1t}), \dots, \tilde{\gamma}_{\kappa 0}(U_{\kappa t}), 0_{1 \times \kappa(p+q-1)})^T$ , then,  $E\|\mathbf{D}_{U_t}\|^2 < \infty$  and  $E\|\boldsymbol{\mu}(Z_t)\| < \infty$ .

**Remark B.2.** The  $\phi$ -irreducibility and aperiodicity in Assumption B1 are key assumptions for deriving geometric ergodicity and subsequently,  $\alpha$ -mixing property. The conditions that imply  $\phi$ -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  $\phi$ -irreducible and aperiodic Markov chain. In addition, Pham (1986) obtained conditions for random coefficient autoregressive (RCA) models to be  $\phi$ -irreducible. In this article, we simply impose the assumptions of  $\phi$ -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  $\phi$ -irreducibility and aperiodicity and we leave this as a future topic. Moreover, the moment conditions  $E\{[\lambda_{\max}(\mathbf{A}_{U_t})]^2\} < 1$  in Assumption B1 is used to bound the random matrices  $\mathbf{A}_{U_t}(Z_t)$ , 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 (B.3), but instead focus on constructing varying interdependences among conditional quantiles.

**Proposition B.1.** Under Assumptions B1 and B2, if  $\mathbb{X}_0$  is initialized from the invariant measure, then,  $\{\mathbb{X}_t\}$  defined in (B.3) is a strictly stationary and  $\alpha$ -mixing process.

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

**Lemma B.1.** Under Assumptions B1 and B2, for any  $\mathbb{W} = (w_1, \dots, w_{\kappa(p+q)})^T$ , we have

$$\|\mathbf{A}_{U_t}(Z_t)\mathbb{W}\|_{\psi,2} \leq \|\mathbf{A}_{U_t}|\mathbb{W}|\|_{\psi,2}. \text{ Here, } \mathbf{A}_{U_t}(Z_t) \text{ is defined in (B.3), } \mathbf{A}_{U_t} \text{ is defined previously and } |\mathbb{W}| = (|w_1|, \dots, |w_{\kappa(p+q)}|)^T.$$

*Proof.* Similar to the proof of Lemma A.1 in Chen and Tsay (1993), let  $\mathbf{A}_{U_t}(Z_t)\mathbb{W} = (d_1, \dots, d_{\kappa(p+q)})^T$  and  $\mathbf{A}_{U_t}|\mathbb{W}| = (g_1, \dots, g_{\kappa(p+q)})^T$ . Then, for  $\iota = \kappa + 1, \dots, \kappa p$  and for  $\iota = \kappa p + \kappa + 1, \dots, \kappa(p+q)$ , we have  $|d_\iota| = g_\iota$ . For  $\iota = 1, \dots, \kappa$  and for  $\iota' = \kappa p + 1, \dots, \kappa p + \kappa$ , by Assumptions B1 and B2,

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

and

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

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

### Proof of Proposition B.1:

*Proof.* By Proposition 3 in Carrasco and Chen (2002) and Lemma 2 in Pham (1986), Assumption B1 implies  $\|\mathbf{A}_{U_t}\|_{\psi,2} < 1$  for all  $U_{it} \in [0, 1]$ . Then, we can find  $0 < \delta < 1$  and  $\varrho > 0$ , such that

$\|\prod_{j=0}^{\varrho-1} \mathbf{A}_{U_{t+j}}\|_{\psi,2} < 1 - \delta$ . Consequently, by Assumption B2 and Lemma B.1, for some constant  $C > 0$ ,

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

where each element of  $\mathbf{D}_{U_t} = (\tilde{\gamma}_{10}(U_{1t}), \dots, \tilde{\gamma}_{\kappa 0}(U_{\kappa t}), 0_{1 \times \kappa(p+q-1)})^T$  is defined in Assumption B2 and the first inequality follows from Jensen's inequality. Notice that  $E\left\|\sum_{j=1}^{\varrho} \left[\prod_{i=j}^{\varrho-1} \mathbf{A}_{U_{t+i}}\right]\right\|$  is bounded and by Assumption B2,  $E\|\mathbf{D}_{U_t}\|$  is bounded, so that  $E\left\|\sum_{j=1}^{\varrho} \left[\prod_{i=j}^{\varrho-1} \mathbf{A}_{U_{t+i}}\right] |\mathbf{D}_{U_{t+j}}|\right\|$  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  $\|\mathbb{X}\| > M$ ,

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

where  $0 < \delta_1 < 1$ . Hence, the compact set  $K = \{\mathbb{X} : \|\mathbb{X}\| \leq M\}$  satisfies that when  $\mathbb{X} \notin K$ ,  $E(\|\mathbb{X}_{t+\varrho}\| \mid \mathbb{X}_t = \mathbb{X}) < (1 - \delta_1)\|\mathbb{X}\|$ . By Lemma 1.1 and Lemma 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  $\alpha$ -mixing.  $\square$

## References

- Bercu, B., and Touati, A. (2008). Exponential Inequalities for Self-normalized Martingales With Applications. *The Annals of Applied Probability*, 18(4), 1848-1869.
- 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, R. and Tsay, R.S. (1993). Functional Coefficient Autoregressive Model. *Journal of the American Statistical Association*, 88(421), 298-308.
- 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.
- 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.
- Welsh, A. (1989). On M-Processes and M-Estimation. *Annals of Statistics*, 17(1), 337-361.
- 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.