Econometric Modeling for Functional-Coefficient VAR Models: Theories and Applications

Xiyuan Liu

Submitted to the graduate degree program in Department of Economics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Zongwu Cai, Chairperson

William A. Barnett

Committee members

John Keating

Shahnaz Parsaeian

Bozenna Pasik-Duncan

Date defended: May 11, 2022

The Dissertation Committee for Xiyuan Liu certifies
that this is the approved version of the following dissertation :

Econometric Modeling for Functional-Coefficient VAR Models: Theories and Applications

Zongwu Cai, Chairperson

Date approved: May 11, 2022

Abstract

This dissertation proposes theories and applications for three new types of functional-coefficient VAR models. The first part of dissertation develops 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. The contributions to literature are four-fold in this part. First, the model setting is general enough to nest many well-known dynamic quantile models in the literature. Second, by allowing coefficients to vary with a smoothing variable, the proposed 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 second part proposes a new class of functional-coefficient factor-augmented predictive VAR (FC-FAVAR) models. Different from the existing literature, this model setting allows both factor loadings of corresponding factor model and coefficients of this predictive VAR model vary with a smoothing economic variable, which adds additional information of variation in the factor structure and economic interpretability to the predictive model. Moreover, both observed variables and unobserved factor regressors in this new model are jointly imposed in a vector autoregressive form. In this way, some important information of model dynamic may be included in these lagged factors, which is helpful to enhance the ability of prediction. Finally, the proposed model is applied in both simulation and empirical

study of one-step ahead prediction, which demonstrate its reliability in forecasting.

In the third part, effects of monetary policy shocks on large amounts of macroeconomic variables are identified by a class of FC-FAVAR models. In the empirical study, I analyze the generalized impulse response functions (GIRF) estimated by the newly proposed model and compare my results with those from classical FAVAR models. The major contributions are two parts. In the empirical study, I provide an alternative way from econometric perspective to reduce price puzzle by using the proposed FC-FAVAR model, without introducing new variables or structure in the conventional macroeconomic model and replacing policy instruments.

Keywords: Conditional quantile models; Dynamic financial network; Functional coefficient models; Nonparametric estimation; VAR modeling; Factor-augmented vector autoregressive; Factor model; Forecasting; Impulse response functions; Price puzzle

JEL Classification: C14, C58, C45, G32, E30, E31

Acknowledgements

I would like to gratefully thank my advisor Professor Zongwu Cai for his intellectual guidance, physical and mental support, and patience. Prof. Cai encouraged me to explore new research topics and always made himself available to discuss about every detail of these topics. He taught me how to think like an independent scholar and deeply re-shape my taste on doing research. I never forget the experiences of being one of participants of Prof. Cai's weekly workshop on Friday. This workshop provides opportunity for PhD students and visiting scholars to present papers and discuss recent advances on econometric and statistical issues. My research interests and dissertation topic were established in this workshop and I was so lucky to have attended at most of the time of my PhD program, as it was cancelled since the COVID-19 pandemic. In addition, Prof. Cai helped me with revising dissertation and slides for defense so carefully, which are definitely hard works and time-consuming. I am also greatly indebted to Prof. Cai for his strong supports on recommendation and revising my application documents during my job searching. Moreover, he and his wife hosted our PhD students and visiting scholars for lunch or dinner many times throughout years, which brought me one of the warmest memories while in the U.S.. Many thanks to Prof. Cai, without his guidance and support, I shall never receive my PhD degree.

I would like to deeply thank my committee members Professor William A. Barnett, Professor John W. Keating and Professor Shahnaz Parsaeian for their warmly helps and suggestions on my research itself, which improve my understanding on macroeconomic study. In addition, I am greatly in indebted to my committee member Professor Pasik-Duncan for her kindness and advices about giving presentation. Meanwhile, I would also like to thank Professor Josephine Lugovskyy and Professor Tarun Sabarwal for their references and helpful advices throughout my experience in the PhD program. Finally, special thanks to Kate Pleskac for her excellent and efficient assistance on achieving my milestones and arranging my final defense.

This dissertation is dedicated to the members of my family. I would like to thank my mother Jiangnan Xie and my father Hongtao Liu, who cultivate me to be curious about this world and support me with love and dedication. I am extremely grateful to the support from my wife Zequn Wu, who always inspires me and give me encourage when I was upset. Our memories of childhood and overcoming hardship together never fade away in my life.

Contents

List of Figures

List of Tables

Chapter 1

Literature Review

1.1 Dynamic Quantile Model

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.

Assume that $\{V_t, y_t\}_{t=-\infty}^{\infty}$ be a strictly stationary sequence and $F(y|v)$ denote the conditional distribution of y_t given $V_t = v$. The conditional quantile function of y_t given $V_t = v$, $q_\tau(v)$, is defined as, for any $0 < \tau < 1$,

$$q_\tau(v) = \inf\{y \in \mathbb{R} : F(y|v) \geq \tau\}.$$

Equivalently, $q_\tau(v)$ can be expressed as

$$q_\tau(v) = \underset{a \in \mathbb{R}}{\operatorname{argmin}} E\{\rho(y_t - a)|V_t = v\},$$

where $\rho_\tau(y) = y[\tau - I(y < 0)]$ is called the "check" (loss) function and $I(A)$ is the indicator function of any set A. 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 conditional Value-at-Risk (CVaR), which can be described as:

$$\mathrm{CVaR}_{\tau,t} = -\inf\{Y \in \mathbb{R} : F(Y|\mathcal{F}_{t-1}) > \tau\},$$

where $\tau \in (0,1)$ is the quantile level, $\mathcal{F}_{t-1}$ is the information set to present all information of the return available at time $t-1$, and $F(\cdot|\mathcal{F}_{t-1})$ represents the conditional distribution function of Y_t given $\mathcal{F}_{t-1}$. Clearly, calculating $\mathrm{CVaR}_{\tau,t}$ is a special procedure for estimating conditional quantiles of financial return distribution. In addition, since $\mathrm{CVaR}_{\tau,t}$ is also a particular quantile of future portfolio value, conditional on current information as discussed in Engle and Manganelli (2004), it is natural to consider the dynamic feature included in conditional quantiles when estimating $\mathrm{CVaR}_{\tau,t}$ and dynamic quantile models can provide a nice tool to achieve this goal. Some early works on dynamic quantile models 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 Gourieroux 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.

1.1.1 Univariate Conditional Autoregressive Value at Risk by Regression Quantiles (CAViaR) models

Let us first look at a simple example of univariate Conditional Autoregressive Value at Risk by Regression Quantiles (CAViaR) models proposed by Engle and Manganelli (2004)

$$q_{\tau,t} = \gamma_{0,\tau} + \sum_{s=1}^{q} \gamma_{s,\tau} q_{\tau,t-s} + \sum_{l=1}^{p} \beta_{l,\tau} |Y_{t-l}|, \tag{1.1}$$

where $q_{\tau,t}$ is the conditional quantile for the return Y_t, and $|\cdot|$ denotes the absolute value. As pointed out by Xiao and Koenker (2009), the CAViaR model has attracted a great deal of research attention in recent two decades. Engle and Manganelli (2004) focused on introducing the CAViaR model instead of on estimating such models. In the CAViaR model, because the regressors $q_{\tau,t-s}$ are latent and are dependent on the unknown parameters, estimation of the CAViaR model is complicated, and conventional nonlinear quantile regression techniques are not directly applicable. Meanwhile, Xiao and Koenker (2009) studied a special case of CAViaR model as follow

$$q_{\tau,t} = \gamma_{0,\tau} + \sum_{s=1}^{q} \gamma_{s} q_{\tau,t-s} + \sum_{l=1}^{p} \beta_{l,\tau} |Y_{t-l}|$$

with $Y_t = \sigma_t \varepsilon_t$ being generated from a linear GARCH (p, q)-type process, which extends from settings in Taylor (1986):

$$\sigma_t = \gamma_0 + \sum_{s=1}^{q} \gamma_s \sigma_{t-s} + \sum_{l=1}^{p} \beta_l |Y_{t-l}|.$$

where $q_{\tau,t} = \sigma_t F_{\varepsilon}^{-1}(\tau)$, σ_t^2 is the conditional variance of Y_t, $F_{\varepsilon}(\cdot)$ is a distribution function of ε_t, $\gamma_{0,\tau} = \gamma_0 F_{\varepsilon}^{-1}(\tau)$, and $\beta_{l,\tau} = \beta_l F_{\varepsilon}^{-1}(\tau)$.

1.1.2 Vector Autoregressive (VAR) for CVaR Models and Tail Dependence

Note that the aforementioned models are only for the univariate case. For the multivariate case, White, Kim and Manganelli (2015) proposed a vector autoregressive (VAR) for CVaR models as follow

$$q_{\tau,t,i} = \gamma_{i0} + \sum_{s=1}^{q} \boldsymbol{\gamma}_{i,s}^T \boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}_{i,l}^T \mathbb{Y}_{t-l}, \quad i = 1, 2, \ldots, \kappa, \tag{1.2}$$

where $q_{\tau,t,i}$ is the conditional quantile for the return Y_{it} of individual i, $\boldsymbol{q}_{\tau,t} = (q_{\tau,t,1}, \ldots, q_{\tau,t,\kappa})^T$, $\mathbb{Y}_t = (|Y_{1t}|, \ldots, |Y_{\kappa t}|)^T$, and $\boldsymbol{\gamma}_{i,s} = (\gamma_{si1}, \ldots, \gamma_{si\kappa})^T$, and $\boldsymbol{\beta}_{i,l} = (\beta_{li1}, \ldots, \beta_{li\kappa})^T$. It is worthwhile to mention that the multivariate dynamic quantile models (1.2) 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). With the help of model (1.2), White et al. (2015) is enable 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.

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 Yílmaz (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.

1.1.3 Financial Network

As a direct extension of the concept of tail dependence, financial network has attracted more and more attentions in recent decade. Indeed, 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.

To be specific, let us consider following framework with constant interdependences among CVaRs studied by White et al. (2015):

$$\mathrm{CVaR}_{it} = \boldsymbol{\gamma}_i^T \mathrm{CVaR}_{t-1}, \quad i = 1, 2, \ldots, \kappa, \tag{1.3}$$

which is a special case of (1.2), where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{i\kappa})^T$ is a vector of constant parameters that represent static interdependences among VaRs, and

$$\mathrm{CVaR}_{t-1} = (\mathrm{CVaR}_{1(t-1)}, \ldots, \mathrm{CVaR}_{\kappa(t-1)})^T$$

is a vector of all returns' CVaRs at time $t-1$. Then, the matrix $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_\kappa)_{\kappa\times\kappa}$ can be regarded as a financial network for measuring transmissions of financial risks. Following this framework, 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. 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 Yílmaz (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. Extensive reviews about financial network can be found in Diebold and Yílmaz (2014) and Härdle et al. (2016).

1.2 Factor-Augmented VAR (FAVAR) Model

It is interesting to see that the multivariate dynamic quantile model (1.2) can be further extended when replacing vector of latent variables $\boldsymbol{q}_{\tau,t}$ by other types of latent vector. For example, if $\boldsymbol{q}_{\tau,t}$ is replaced by a vector of unobserved latent factors $\boldsymbol{f}_t$, then model (1.2) becomes a factor-augmented VAR (FAVAR) model constructed in Bernanke, Boivin and Eliasz (2005). In the next subsection, I will give detailed literature review about FAVAR models and their extensions.

1.2.1 Factor-Augmented VAR (FAVAR) Model with Fixed Coefficients

Linear vector autoregression (VAR) models and their extensions such as vector autoregressive moving average (VARMA) models and VARs with exogenous variables (VARX) were well developed in last century for studying the effects of monetary policy shocks on macroeconomic variables and modeling the dynamic interdependences among them. These models mainly arise as powerful tool-kits for macroeconomists but only impose minimal restrictions on the identification of large-scale macroeconomic models (Sims, 1980).

Despite the popularity, linear VAR models assume that the variables entered in econometric models are observable. Nevertheless, Bernanke et al. (2005) claimed that the assumption that both the central bank and the econometrician observe all the elements for estimating VAR model is too strong. In addition, since the standard VAR in literature usually did not contain more than six to eight variables, as argued in Bernanke et al. (2005), such low-dimensional linear vector VAR models may not include adequate information used by both central bank and private sectors. As an alternative, Stock and Watson (2002) seminally introduced the method of factor-augmented forecasts (also known as "diffusion index forecasts") in the VAR literature to exploit the information in a large set of macroeconomic variables. After this work, factor-augmented methods are being used by an increasing number of researchers and begin to fuse with linear VAR models and their variants. Pioneering contributions include the factor-augmented vector autoregressions (FAVAR) proposed in Bernanke et al. (2005) and the asymptotic theory for the estimated parameters of the factor-augmented regressions in Bai and Ng (2006). The linear FAVAR model in Bernanke et al. (2005) assumes the following form

$$P_t = \gamma_0 + \Gamma_1 P_{t-1} + \Gamma_2 P_{t-2} + \cdots + \Gamma_q P_{t-q} + \boldsymbol{\varepsilon}_t, \tag{1.4}$$

where $P_t = (\boldsymbol{f}_t^T, \boldsymbol{y}_t^T)^T$, with $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{rt})^T$ being a $r \times 1$ vector of unobservable factors and $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{m,t})^T$ being a $m \times 1$ vector of observable economic variables. In addition, let $Q = m + r$, then, $\gamma_0 = (\gamma_{10}, \ldots, \gamma_{Q0})^T$ is denoted as a vector of scalar intercepts, Γ_k is a $Q \times Q$ coefficient matrix for $1 \leq k \leq q$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{Q,t})^T$ is a vector of error terms. Furthermore, let $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{Nt})^T$ be a $N \times 1$ vector of available predictive variables at time t for $1 \leq t \leq n$, Bernanke et al. (2005) assumed that $\boldsymbol{x}_t$ is affected by P_t in model (1.4) through following equation with factors

$$\boldsymbol{x}_t = \boldsymbol{B_f} \boldsymbol{f}_t + \boldsymbol{B_y} \boldsymbol{y}_t + \boldsymbol{u}_t, \tag{1.5}$$

where $\boldsymbol{B_f}$ is a $N \times r$ matrix of factor loadings, $\boldsymbol{B_y}$ is a $N \times m$ matrix of coefficients, and $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})^T$ is a $N \times 1$ vector of idiosyncratic errors. Note that the number N is large and it is commonly assumed to be much greater than the number of factors and observed variables ($r + m \ll N$). Notice that model (1.5) can be transformed into a factor model with only unobserved $\boldsymbol{f}_t$ by imposing an orthogonality restriction between $\boldsymbol{f}_t$ and observed $\boldsymbol{y}_t$, see, for example, Bai et al. (2016) and Yamamoto (2019). In recent years, there has been increasing interest on studying factor models, see, for example, Chamberlain and Rothschild (1983), Fama and French (1992), Bai and Ng (2002), Fan, Liao and Mincheva (2013), Fan and Liao (2020) and the references therein. It is well known in the literature of dynamic factor models that the information from a large number of time series can be summarized by a relatively small set of estimated factors, see, e.g., Stock and Watson (2002) and Bernanke and Boivin (2003). In the further extensions, Dufour and Stevanović (2013) considered the combination of vector autoregressive moving-average (VARMA) models and factor-augmented techniques. Moreover, Bai, Li and Lu (2016) derived the inferential theory that corresponds to a maximum likelihood estimation for FAVAR models.

So far, the aforementioned papers are based on the assumption that the coefficients of the factor-augmented regression models are constant over time. However, the structural instability of factor-augmented models was also witnessed by numerous studies. For instance, Corradi and Swanson (2014) constructed a test for the joint hypothesis of structural stability of both factor loadings and coefficients in factor-augmented forecasting model.

1.2.2 Functional Coefficients Factor-Augmented Forecasting Model

To address inherent issues in static factor-augmented models, recently, Li, Tosasukul and Zhang (2020) proposed a univariate factor-augmented predictive regression model with functional coefficients, which allows the coefficients to vary with a variable. Specifically, for $1 \leq j \leq M$, define

$$y_{j,t} = \gamma_{j0}(Z_{jt}) + \sum_{d=1}^{q_f} \boldsymbol{\gamma}_{j,d,\boldsymbol{f}}^T(Z_{jt})\boldsymbol{f}_{t-d} + \sum_{c=1}^{q_y} \boldsymbol{\gamma}_{j,c,y}^T(Z_{jt})\boldsymbol{y}_{t-c} + v_{j,t}, \tag{1.6}$$

where $\gamma_{j0}(\cdot)$ is a scalar function, $\boldsymbol{\gamma}_{j,d,\boldsymbol{f}}(\cdot) = (\gamma_{dj1,\boldsymbol{f}}(\cdot), \ldots, \gamma_{djr,\boldsymbol{f}}(\cdot))^T$ is a $r \times 1$ vector of functional coefficients, $\boldsymbol{\gamma}_{j,c,y}(\cdot) = (\gamma_{cj1,y}(\cdot), \ldots, \gamma_{cjm,y}(\cdot))^T$ is a $m \times 1$ vector of functional coefficients, Z_{jt} is an observable scalar smoothing variable, and $v_{j,t}$ is an error term. Notice that the model (1.6) covers the model in Yan and Cheng (2022), who studied a parametric factor-augmented forecasting model in the presence of threshold effects. Of course, model (1.6) also includes the threshold models without factors studied in Tsay (1998), where predictive residuals are used to construct a test statistic to detect threshold nonlinearity in a vector time series. In their paper, Tsay (1998) applied their modeling procedure to study U.S. monthly interest rates and two daily river flow series of Iceland.

1.3 Overview

The rest of this dissertation is organized as follows. In Chapter 2, I 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, I develop an easy-to-implement two-stage procedure to estimate functionals in the dynamic network system by the local linear smoothing technique. I 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.

Chapter 3 proposes a new class of functional-coefficient factor-augmented predictive VAR models, where both factor loadings of corresponding factor model and coefficients of this predictive VAR model vary with a smoothing variable. To estimate this new model, I develop a simple-to-implement procedure which consists of two steps. In step one, the unobserved factor regressors in this predictive model are estimated via a local principal component analysis (local PCA) method. After obtaining the estimated factor regressors, a local linear smoothing method is applied to estimate the coefficient functions in the predictive model for one-step ahead forecasting, and the corresponding prediction interval is constructed by a wild bootstrap procedure. Finally, a simulation study and an empirical application of forecasting the consumer price index (CPI) in the U.S. demonstrate that my estimation procedure is reliable and works reasonably well.

In Chapter 4, effects of monetary policy shocks on large amounts of macroeconomic variables are identified by a class of functional-coefficient factor-augmented vector autoregressive (FC-FAVAR) models proposed in chapter 3, which allows coefficients of classical FAVAR models to vary with some variable. In the empirical study, I analyze the generalized impulse response functions estimated by the newly proposed model and compare my results with those from classical FAVAR models. My empirical finding is that the proposed new model has an ability to reduce the well-known price puzzle without adding new variables into the dataset.

Chapter 2

A Nonparametric Dynamic Network via Multivariate Quantile Autoregressions

2.1 Introduction

In this article, I propose a nonparametric approach involving multivariate dynamic quantile models with nonlinear structures. Different from previous studies, I 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, I 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 (2.1) 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, I adapt one of modeling methods to analyze dynamic quantiles, called the functional coefficient modeling approach. Compared with the existing literature, my approach is different mainly in three parts. First, I provide a kernel-based estimation framework for a new type of dynamic quantile model, which imposes relatively less restriction on model's structure. Second, my 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 my motivations for this study comes from analyzing the dynamic mechanism of financial network in international equity markets. Compared to the literature thus far, I 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 Yílmaz (2014). Moreover, in order to illustrate overall patterns of time-varying network of risk across institutions, the main interest in this chapter 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, I 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. The empirical study in this chapter 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 2.4.

Lastly, my contributions to the literature can be summarized as follows. First, the model setting in this chapter (see (2.1) 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 chapter is organized as follows. In Section 2.2, 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 2.3 to illustrate the finite sample performance of the proposed estimation procedure. In Section 2.4, the 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 2.5 and all the technical proofs are gathered in the Appendix A. Throughout this chapter, $0_{a\times b}$ stands for the $(a \times b)$ matrix of zeros and I_a is the $(a \times a)$ identity matrix.

2.2 FCVAR Model for Dynamic Quantiles

2.2.1 Model Setup

Let Y_{it} $(1 \le i \le \kappa,\ 1 \le t \le n)$, a scalar dependent variable, be the ith observation at time t, $\mathcal{F}_{i,t-1}$ represent information set up to time $t-1$ for $1 \le i \le \kappa$, and $q_{\tau,t,i}$ be the

τth conditional quantile of Y_{it} given $\mathcal{F}_{i,t-1}$. Then, I study the following functional-coefficient VAR model for dynamic quantiles, termed as FCVAR-DQ model, given by, for $1 \le i \le \kappa$ and $1 \le t \le n$,

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_{it}) + \sum_{s=1}^{q} \boldsymbol{\gamma}_{i,s,\tau}^T(Z_{it})\boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}_{i,l,\tau}^T(Z_{it})\mathbb{Y}_{t-l} \tag{2.1}$$

for some p and q, where $\boldsymbol{q}_{\tau,t} = (q_{\tau,t,1}, \dots, q_{\tau,t,\kappa})^T$ 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 τ, 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 τ 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 (2.1) by allowing τ in $q_{\tau,t,i}$ to vary across different equations, only with mild changes on asymptotic theory in this paper. Thus, in order to meet our empirical motivation, all of $\tau's$ in model (2.1) are the same throughout this article.

Importantly, in the case of estimating dynamic financial network in empirical studies, by following White et al. (2015), I 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 chapter, for easy exposition, my focus is on the simple case that $q = p = 1$, $\kappa = \kappa_1$, $\mathbb{Y}_t = (|Y_{1t}|, \ldots, |Y_{\kappa t}|)^T$, and $Z_{it} = Z_t$ for all $1 \le i \le \kappa$. Then, model (2.1) can be rewritten as

$$q_{\tau,t,i} = \boldsymbol{g}_{i,\tau}^T(Z_t)\boldsymbol{X}_t, \tag{2.2}$$

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

It is worthwhile to note that if $q_{\tau,t,i}$ in model (2.2) is defined as VaR of return Y_{it}, then, $\{\gamma_{ij,\tau}(Z_t)\}_{i=1,j=1}^{\kappa}$ in model (2.2) 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

$$\boldsymbol{\Gamma}_{1,\tau}(Z_t) = \left(\gamma_{ij,\tau}(Z_t)\right)_{\kappa\times\kappa}. \tag{2.3}$$

Then, (2.2) can be expressed as a matrix form, which, indeed, is a FCVAR model for $\boldsymbol{q}_{\tau,t}$ with exogenous variables,

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

where $\boldsymbol{\gamma}_{0,\tau}(Z_t)$ and $\boldsymbol{\Gamma}_{\beta,1,\tau}(Z_t)$ are defined obviously. Therefore, $\boldsymbol{\Gamma}_{1,\tau}(Z_t)$ in (2.3) can serve as a dynamic network system changing with both τ and some information variable Z_t, and it is in a nonparametric nature, so that it is a nonparametric dynamic network. Notice that the general setting in the dynamic network system (2.3) 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 2.2.1

later), it is assumed throughout this article that the process $\{(Y_{it}, x_{it}, Z_t)\}$ in model (2.1) is strictly stationary and α-mixing (strongly mixing). Indeed, in the Appendix (see Appendix B), I 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 (2.1) is strictly stationary and α-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 α-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 α-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 my main interests in this article are to derive the asymptotic theory for model (2.2) and estimate a new class of dynamic financial network. Therefore, I provide some sufficient conditions that imply these important probabilistic properties and corresponding rigorously theoretical justifications in the Appendix (see Appendix B).

Remark 2.2.1. *(Special Cases) The proposed FCVAR-DQ model* (2.1) *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* (2.1), Y_{it} *in* (2.1) *takes a simple form as* $Y_{it} = \sigma_{it}\, e_{it}$, *where* σ_{it}^2 *is the conditional variance of* Y_{it} *and* e_{it} *is an 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 κ ($\kappa \geq 1$) returns extended from the setting in Taylor (1986) as follows*

$$\sigma_{it} = \gamma_{i0}(Z_t) + \sum_{s=1}^{q} \boldsymbol{\gamma}_{i,s}^T(Z_t)\boldsymbol{\Sigma}_{t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}_{i,l}^T(Z_t)\mathbb{Y}_{t-l}, \tag{2.4}$$

where $\boldsymbol{\Sigma}_t = (\sigma_{it}, \ldots, \sigma_{\kappa t})^T$ *and* $\mathbb{Y}_t = (|Y_{1t}|, \ldots, |Y_{\kappa t}|)^T$, *then, model* (2.1) *reduces to following dynamic quantile model:*

$$q_{\tau,t,i} = \gamma_{i0,\tau}(Z_t) + \sum_{s=1}^{q} \boldsymbol{\gamma}_{i,s}^T(Z_t)\boldsymbol{q}_{\tau,t-s} + \sum_{l=1}^{p} \boldsymbol{\beta}_{i,l,\tau}^T(Z_t)\mathbb{Y}_{t-l}, \tag{2.5}$$

where $\gamma_{i0,\tau}(\cdot) = \gamma_{i0}(\cdot)F_e^{-1}(\tau)$, $\boldsymbol{\gamma}_{i,s}(\cdot) = (\gamma_{si1}(\cdot), \ldots, \gamma_{si\kappa}(\cdot))^T$ *and* $\boldsymbol{\beta}_{i,l,\tau}(\cdot) = (\beta_{li1,\tau}(\cdot), \ldots, \beta_{li\kappa,\tau}(\cdot))^T$ *with* $\beta_{lij,\tau}(\cdot) = \beta_{lij}(\cdot)F_e^{-1}(\tau)$. *Notice that if* $\gamma's$ *and* $\beta's$ *in* (2.5) *are constant, model* (2.5) *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* $\boldsymbol{q}_{\tau,t}$ *would be observable and all coefficients are threshold functions, model* (2.1) *covers the model in Tsay (1998).*

Remark 2.2.2. *(Monotonicity). The issue of monotonicity is frequently discussed for the quantile autoregression model. A specific case for the monotonicity of* (2.1) *to hold is that* $\{\boldsymbol{\gamma}_{i,s,\tau}(Z_t)\}_{i=1,s=1}^{\kappa,q}$ *and* $\{\boldsymbol{\beta}_{i,l,\tau}(Z_t)\}_{i=1,l=1}^{\kappa,p}$ *are all monotone increasing functions with respect to* τ, *and* $\mathbb{Y}_t$ *is a positive random vector. In other cases, the assumption of monotonicity can be satisfied by conducting certain data transformation techniques; see Koenker and Xiao (2006) and Fan and Fan (2006) for detailed discussions.*

Remark 2.2.3. *(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* (2.2). *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 2.4 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.*

2.2.2 Two-stage Estimation Procedure

Since the estimation procedures for (2.1) and (2.2) are the same, I aim at estimating functional coefficients $\boldsymbol{g}_{i,\tau}(\cdot)$ in the model defined in (2.2) for simplicity. Because $q_{\tau,t-1,i}$ in $\boldsymbol{X}_t$ depends on unknown functional coefficients $\boldsymbol{g}_{i,\tau}(\cdot)$, model (2.2) is more complicated than functional coefficient models with observed data. My procedures consist of two steps. The first is to estimate latent $q_{\tau,t-1,i}$, and then I perform locally weighted estimation for functional coefficients using the estimated $q_{\tau,t-1,i}$ from the first step. In this paper, I only focus on estimating functional coefficients in (2.2), rather than jointly forecasting $q_{\tau,t,i}$ or doing impulse response analysis. So, it is sufficient to estimate $\boldsymbol{g}_{i,\tau}(\cdot)$ in an equation-by-equation way for different i. Thus, by abuse of notation, i will be dropped in what follows.

Given (2.1) and (2.2), let $\boldsymbol{\gamma}_{0,\tau}(Z_t)$ define as earlier as $(\gamma_{10,\tau}(Z_t), \ldots, \gamma_{\kappa 0,\tau}(Z_t))^T$ 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 $s = 1, \ldots, q$ and $l = 1, \ldots, p$. 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 2.2.3, ensuring the invertibility of $\mathcal{B}_\tau(\mathcal{L})$, model (2.1) becomes to the following formulation

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

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 ith row of matrix $C_{0,t,\tau}\boldsymbol{\gamma}_{0,\tau}(Z_t)$ and $\boldsymbol{\alpha}_{l,\tau}(\cdot) = (\alpha_{l1,\tau}(\cdot), \ldots, \alpha_{l\kappa,\tau}(\cdot))^T$ be the ith row of matrix $C_{l,t,\tau}$. Therefore, with the definitions of $\alpha_{0,\tau}(\cdot)$ and $\boldsymbol{\alpha}_{l,\tau}(\cdot)$, I can first approximate the latent $q_{\tau,t}$ by using a functional-coefficient quantile function:

$$q_{\tau,t} = \alpha_{0,\tau}(Z_t) + \sum_{l=1}^{\infty} \boldsymbol{\alpha}_{l,\tau}^T(Z_t)\mathbb{Y}_{t-l}, \tag{2.6}$$

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 $\max_{1\le t\le n} |\alpha_{lj,\tau}(Z_t)| \le c\rho^l$ for $j = 1, \ldots, \kappa$. Since $\alpha_{lj,\tau}(\cdot)$ decreases geometrically, by choosing an appropriate $m_n = m(n) = m$, I study following truncated equation (2.7) with increasing dimension of covariates:

$$q_{\tau,t} = \alpha_{0,\tau}(Z_t) + \sum_{l=1}^{m_n} \boldsymbol{\alpha}_{l,\tau}^T(Z_t)\mathbb{Y}_{t-l} \equiv \boldsymbol{W}_t^T\boldsymbol{\alpha}_\tau(Z_t) = q_\tau(Z_t, \boldsymbol{W}_t), \tag{2.7}$$

where $\boldsymbol{W}_t = (1, \mathbb{Y}_{t-1}^T, \ldots, \mathbb{Y}_{t-m}^T)^T$ is a $(\kappa m + 1) \times 1$ vector of covariates and $\boldsymbol{\alpha}_\tau(\cdot) = (\alpha_{0,\tau}(\cdot), \boldsymbol{\alpha}_{1,\tau}^T(\cdot), \ldots, \boldsymbol{\alpha}_{m,\tau}^T(\cdot))^T$ is a $(\kappa m+1)\times 1$ vector of functional coefficients. Note that (2.7) can be regarded as an approximation of (2.6) 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 2.2.3, 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 approximated 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 rth derivative of $\boldsymbol{\alpha}_\tau(\cdot)$. Thus, $q_{\tau,t} \approx \sum_{r=0}^{w} \boldsymbol{W}_t^T\boldsymbol{\delta}_{r,\tau}(Z_t - z_0)^r$, where $\boldsymbol{\delta}_{r,\tau} = \boldsymbol{\alpha}_\tau^{(r)}(z_0)/r!$. Hence, $\hat{\boldsymbol{\delta}} = \text{argmin}_{\boldsymbol{\delta}} Q(\boldsymbol{\delta})$, where $Q(\boldsymbol{\delta})$ is the locally weighted loss function for fixed κ, given by

$$Q(\boldsymbol{\delta}) = \sum_{t=m+1}^{n} \rho_\tau \left\{ Y_t - \sum_{r=0}^{w} \boldsymbol{W}_t^T\boldsymbol{\delta}_r(Z_t - z_0)^r \right\} K_{h_1}(Z_t - z_0), \tag{2.8}$$

$\rho_\tau(y) = y[\tau - I(y < 0)]$ 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 I smooth locally around Z_t and consider a local linear estimation, the locally weighted loss function (2.8) becomes to the following

$$Q_1(\boldsymbol{\delta}) = \sum_{\mathfrak{s}=m+1\neq t}^{n} \rho_\tau \left\{ Y_{\mathfrak{s}} - \sum_{r=0}^{1} \boldsymbol{W}_{\mathfrak{s}}^T \boldsymbol{\delta}_r (Z_{\mathfrak{s}} - Z_t)^r \right\} K_{h_1}(Z_{\mathfrak{s}} - Z_t). \tag{2.9}$$

After yielding $\hat{\boldsymbol{\delta}}_{0,\tau}$ at τ by minimizing (2.9), $q_{\tau,t}$ can be estimated by $\hat{q}_{\tau,t} = \boldsymbol{W}_t^T \hat{\boldsymbol{\delta}}_{0,\tau}$.

Remark 2.2.4. *(Truncation parameter $m(n)$). Welsh (1989) and He and Shao (2000) studied nonlinear 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 $m_n^2 \log(p_n m_n)/n \to 0$, 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, I am interested in studying varying interdependences among conditional quantiles, rather than determining the optimal number for m. In addition, I focus on estimating (2.7) using kernel-based approaches, which is necessary in order to obtain asymptotic properties for functional coefficients. Under Assumption A10 in Section 2.2.3, 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 2.3 and the empirical analysis in Section 2.4.*

Remark 2.2.5. *It is necessary to emphasize that $\alpha_{0,\tau}(\cdot)$ and each component of $\boldsymbol{\alpha}_{l,\tau}(\cdot)$ in (2.6) 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, I 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 $\alpha_{0,\tau}(Z_t, Z_{t-1}, \ldots, Z_{t-l})$ and*

$\sum_{l=1}^{\infty} \boldsymbol{\alpha}_{l,\tau}^T(Z_t, Z_{t-1}, \ldots, Z_{t-l})\mathbb{Y}_{t-l}$ *in* (2.6) *for notational simplicity.*

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

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 (2.9). Then, latent $q_{\tau,t}$ is approximated by $\hat{q}_{\tau,t} = \boldsymbol{W}_t^T \hat{\boldsymbol{\delta}}_{0,\tau}$.

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

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

$\boldsymbol{g}_\tau(\cdot)$ 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

$$Q_2(\Theta) = \sum_{t=1}^{n} \rho_\tau \left\{ Y_t - \sum_{r=0}^{1} \hat{\boldsymbol{X}}_t^T \Theta_{r,\tau}(Z_t - z_0)^r \right\} K_{h_2}(Z_t - z_0) \tag{2.10}$$

with $\Theta_{r,\tau} = \boldsymbol{g}_\tau^{(r)}(\cdot)/r!$. Similar to (2.9), $K_{h_2}(u) = K(u/h_2)/h_2$ and h_2 is the bandwidth used for this step, which is different from the bandwidth h_1 used in (2.9); see Remark 2.2.6 later in Section 2.2.3 for more discussions. A further improvement can be achieved by applying iteration to the foregoing two-stage procedures.

2.2.3 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 interdependences 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 2.2.2 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* $\boldsymbol{g}_\tau(\cdot)$ *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* $\boldsymbol{W}$ *has an everywhere positive conditional density* $f_{Y|Z,\boldsymbol{W}}(\cdot)$*, which is bounded and satisfies the Lipschitz continuity condition. Here,* $\boldsymbol{W}_t$ *is defined in (2.7). 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* α*-mixing coefficient* $\alpha(t)$ *which satisfies* $\sum_{t=1}^{\infty} t^{\iota} \alpha^{(\delta-2)/\delta}(t) < \infty$ *for some positive real number* $\delta > 2$ *and* $\iota > (\delta-2)/\delta$.

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* $\boldsymbol{B}_n = \frac{1}{n} \sum_{t=m+1}^{n} \boldsymbol{W}_t \boldsymbol{W}_t^T$ *and denote the maximum and minimum eigenvalues of* $\boldsymbol{B}_n$ *as* $\lambda_{\max}(\boldsymbol{B}_n)$ *and* $\lambda_{\min}(\boldsymbol{B}_n)$*. Then,* $\liminf_{n\to\infty} \lambda_{\min}(\boldsymbol{B}_n) > 0$, $\limsup_{n\to\infty} \lambda_{\max}(\boldsymbol{B}_n) < \infty$. *It is assumed that* $E\|\boldsymbol{W}_t\|^{\delta^*} \leq C m^{\delta^*/2}$ *with* $\delta^* > \delta$.

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

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

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

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

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

Remark 2.2.6. *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 α-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 $\boldsymbol{B}_n$ to $E[\boldsymbol{W}_t\boldsymbol{W}_t^T]$, when $\boldsymbol{W}_t$ is α-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 $mh_1 \to 0$ 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 2.2.1; see Cai and Xu (2008) for details on nonparametric quantile regressions models for α-mixing time series.

Before stating the asymptotic behavior of $\hat{\boldsymbol{g}}_\tau(z_0)$ in the following theorem, for notational simplicity, it needs to define some notations. Define

$$\Omega^*(z_0) \equiv E\left[\boldsymbol{X}_t\boldsymbol{X}_t^T f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0, \boldsymbol{X}_t))|Z_t = z_0\right]$$

with $q_\tau(z_0, \boldsymbol{X}_t) = \boldsymbol{g}_\tau^T(z_0)\boldsymbol{X}_t$ and $f_{Y|Z,\boldsymbol{X}}(\cdot)$. In addition, let $\boldsymbol{\Xi}(z_0) \equiv \tau(1-\tau)\nu_0[\Omega(z_0) - H_1(z_0) + H_2(z_0)]$, where $\Omega(z_0) \equiv E[\boldsymbol{X}_t\boldsymbol{X}_t^T|Z_t = z_0]$,

$$H_1(z_0) = E[\boldsymbol{X}_t\boldsymbol{W}_t^T|Z_t = z_0](\boldsymbol{D}^*(z_0))^{-1}\Gamma^T(z_0) + \Gamma(z_0)(\boldsymbol{D}^*(z_0))^{-1}E[\boldsymbol{W}_t\boldsymbol{X}_t^T|Z_t = z_0],$$

$H_2(z_0) = \Gamma(z_0)(\boldsymbol{D}^*(z_0))^{-1}\boldsymbol{D}(z_0)(\boldsymbol{D}^*(z_0))^{-1}\Gamma^T(z_0)$, and $\Gamma(z_0) \equiv E\left\{ f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0, \boldsymbol{X}_t))\boldsymbol{X}_t\boldsymbol{g}_\tau^T(z_0)\,\boldsymbol{\Pi}_t \,\middle|\, Z_t = z_0\right\}$ is a $(2\kappa+1)\times(\kappa m+1)$ matrix, with $\boldsymbol{\Pi}_t^T = (0^T_{1\times(\kappa m+1)}, \boldsymbol{W}_t, \ldots, \boldsymbol{W}_t, 0^T_{\kappa\times(\kappa m+1)})$. Now, the asymptotic normality of $\hat{\boldsymbol{g}}_\tau(z_0)$ is presented in the following theorem with its detailed proof relegated to the Appendix (see Appendix A).

Theorem 2.2.1. *(Asymptotic Normality) Under Assumptions A1–A12, we have*

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

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

Remark 2.2.7. *It is not surprising to see from Theorem 2.2.1 that the asymptotic bias* $h_2^2\mu_2\boldsymbol{g}_\tau^{(2)}(z_0)/2$ *does not depend on* h_1. *Indeed, since the estimation in the step one is undersmoothed 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 nonparametric estimation,* $\boldsymbol{\Xi}(z_0)$ *in the asymptotic variance term contains additional two terms* $H_1(z_0)$ *and* $H_2(z_0)$, *which involve* $\boldsymbol{W}_t$ *in the first step. This formation of asymptotic variance appears because of the fact that* $\hat{\boldsymbol{X}}_t$ *contains* $\hat{q}_{\tau,t-1}$, *which is estimated in the step one of my two-stage approaches and therefore includes information of* $\boldsymbol{W}_t$. *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 2.2.8. *(Bandwidth Selection) Finally, I 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* (2.10), *it is about selecting the bandwidth in the context of estimating the coefficient functions in the quantile regression. Therefore, I recommend the method proposed in Cai and Xu (2008) for selecting* h_2 *in* (2.10), *which is used in our simulation study in Section 2.3.*

2.2.4 Covariance Estimate

For constructing confidence intervals for the estimated functional coefficients in the empirical study, it turns to discussing how to obtain consistent estimator of the asymptotic covariance matrix $\Sigma_\tau(z_0)$. To this end, one needs to estimate $\boldsymbol{D}(z_0)$, $\boldsymbol{D}^*(z_0)$, $\Gamma(z_0)$, $H_1(z_0)$, $H_2(z_0)$, $\Omega(z_0)$ and $\Omega^*(z_0)$ consistently. For this purpose, define $\hat{\boldsymbol{D}}(z_0) = \sum_{t=1}^n \boldsymbol{W}_t \boldsymbol{W}_t^T K_{h_1}(Z_t - z_0)/n$ and $\hat{\boldsymbol{D}}^*(z_0) = \sum_{t=1}^n w_{1t} \boldsymbol{W}_t \boldsymbol{W}_t^T K_{h_1}(Z_t - z_0)/n$, where $w_{1t} = I(\boldsymbol{W}_t^T \hat{\boldsymbol{\alpha}}_\tau(z_0) - \delta_{1n} < Y_t \leq \boldsymbol{W}_t^T \hat{\boldsymbol{\alpha}}_\tau(z_0) + \delta_{1n})/(2\delta_{1n})$ for any $\delta_{1n} \to 0$ as $n \to \infty$. Similar to the proof in Cai and Xu (2008), one can show that $\hat{\boldsymbol{D}}(z_0) = f_z(z_0)\boldsymbol{D}(z_0) + o_p(1)$ and $\hat{\boldsymbol{D}}^*(z_0) = f_z(z_0)\boldsymbol{D}^*(z_0) + o_p(1)$, respectively. Also, let $\boldsymbol{E}_{xw}(z_0) = \sum_{t=1}^n \hat{\boldsymbol{X}}_t \boldsymbol{W}_t^T K_{h_2}(Z_t - z_0)/n$. 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: $\hat{\Gamma}(z_0) = \sum_{t=1}^n w_{2t} \hat{\boldsymbol{X}}_t \hat{\boldsymbol{g}}_\tau^T(z_0) \boldsymbol{\Pi}_t K_{h_2}(Z_t - z_0)/n$, $\hat{H}_1(z_0) = \boldsymbol{E}_{xw}(z_0)(\hat{\boldsymbol{D}}^*(z_0))^{-1}\hat{\Gamma}^T(z_0) + \hat{\Gamma}(z_0)(\hat{\boldsymbol{D}}^*(z_0))^{-1}(\boldsymbol{E}_{xw}(z_0))^T$, $\hat{\Omega}(z_0) = \sum_{t=1}^n \hat{\boldsymbol{X}}_t \hat{\boldsymbol{X}}_t^T K_{h_2}(Z_t - z_0)/n$, $\hat{H}_2(z_0) = \hat{\Gamma}(z_0)(\hat{\boldsymbol{D}}^*(z_0))^{-1} \times \hat{\boldsymbol{D}}(z_0)(\hat{\boldsymbol{D}}^*(z_0))^{-1}\hat{\Gamma}^T(z_0)$, and $\hat{\Omega}^*(z_0) = \sum_{t=1}^n w_{2t} \hat{\boldsymbol{X}}_t \hat{\boldsymbol{X}}_t^T K_{h_2}(Z_t - z_0)/n$, where $w_{2t} = I(\hat{\boldsymbol{g}}_\tau^T(z_0)\hat{\boldsymbol{X}}_t - \delta_{2n} < Y_t \leq \hat{\boldsymbol{g}}_\tau^T(z_0)\hat{\boldsymbol{X}}_t + \delta_{2n})/(2\delta_{2n})$ for any $\delta_{2n} \to 0$. 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, $\hat{\Sigma}_\tau(z_0) = (\hat{\Omega}^*(z_0))^{-1}\hat{\boldsymbol{\Xi}}(z_0)(\hat{\Omega}^*(z_0))^{-1}$ is a consistent estimate of $\Sigma_\tau(z_0)$, where $\hat{\boldsymbol{\Xi}}(z_0) = \tau(1-\tau)\nu_0[\hat{\Omega}(z_0) - \hat{H}_1(z_0) + \hat{H}_2(z_0)]$ is the consistent estimate of $\boldsymbol{\Xi}(z_0)$ with $\hat{\Omega}(z_0)$, $\hat{H}_1(z_0)$ and $\hat{H}_2(z_0)$ given above.

2.3 A Monte Carlo Simulation Study

In this section, I 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, I 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, I 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, I 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 $i = 1$, $\gamma_{i1,\epsilon_{it}}(z) = 0.1\left\{1+\exp(-4z)\right\}^{-1}\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) = (0.1\sin(-0.5\pi z) + 0.1)\chi_{it}$, $\gamma_{i3,\epsilon_{it}}(z) = (0.04z^2)\epsilon_{it}$, $\gamma_{i4,\chi_{it}}(z) = (-0.04z^2 + 0.15)\chi_{it}$, $\beta_{i1}(z) = 0.1\sin(0.5\pi z) + 0.1$, $\beta_{i2}(z) = 0.1\sin^2(z)$, $\beta_{i3}(z) = 0.02\exp(-z)$, and $\beta_{i4}(z) = 0.1\cos^2(z)$. For $i = 2$, $\gamma_{i1,\epsilon_{it}}(z) = (0.1\sin(-0.5\pi z)+0.1)\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) = 0.1\left\{1+\exp(-4z)\right\}^{-1}\chi_{it}$, $\gamma_{i3,\epsilon_{it}}(z) = (-0.04z^2+0.15)\epsilon_{it}$, $\gamma_{i4,\chi_{it}}(z) = (0.04z^2)\chi_{it}$, $\beta_{i1}(z) = 0.1\sin^2(z)$, $\beta_{i2}(z) = 0.1\sin(0.5\pi z)+0.1$, $\beta_{i3}(z) = 0.1\cos^2(z)$, and $\beta_{i4}(z) = 0.02\exp(-z)$. For $i = 3$, $\gamma_{i1,\epsilon_{it}}(z) = 0.1\left\{1+2\,\exp(-2z)\right\}^{-1}\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) =$

$(0.1\sin(-0.6\pi z)+0.1)\chi_{it}$, $\gamma_{i3,\epsilon_{it}}(z) = (0.04z^2)\epsilon_{it}$, $\gamma_{i4,\chi_{it}}(z) = (-0.04z^2+0.15)\chi_{it}$, $\beta_{i1}(z) = 0.1\sin(0.6\pi z)+0.1$, $\beta_{i2}(z) = 0.1\sin^2(z)$, $\beta_{i3}(z) = 0.02\exp(-z)$, and $\beta_{i4}(z) = 0.1\cos^2(z)$. For $i = 4$, $\gamma_{i1,\epsilon_{it}}(z) = (0.1\sin(-0.6\pi z)+0.1)\epsilon_{it}$, $\gamma_{i2,\chi_{it}}(z) = 0.1\left\{1+2\exp(-2z)\right\}^{-1}\chi_{it}$, $\gamma_{i3,\epsilon_{it}}(z) = (-0.04z^2+0.15)\epsilon_{it}$, $\gamma_{i4,\chi_{it}}(z) = (0.04z^2)\chi_{it}$, $\beta_{i1}(z) = 0.1\sin^2(z)$, $\beta_{i2}(z) = 0.1\sin(0.6\pi z)+0.1$, $\beta_{i3}(z) = 0.1\cos^2(z)$, and $\beta_{i4}(z) = 0.02\exp(-z)$. Finally, ε_{it} are mutually i.i.d. from $\mathcal{N}(0,1)$. Thus, for $1 \le i \le 4$, my 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 (2.1) or (2.2) 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 $1 \le i,j \le 4$, with $\Phi(\cdot)$ being the distribution function of the standard normal. Therefore, $\gamma_{i0,\tau}(\cdot)$, $\gamma_{ij,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$ are functions of τ, suggesting different covariate effects at different levels of τ.

To assess the finite sample performance of the proposed nonparametric estimators, I 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

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

where $\gamma_\tau(\cdot)$ can be either $\gamma_{ij,\tau}(\cdot)$ or $\gamma_{i0,\tau}(\cdot)$, both $\hat{\gamma}_\tau(\cdot)$ and $\hat{\beta}_{ij,\tau}(\cdot)$ are local linear quantile estimates of $\gamma_\tau(\cdot)$ and $\beta_{ij,\tau}(\cdot)$, respectively, and $\{z_k = 0.1(k-1)-1.75 : 1 \le k \le n_0 = 36\}$ are the grid points. Also note that in this example, $q_{\tau,t,i} = \sigma_{it}F_\varepsilon^{-1}(\tau) = 0$ when $\tau = 0.5$, which leads the quantile regression problem to be ill-posed so that the results for $\tau = 0.5$ are omitted. Therefore, I only consider τ's level to be 0.05, 0.15, 0.85 and 0.95 and the sample sizes are n = 500, 1500 and 4000. For each setting, I replicate simulation 500 times and compute the median and standard deviation (in parentheses) of 500 MADE values and the results are reported in Tables 2.1-2.4

One can see clearly from Tables 2.1-2.4 that both median and standard deviation of 500 MADE values steadily decrease as the sample size increases for all four values of τ. Moreover, the performances for $\gamma_{i0,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$ at $\tau = 0.15$ and $\tau = 0.85$ 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 $\tau = 0.15$ and $\tau = 0.85$ are not necessarily superior to that for $\tau = 0.05$ and 0.95.

Finally, I illustrate the finite sample performance for the consistent covariance estimation given in Section 2.2.4 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 2.2.4. Then, I compute the average of empirical coverage rates (AECR) of 95% pointwise CI of $\gamma_{ij,\tau}(\cdot)$ and $\beta_{ij,\tau}(\cdot)$ without the asymptotic bias correction for $1 \le i, j \le 4$, defined as,

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

where $se(\hat{\gamma}_{ij,\tau}(\cdot)) = \left[\widehat{Var}(\hat{\gamma}_{ij,\tau}(\cdot))/nh_2\right]^{1/2}$, $I_b\{\gamma_{ij,\tau}(\cdot) \in \hat{\gamma}_{ij,\tau}(\cdot) \pm 1.96 \times se(\hat{\gamma}_{ij,\tau}(\cdot))\}$ 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 bth 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 2.5, for $n = 4000$ and $\tau = 0.05$, 0.15, 0.85 and 0.95. From Table 2.5, 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.

2.4 A Real Example

2.4.1 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, I consider the interaction between current and one-day lagged VaR. In particular, I 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, my network can be written as following equation system:

$$\mathrm{VaR}_{it} = \boldsymbol{\gamma}_{i,\tau}^T(Z_{t-1})\mathrm{VaR}_{t-1}, \quad i = 1, 2, \ldots, \kappa, \tag{2.11}$$

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

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

for $i = 1, 2, \cdots, \kappa$ at a given $\tau \in (0, 1)$. Here, $\mathcal{F}_{i,t-1}$ is the information set to present all information of the ith 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 $\boldsymbol{\gamma}_{i,\tau}(\cdot) = (\gamma_{i1,\tau}(\cdot), \ldots, \gamma_{i\kappa,\tau}(\cdot))^T$ is a $\kappa \times 1$ vector of functional coefficients. Then, I extract the quantile estimation of functional coefficients from equation system (2.11) 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 τ-th quantile level, and is driven by the smoothing variable Z_{t-1}. Here, taking absolute value on each $\hat{\gamma}_{ij,\tau}(Z_{t-1})$ enables us to calculate and analyze indicators of connectedness, and details will be reported in Section 2.4.3 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 (2.3). Therefore, my 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.

2.4.2 Data

My 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. I model the ith 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 π_{it} is ith index level on the tth 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, I 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 (2.11), due to the computational burdens, we only consider risk co-dependences among four major markets' indexes.

As for the smoothing variable Z_t, I choose $Z_t = 10\log(D_t/D_{t-1})$, where D_t is the U.S. dollar index on the tth day and can be downloaded from the Federal Reserve Bank of St. Louis. The U.S. dollar index measures value of U.S. dollar against the currencies of a broad group of major U.S. trading partners, higher values of the index indicate a stronger U.S. dollar. This choice of smoothing variable is reasonable, because the exchange rate has been regarded as an important factor associated with international transmission of risk in many empirical studies. For instance, Menkhoff, Sarno, Schmelling and Schrimpf (2012) discussed the relation between innovations in global foreign exchange volatility and excess returns arising from strategies of carry trade, through which the risk spillover transmits from one country to others. In addition, Yang and Zhou (2017) showed that volatility spillover intensity increases with U.S. dollar depreciation. I 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.

2.4.3 Empirical Results

The empirical analysis in this section includes two steps: First, I estimate $\gamma_{ij,\tau}(Z_{t-1})$ for each market index in the equation system (2.11) under $\tau = 0.05$. Second, I 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 (2.11) varies significantly with Z_{t-1} or not. To this end, I estimate each $\gamma_{ij,\tau}(Z_{t-1})$ and corresponding 95% pointwise confidence intervals with the asymptotic bias ignored. Figure 2.1 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 2.1 represents the estimates of the $\gamma_{ij,\tau}(\cdot)$ for $1 \leq i \leq 4$ and $1 \leq j \leq 4$ in (2.11) 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 2.1, we clearly see that these coefficient functions vary significantly over the interval $[-0.075, 0.075]$, which means that I can not use fixed-coefficient dynamic quantiles models to fit the data.

Next, I consider analyzing the matrix $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$, in which each entry is $|\gamma_{ij,\tau}(Z_{t-1})|$. To simplify notation, Z_{t-1} and τ are dropped from $|\hat{\gamma}_{ij,\tau}(Z_{t-1})|$ and $|\hat{\gamma}_{ji,\tau}(Z_{t-1})|$ in the matrix $|\hat{\mathbf{\Gamma}}_{1,\tau}(Z_{t-1})|$, 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), I 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 2.2 and 2.3 display the corresponding results along the same values of Z_{t-1}, under $\tau = 0.05$, respectively. In Figure 2.2, 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 2.3, the vertical axis measures the total connectedness appeared in international equity markets and the horizontal axises in both figures are the same as those in each panel of Figure 2.1.

Figure 2.2 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 2.3 sheds light on the variation of risk spillover in the global financial market. Observed that in Figure 2.3, 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 2.3 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.

2.5 Conclusion

In this chapter, I 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 my 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 my 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 proposed 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 my econometric model. Second, the asymptotic properties of functional coefficients in my 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 (2.2) is time, then the model in (2.2) 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. I leave these important issues, together with some possible extensions as mentioned earlier in this chapter, as future research topics.

Table 2.1: 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 $1 \leq j \leq 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(γ_{10})	MADE(γ_{20})	MADE(γ_{10})	MADE(γ_{20})	MADE(γ_{10})	MADE(γ_{20})
0.05	0.649 (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)
0.85	0.411 (0.052)	0.414 (0.053)	0.350 (0.030)	0.352 (0.032)	0.310 (0.022)	0.313 (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(γ_{30})	MADE(γ_{40})	MADE(γ_{30})	MADE(γ_{40})	MADE(γ_{30})	MADE(γ_{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)
0.85	0.393 (0.057)	0.414 (0.048)	0.352 (0.032)	0.351 (0.030)	0.306 (0.020)	0.309 (0.021)
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(γ_{11})	MADE(γ_{12})	MADE(γ_{11})	MADE(γ_{12})	MADE(γ_{11})	MADE(γ_{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)
0.85	0.123 (0.057)	0.148 (0.070)	0.093 (0.046)	0.110 (0.047)	0.088 (0.035)	0.103 (0.037)
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(γ_{13})	MADE(γ_{14})	MADE(γ_{13})	MADE(γ_{14})	MADE(γ_{13})	MADE(γ_{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)
0.85	0.119 (0.059)	0.159 (0.075)	0.099 (0.044)	0.118 (0.051)	0.082 (0.035)	0.090 (0.037)
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(γ_{21})	MADE(γ_{22})	MADE(γ_{21})	MADE(γ_{22})	MADE(γ_{21})	MADE(γ_{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)
0.85	0.140 (0.064)	0.114 (0.055)	0.114 (0.048)	0.084 (0.040)	0.097 (0.037)	0.067 (0.032)
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(γ_{23})	MADE(γ_{24})	MADE(γ_{23})	MADE(γ_{24})	MADE(γ_{23})	MADE(γ_{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)
0.85	0.125 (0.064)	0.125 (0.058)	0.105 (0.049)	0.096 (0.044)	0.083 (0.035)	0.082 (0.036)
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 2.2: Simulation results for $\gamma_{ij,\tau}(\cdot)$ for $i = 3, 4$ and for $1 \leq j \leq 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(γ_{31})	MADE(γ_{32})	MADE(γ_{31})	MADE(γ_{32})	MADE(γ_{31})	MADE(γ_{32})
0.05	0.146 (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)
0.85	0.121 (0.055)	0.149 (0.067)	0.095 (0.046)	0.114 (0.047)	0.078 (0.033)	0.106 (0.039)
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(γ_{33})	MADE(γ_{34})	MADE(γ_{33})	MADE(γ_{34})	MADE(γ_{33})	MADE(γ_{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)
0.85	0.122 (0.057)	0.160 (0.074)	0.099 (0.042)	0.118 (0.053)	0.079 (0.037)	0.090 (0.033)
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(γ_{41})	MADE(γ_{42})	MADE(γ_{41})	MADE(γ_{42})	MADE(γ_{41})	MADE(γ_{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)
0.85	0.143 (0.062)	0.110 (0.060)	0.113 (0.046)	0.079 (0.035)	0.097 (0.037)	0.063 (0.029)
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(γ_{43})	MADE(γ_{44})	MADE(γ_{43})	MADE(γ_{44})	MADE(γ_{43})	MADE(γ_{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)
0.85	0.136 (0.066)	0.124 (0.061)	0.107 (0.049)	0.099 (0.040)	0.086 (0.033)	0.080 (0.032)
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 2.3: Simulation results for $\beta_{ij,\tau}(\cdot)$ for $i = 1, 2$ and for $1 \leq j \leq 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(β_{11})	MADE(β_{12})	MADE(β_{11})	MADE(β_{12})	MADE(β_{11})	MADE(β_{12})
0.05	0.215 (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)
0.85	0.134 (0.059)	0.143 (0.074)	0.080 (0.034)	0.088 (0.039)	0.053 (0.023)	0.057 (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(β_{13})	MADE(β_{14})	MADE(β_{13})	MADE(β_{14})	MADE(β_{13})	MADE(β_{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)
0.85	0.130 (0.066)	0.133 (0.072)	0.075 (0.038)	0.082 (0.040)	0.049 (0.022)	0.055 (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(β_{21})	MADE(β_{22})	MADE(β_{21})	MADE(β_{22})	MADE(β_{21})	MADE(β_{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)
0.85	0.135 (0.069)	0.136 (0.072)	0.081 (0.034)	0.084 (0.038)	0.052 (0.019)	0.058 (0.022)
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(β_{23})	MADE(β_{24})	MADE(β_{23})	MADE(β_{24})	MADE(β_{23})	MADE(β_{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)
0.85	0.132 (0.064)	0.130 (0.067)	0.078 (0.035)	0.085 (0.039)	0.052 (0.022)	0.055 (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 2.4: Simulation results for $\beta_{ij,\tau}(\cdot)$ for $i = 3, 4$ and for $1 \leq j \leq 4$.

τ	$n = 500$		$n = 1500$		$n = 4000$	
	MADE(β_{31})	MADE(β_{32})	MADE(β_{31})	MADE(β_{32})	MADE(β_{31})	MADE(β_{32})
0.05	0.218 (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)
0.85	0.133 (0.063)	0.133 (0.064)	0.088 (0.038)	0.083 (0.039)	0.058 (0.024)	0.054 (0.023)
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(β_{33})	MADE(β_{34})	MADE(β_{33})	MADE(β_{34})	MADE(β_{33})	MADE(β_{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)
0.85	0.131 (0.058)	0.134 (0.065)	0.080 (0.035)	0.082 (0.039)	0.050 (0.021)	0.055 (0.023)
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(β_{41})	MADE(β_{42})	MADE(β_{41})	MADE(β_{42})	MADE(β_{41})	MADE(β_{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)
0.85	0.130 (0.066)	0.141 (0.068)	0.081 (0.037)	0.091 (0.037)	0.050 (0.020)	0.057 (0.022)
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(β_{43})	MADE(β_{44})	MADE(β_{43})	MADE(β_{44})	MADE(β_{43})	MADE(β_{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)
0.85	0.130 (0.064)	0.125 (0.063)	0.079 (0.034)	0.079 (0.039)	0.047 (0.023)	0.050 (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 2.5: 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 $n = 4000$.

τ	Coverage Rates of $\hat{\gamma}_{ij,\tau}(\cdot)$				Coverage Rates of $\hat{\beta}_{ij,\tau}(\cdot)$			
	$\hat{\gamma}_{11,\tau}$	$\hat{\gamma}_{12,\tau}$	$\hat{\gamma}_{13,\tau}$	$\hat{\gamma}_{14,\tau}$	$\hat{\beta}_{11,\tau}$	$\hat{\beta}_{12,\tau}$	$\hat{\beta}_{13,\tau}$	$\hat{\beta}_{14,\tau}$
0.05	0.959	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
0.85	0.954	0.943	0.953	0.937	0.943	0.956	0.951	0.949
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
0.85	0.941	0.943	0.943	0.952	0.954	0.953	0.959	0.956
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
0.85	0.950	0.941	0.952	0.942	0.952	0.956	0.949	0.951
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
0.85	0.946	0.947	0.941	0.948	0.957	0.952	0.958	0.955
0.95	0.936	0.947	0.929	0.945	0.944	0.941	0.949	0.942

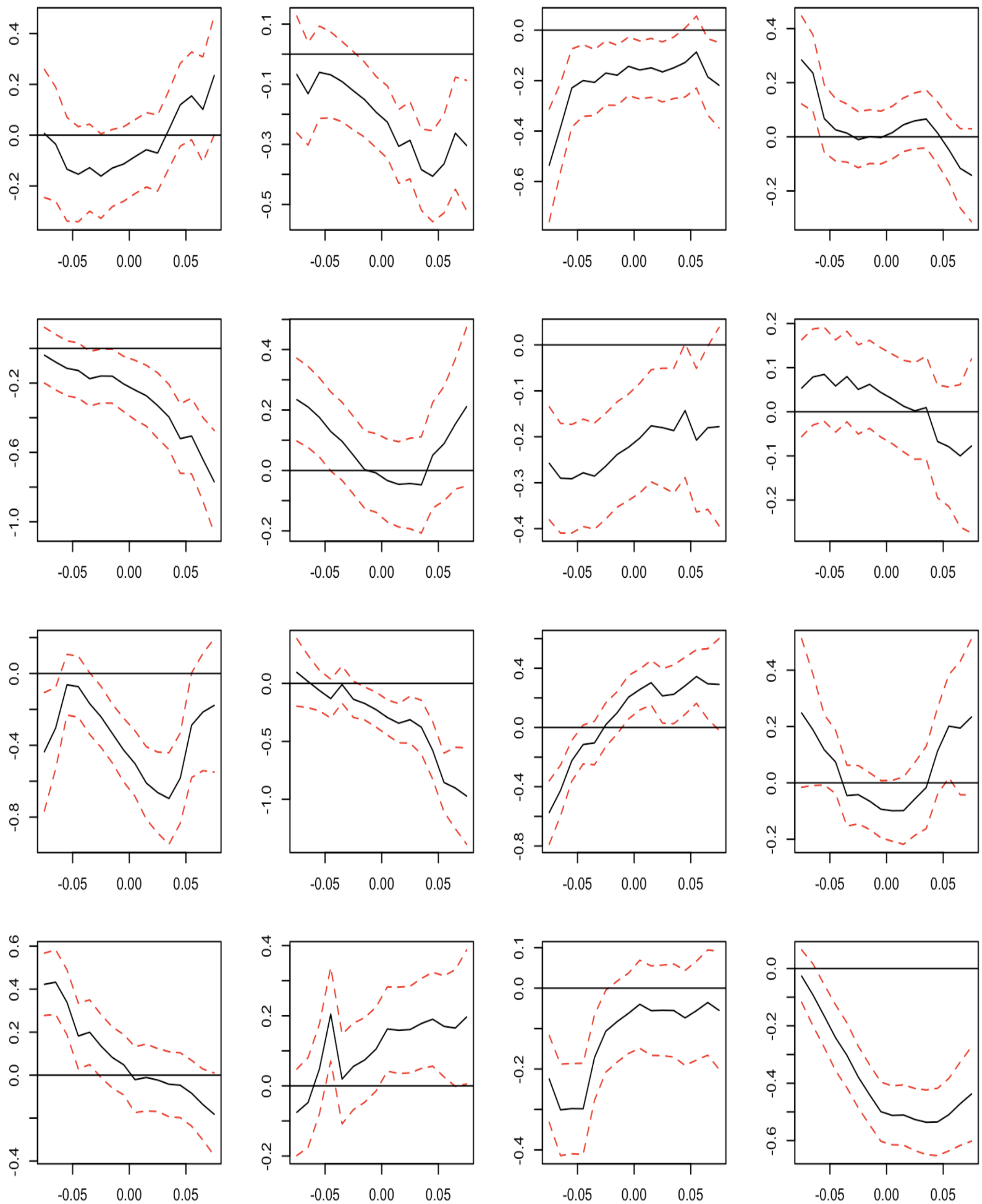

Figure 2.1: Plots of the estimated coefficient functions $\gamma_{ij,\tau}(\cdot)$ for $1 \leq i \leq 4$ and $1 \leq j \leq 4$ in (2.11) in the main article under $\tau = 0.05$ (black solid lines), in which ij-th panel represents the result for $\gamma_{ij,\tau}(\cdot)$, respectively. The red dashed lines in each panel indicate the 95% pointwise confidence interval for the estimate with the asymptotic bias ignored.

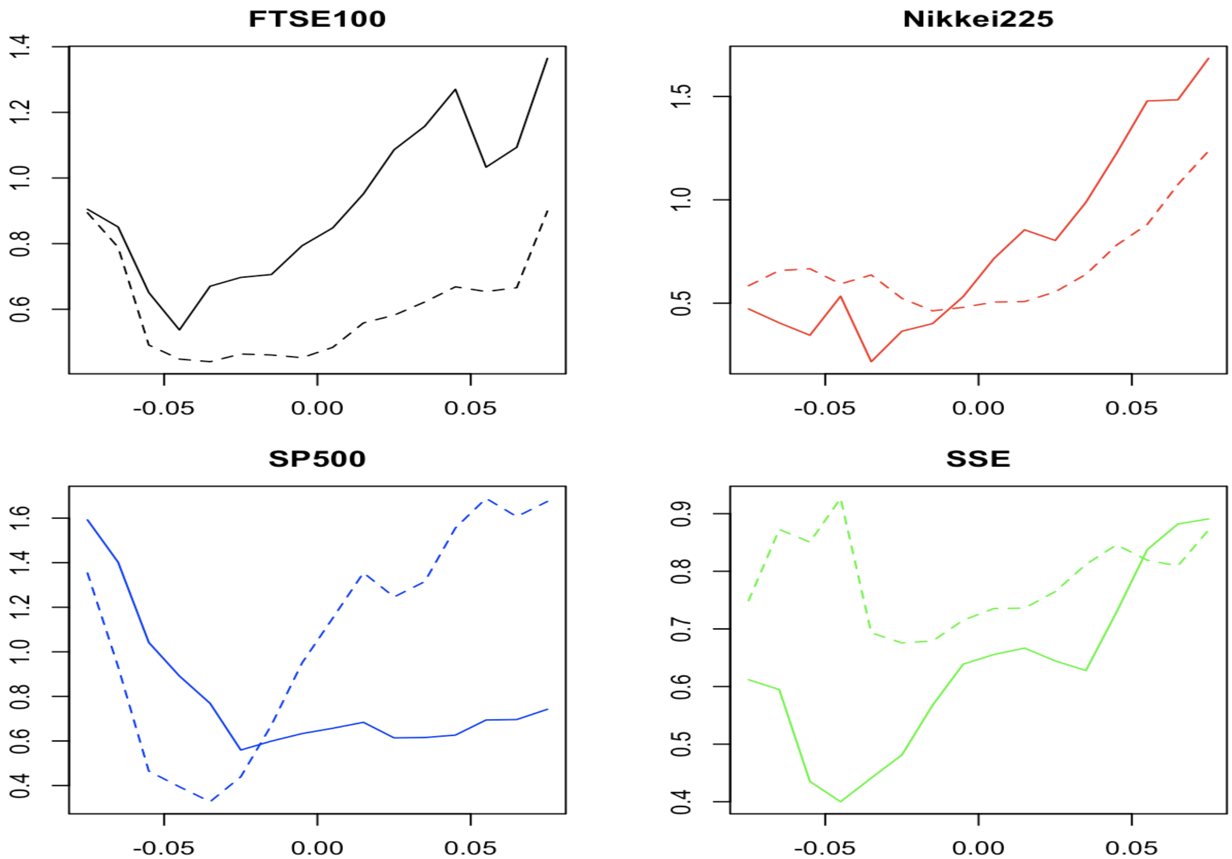

Figure 2.2: Connectedness with respect to outgoing links and connectedness with respect to incoming links for four market indexes with $\tau = 0.05$. The solid line in each panel represents values of connectedness with respect to outgoing links and the dashed line in each panel indicates values of connectedness is for incoming link.

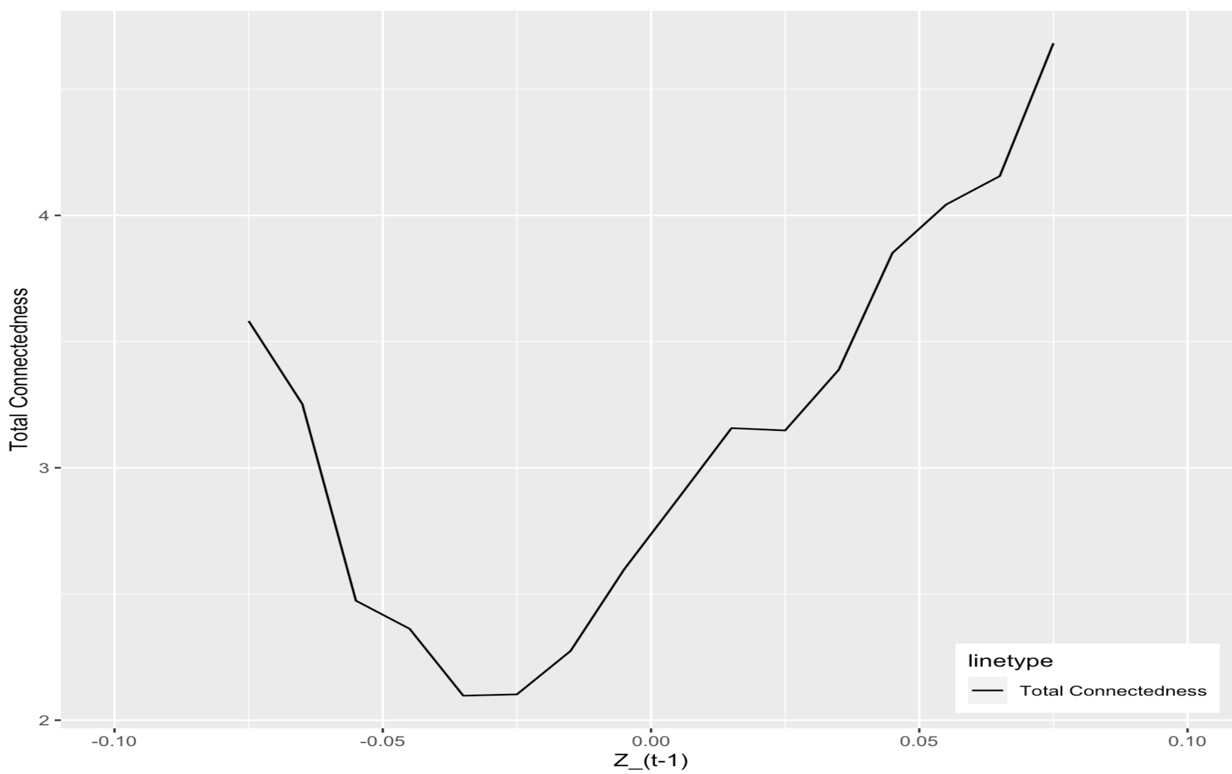

Figure 2.3: Total connectedness in international equity markets with $\tau = 0.05$.

Chapter 3

A Functional Coefficient Factor-Augmented Predictive VAR Model with Dynamic Factor Loadings

3.1 Introduction

In the existing literature, functional-coefficient factor-augmented forecasting models and its variants assumed that latent factors are extracted from a factor model with fixed factor loadings. However, this assumption can be restrictive given that financial and macroeconomic datasets often span a long time period. Indeed, during a long period of time, the interdependences among economic and/or financial variables may be subject to structure changes caused by institutional changes, business cycles, technological advances and preference switching; see, for example, Stock and Watson (2002, 2009), Su and Wang (2017) and Pelger and Xiong (2021) and the references therein. In order to fully capture the nonlinear relationships between these variables and the latent factors, it is natural to allow the factor loadings to vary with a smoothing variable which contains information of economic changes. Failure to consider the structure changes in factor loadings can lead to a misleading estimate for latent factors and consequently, result in unreliable results of forecasting and inference when using factor-augmented forecasting models.

In this chapter, I propose a functional coefficient FAVAR, termed as FC-FAVAR, predictive model (will be presented in (3.3) later) to fill the gaps in literature. Unlike conventional FAVAR and functional coefficient factor-augmented forecasting model, I capture nonlinearities in data by using a functional coefficient setting, where both the factor loadings and coefficients in the predictive model are allowed to vary with a variable of general economy. Actually, as elaborated by Cai, Das, Xiong and Wu (2006) and Cai (2010), a functional-coefficient model can be a good approximation to a fully nonparametric model and has a great ability to capture heteroscedasticity; see Cai (2010) for more details. In the last two decades, the functional-coefficient modeling approach has received much attention on time series studies, to name just a few, Chen and Tsay (1993), Cai, Fan and Yao (2000), Cai (2007), Dahlhaus and Subba Rao (2006), Chen and Hong (2016).

The estimation of this new model relies on a two-step procedure. In step one, a local version of principal component analysis (PCA) as in Su and Wang (2017) and Pelger and Xiong (2021) is applied to estimate unobservable factor regressors and the number of factor is determined by a BIC-type information criterion which is similar to that in Su and Wang (2017). Different from the classical PCA, the high-dimensional dataset is first transformed by a kernel projection on a given point of smoothing variable in the time direction. Then, conventional PCA is used on this projected data to obtain latent factors. The estimated latent factors are next introduced in the second step as parts of auto-regressors and the coefficient functions of FC-FAVAR model are estimated by a local linear approach, which has been throughout discussed in Cai et al. (2000). With the estimated model at hand, I develop an one-step ahead forecasting for the observed auto-regressors and the corresponding prediction interval is constructed by a wild bootstrap procedure proposed in Li et al. (2020).

Contributions of this chapter are two folds. First of all, compared to existing literature, the unobserved factors in this model comes from a factor model with dynamic factor loadings, which adds additional information of variation in the factor structure and economic interpretability to the predictive model. Secondly, this VAR predictive model allows both observed variables and unobserved factor regressors to be jointly imposed in a vector autoregressive form. More specifically, I allow latent factors in this VAR model to also include the same number of lags as that of observed auto-regressors. I think that some important information of model dynamic may be included in these lagged factors, which is helpful to enhance the ability of prediction. These merits will be demonstrated by a simulation study and an empirical application.

The rest of this chapter is organized as follows. In Section 3.2, the model setup is presented for the FC-FAVAR model, and a two-stage procedure for estimating functional coefficients as well as a wild bootstrap procedure for constructing prediction interval are also discussed in this section. A simulation study of one-step ahead forecasting is presented in Section 3.3 to examine the performance of prediction. In Section 3.4, my model is applied to forecasting the consumer price index (CPI) in the U.S.. Section 3.5 concludes the paper. Finally, some assumptions for inducing probabilistic properties of FC-FAVAR model are gathered in Appendix C.

3.2 Econometric Modeling

3.2.1 Functional Coefficient FAVAR Model

Let $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{Nt})^T$ be a $N \times 1$ vector of available predictive variables at time t for $1 \le t \le n$. For $1 \le j \le m$ with $m \ge 1$, consider following factor-augmented forecasting model with functional coefficients

$$y_{j,t} = \gamma_{j0}(Z_{jt}) + \sum_{d=1}^{q_f} \boldsymbol{\gamma}_{j,d,\boldsymbol{f}}^T(Z_{jt}) \boldsymbol{f}_{t-d} + \sum_{c=1}^{q_y} \boldsymbol{\gamma}_{j,c,\boldsymbol{y}}^T(Z_{jt}) \boldsymbol{y}_{t-c} + \sum_{l=1}^{p} \boldsymbol{\beta}_{j,l}^T(Z_{jt}) W_{t-l} + v_{j,t} \tag{3.1}$$

for some q_g, q_f and p, where $\gamma_{j0}(\cdot)$ is a scalar function, $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{m,t})^T$ is a $m \times 1$ vector of observable economic variables that are contained in $\boldsymbol{x}_t$, $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{rt})^T$ is a $r \times 1$ vector of unobservable factors, and W_t is a $\kappa \times 1$ vector of observable covariates, including possibly some or all of $\{y_{j,t}\}_{j=1}^m$ and/or some exogenous variables. In addition, both $\boldsymbol{\gamma}_{j,c,\boldsymbol{y}}(\cdot) = (\gamma_{cj1,\boldsymbol{y}}(\cdot), \ldots, \gamma_{cjm,\boldsymbol{y}}(\cdot))^T$ and $\boldsymbol{\gamma}_{j,d,\boldsymbol{f}}(\cdot) = (\gamma_{dj1,\boldsymbol{f}}(\cdot), \ldots, \gamma_{djr,\boldsymbol{f}}(\cdot))^T$ are $m \times 1$ and $r \times 1$ vectors of functional coefficients, respectively. Finally, $\boldsymbol{\beta}_{j,l}(\cdot) = (\beta_{lj1}(\cdot), \ldots, \beta_{lj\kappa}(\cdot))^T$ is a $\kappa \times 1$ vector of functional coefficients and $v_{j,t}$ is an error term. Here, Z_{jt} is an observable scalar smoothing variable, which might be one part of W_{t-l} and/or time or other exogenous variables or their lagged variables. Of course, Z_{jt} can also be an economic index to characterize economic activities. It is worthwhile to note that Z_{jt} 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_{jt} in very high dimension are often not practically useful due to the curse of dimensionality; see Cai et al. (2000) for details.

Importantly, in the case of estimating high dimensional VAR models with functional coefficients in my empirical studies, I assume that $\boldsymbol{y}_t$ and $\boldsymbol{f}_t$ jointly follow a VAR process. In addition, for easy exposition, I let $p = 0$ and $q_y = q_f \equiv q$, and the smoothing variable Z_{jt} is allowed to vary only across different time periods but keeps constant over individual units. Therefore, model (3.1) can be written as a VAR model with functional coefficients. In particular, by letting $Z_{jt} = Z_t$ for all $1 \le j \le m$, $1 \le \iota \le Q$ and $1 \le \ell \le Q$ with $Q = m + r$, the proposed FC-FAVAR model is

$$\mathbb{P}_t = \boldsymbol{\gamma}_0(Z_t) + \boldsymbol{\Phi}(Z_t)\mathbb{P}_{t-1} + \boldsymbol{\Upsilon}_t, \tag{3.2}$$

where $\mathbb{P}_t = (P_t^T, \ldots, P_{t-q+1}^T)^T$ with $P_t = (\boldsymbol{f}_t^T, \boldsymbol{y}_t^T)^T$, $\boldsymbol{\gamma}_0(\cdot) = (\gamma_{10}(\cdot), \ldots, \gamma_{Q0}(\cdot), 0_{1\times(Qq-Q)})^T$ is a vector of scalar function $\gamma_{\iota 0}(\cdot)$, and $\boldsymbol{\Upsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{Q,t}, 0_{1\times(Qq-Q)})^T$ is a vector of error terms. In addition, $\boldsymbol{\Phi}(Z_t)$ is a functional coefficient matrix and is expressed as follows

$$\boldsymbol{\Phi}(Z_t) = \begin{pmatrix} \Gamma_1(Z_t) & \Gamma_2(Z_t) & \ldots & \Gamma_{q-1}(Z_t) & \Gamma_q(Z_t) \\ I_Q & 0 & \ldots & 0 & 0 \\ 0 & I_Q & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I_Q & 0 \end{pmatrix},$$

where I_Q is a $Q \times Q$ identity matrix and $\Gamma_k(Z_t) = (\gamma_{k\iota\ell,P}(Z_t))_{Q\times Q}$ is a $Q \times Q$ matrix with

$\gamma_{k\iota\ell,P}(\cdot)$ being the functional coefficient for $1 \leq k \leq q$. Notice that process P_t in (3.2) is presented as

$$P_t = \gamma_0(Z_t) + \Gamma_1(Z_t)P_{t-1} + \Gamma_2(Z_t)P_{t-2} + \cdots + \Gamma_q(Z_t)P_{t-q} + \boldsymbol{\varepsilon}_t, \tag{3.3}$$

where $\gamma_0(\cdot) = (\gamma_{10}(\cdot), \ldots, \gamma_{Q0}(\cdot))^T$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{Q,t})^T$.

With models (3.3) at hand, in order to fully capture nonlinear features in a high-dimensional dataset, I further assume that $\boldsymbol{f}_t$ comes from a factor model with dynamic loadings as follow

$$\boldsymbol{x}_t = \boldsymbol{B}_{\boldsymbol{f}}(Z_t)\boldsymbol{f}_t + \boldsymbol{B}_{\boldsymbol{y}}(Z_t)\boldsymbol{y}_t + \boldsymbol{u}_t, \tag{3.4}$$

where $\boldsymbol{f}_t$ is a $r \times 1$ vector of latent factors, $\boldsymbol{B}_{\boldsymbol{f}}(Z_t) = (\boldsymbol{b}_{\boldsymbol{f},1}(Z_t), \ldots, \boldsymbol{b}_{\boldsymbol{f},N}(Z_t))^T$ is a $N \times r$ matrix of dynamic loadings, $\boldsymbol{B}_{\boldsymbol{y}}(Z_t) = (\boldsymbol{b}_{\boldsymbol{y},1}(Z_t), \ldots, \boldsymbol{b}_{\boldsymbol{y},N}(Z_t))^T$ is a $N \times m$ matrix of functional coefficients and $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})^T$ is a $N \times 1$ vector of idiosyncratic errors. Here, Z_t in factor model (3.4) is the same as that in model (3.3). To demonstrate high-dimensional setting, the number N is large and it is commonly assumed to be much greater than the number of factors and observed variables ($m + r \ll N$). In this case, following from Su and Wang (2017) and Pelger and Xiong (2021), I apply a local PCA method for estimating $\boldsymbol{f}_t$ in model (3.4) and the estimation procedures will be presented later. Note that $\boldsymbol{f}_t$ can also be estimated directly through a locally common correlated effect (LCCE) approach proposed in Cai, Fang and Xu (2022) when N is large, which can make computation much easier. As an alternative, the methods of diversified projections (DP) established in Fan and Liao (2020) can also be applied to estimating $\boldsymbol{f}_t$. The estimation procedures using LCCE or DP are also very interesting and I leave them as topics for future research.

Clearly, the model in (3.3) covers many well known models in literature as a special case. In particular, when $\boldsymbol{y}_{t-k}$ in P_{t-k} is assume to have no effect on $\boldsymbol{f}_t$ in P_t for $1 \leq k \leq q$, then the model in (3.3) includes the model in Li et al. (2020). In addition, when $m = 1$ (univariate case) and factor part is excluded, this model nests that in Chen and Tsay (1993), Cai et al. (2000) and Cai (2010), and the model in Cai (2007) for Z_{jt} being time. In addition, if Z_{jt} is time, then model (3.3) is called time-varying FAVAR model, which includes static FAVAR model in Bernanke et al. (2005), Bai et al. (2016) and Yamamoto (2019), as well as the threshold FAVAR model in Yan and Cheng (2022).

Remark 3.2.1. *(Strictly stationary and α-mixing). To apply estimation procedures in this chapter, one has to show that the model given in* (3.2) *can generate strictly stationary and α-mixing process. It is well-established that a geometrically ergodic Markov process initiated from its invariant distribution is (strictly) stationary and α-mixing (Pham, 1986). Notice that model (3.2) can also be expressed as a vector valued Markov model. Thus, it is common practice to prove ergodicity to establish the stationarity for FC-FAVAR models and I present an assumption that induces strictly stationary and α-mixing for process* $\{(P_t, Z_t)\}$ *in Appendix C. Notice that the detailed proof of this stationarity is similar to that in Cai and Liu (2022) and omitted.*

Remark 3.2.2. *(Selection of Z_t). Of importance is to choose an appropriate smoothing variable Z_t in applying the functional-coefficient FAVAR model in* (3.3). *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 3.4 by choosing Z_t to be the monthly series of the first difference of logarithms of consumer price index (CPI) in the U.S.. Without any prior information, it is pertinent to choose Z_t in terms of some data-driven methods such as the Akaike information criterion (AIC), cross-validation (CV), and other criteria. Ideally, one would choose Z_t as a linear function of given explanatory variables according to some optimal criterion or an economic index based on economic theory or background. Nevertheless, here I would recommend using a simple and practical approach proposed by Cai et al. (2000) in practice.*

3.2.2 Two-stage Estimation Procedures

My estimation procedures consist of two steps. The first step is to estimate vector of latent factors $\boldsymbol{f}_t$ in (3.4), and then I perform locally weighted estimation for functional coefficients in (3.3) using the estimated $\hat{\boldsymbol{f}}_t$ from the first step. Compared to the estimation procedures in the existing literature, which relies on classical PCA, the latent factors $\hat{\boldsymbol{f}}_t$ in this chapter are estimated via a local PCA method. In particular, for $1 \leq i \leq N$, (3.4) can be written as

$$x_{it} = \boldsymbol{b}_{\boldsymbol{f},i}^T(Z_t)\boldsymbol{f}_t + \boldsymbol{b}_{\boldsymbol{y},i}^T(Z_t)\boldsymbol{y}_t + u_{it}. \tag{3.5}$$

Now, for $1 \leq \mathfrak{s} \neq \mathfrak{r} \leq n$, I construct the residual $cc_{i\mathfrak{r}} = x_{i\mathfrak{r}} - \hat{\boldsymbol{b}}_{\boldsymbol{y},i}^T(Z_\mathfrak{r})\boldsymbol{y}_\mathfrak{r}$, where $\hat{\boldsymbol{b}}_{\boldsymbol{y},i}(Z_\mathfrak{r})$ is the minimizer of a locally weighted loss function $G(\boldsymbol{b}_{\boldsymbol{y}})$, given by

$$G(\boldsymbol{b}_{\boldsymbol{y}}) = \sum_{i=1}^{N}\sum_{\mathfrak{s}\neq\mathfrak{r}}^{n}[x_{i\mathfrak{s}} - \boldsymbol{b}_{\boldsymbol{y},i}^T(Z_\mathfrak{r})\boldsymbol{y}_\mathfrak{s}]^2 K_{h_1}(Z_\mathfrak{s} - Z_\mathfrak{r}). \tag{3.6}$$

Here, $K(\cdot)$ is a kernel function and $K_{h_1}(u) = K(u/h_1)/h_1$ with $h_1 = h_1(n)$ being a sequence of positive numbers tending to zero and controls the amount of smoothing used in estimation.

Next, given $cc_{i\mathfrak{r}}$, for $1 \leq \mathfrak{r} \neq t \leq n$, $\boldsymbol{b}_{\boldsymbol{f},i}(Z_\mathfrak{r})$ can be approximated at each $Z_\mathfrak{t}$ as $\boldsymbol{b}_{\boldsymbol{f},i}(Z_\mathfrak{r}) \approx \boldsymbol{b}_{\boldsymbol{f},i}(Z_t)$, when $Z_\mathfrak{r} \approx Z_t$. Then, it follows that

$$cc_{i\mathfrak{r}} \approx \boldsymbol{b}_{\boldsymbol{f},i}^T(Z_t)\boldsymbol{f}_\mathfrak{r} + u_{i\mathfrak{r}}, \tag{3.7}$$

when $Z_\mathfrak{r} \approx Z_t$. To estimate $\{\boldsymbol{b}_{\boldsymbol{f},i}(Z_t)\}_{i=1}^N$ and $\{\boldsymbol{f}_\mathfrak{r}\}_{\mathfrak{r}=1}^n$, I apply local PCA by solving following locally weighted loss function

$$\min_{\{\boldsymbol{b}_{\boldsymbol{f},i}(Z_t)\}_{i=1}^N, \{\boldsymbol{f}_\mathfrak{r}\}_{\mathfrak{r}=1}^n} (Nn)^{-1}\sum_{i=1}^{N}\sum_{\mathfrak{r}\neq t}^{n}[cc_{i\mathfrak{r}} - \boldsymbol{b}_{\boldsymbol{f},i}^T(Z_t)\boldsymbol{f}_\mathfrak{r}]^2 K_{h_1}(Z_\mathfrak{r} - Z_t), \tag{3.8}$$

subject to certain identification restrictions to be specified later.

As argued by Su and Wang (2017), the solution of (3.8) can be obtained via a conventional PCA method. Indeed, multiplying both sides of (3.7) by $k_{h_1,\mathfrak{r}t}^{1/2} \equiv K_{h_1}^{1/2}(Z_\mathfrak{r} - Z_t)$ yields

$$k_{h_1,\mathfrak{r}t}^{1/2} cc_{i\mathfrak{r}} \approx \boldsymbol{b}_{\boldsymbol{f},i}^T(Z_t) k_{h_1,\mathfrak{r}t}^{1/2} \boldsymbol{f}_\mathfrak{r} + k_{h_1,\mathfrak{r}t}^{1/2} u_{i\mathfrak{r}}. \tag{3.9}$$

Define the $n \times N$ matrices $cc^{(t)} = (cc_1^{(t)}, \ldots, cc_N^{(t)})$ and $u^{(t)} = (u_1^{(t)}, \ldots, u_N^{(t)})$, where $cc_i^{(t)} = (k_{h_1,\mathfrak{r}1}^{1/2} cc_{i1}, \ldots, k_{h_1,\mathfrak{r}n}^{1/2} cc_{in})^T$ and $u_i^{(t)} = (k_{h_1,\mathfrak{r}1}^{1/2} u_{i1}, \ldots, k_{h_1,\mathfrak{r}n}^{1/2} u_{in})^T$. Let

$$\boldsymbol{f}^{(t)} = (k_{h_1,\mathfrak{r}1}^{1/2} \boldsymbol{f}_1, \ldots, k_{h_1,\mathfrak{r}n}^{1/2} \boldsymbol{f}_n)^T.$$

Therefore, (3.9) can be written in following matrix notation:

$$cc^{(t)} \approx \boldsymbol{f}^{(t)} \boldsymbol{B}_{\boldsymbol{f}}^T(Z_t) + u^{(t)},$$

where $\boldsymbol{B}_{\boldsymbol{f}}(Z_t) = (\boldsymbol{b}_{\boldsymbol{f},1}(Z_t), \ldots, \boldsymbol{b}_{\boldsymbol{f},N}(Z_t))^T$ is a $N \times r$ matrix of functional loadings. Then, the minimization problem (3.8) becomes

$$\min_{\boldsymbol{B}_{\boldsymbol{f}}(Z_t), \boldsymbol{f}^{(t)}} tr\left[(cc^{(t)} - \boldsymbol{f}^{(t)} \boldsymbol{B}_{\boldsymbol{f}}^T(Z_t))(cc^{(t)} - \boldsymbol{f}^{(t)} \boldsymbol{B}_{\boldsymbol{f}}^T(Z_t))^T\right]. \tag{3.10}$$

Under the identification restrictions that $\boldsymbol{f}^{(t)T} \boldsymbol{f}^{(t)}/n = I_r$ and $\boldsymbol{B}_{\boldsymbol{f}}^T(Z_t) \boldsymbol{B}_{\boldsymbol{f}}(Z_t)$ is a diagonal matrix, the estimated factor matrix, denoted by $\hat{\boldsymbol{f}}^{(t)} = (\hat{\boldsymbol{f}}_1^{(t)}, \ldots, \hat{\boldsymbol{f}}_n^{(t)})^T$, is $\sqrt{n}$ times eigenvectors corresponding to the r largest eigenvalues of the $n \times n$ matrix $cc^{(t)} cc^{(t)T}$, arranged in descending order, and $\hat{\boldsymbol{B}}_{\boldsymbol{f}}^T(Z_t) = (\hat{\boldsymbol{b}}_{\boldsymbol{f},1}(Z_t), \ldots, \hat{\boldsymbol{b}}_{\boldsymbol{f},N}(Z_t))^T = \hat{\boldsymbol{f}}^{(t)T} cc^{(t)}/n$ are the estimators of the corresponding functional factor loadings. Notice that the estimated factor $\hat{\boldsymbol{f}}_\mathfrak{r}^{(t)}$ is only consistent for a rotational version of the weighted factor $k_{h,\mathfrak{r}t}^{1/2} \boldsymbol{f}_\mathfrak{r}$ for $1 \leq \mathfrak{r} \leq n$. To obtain a consistent estimator of $\boldsymbol{f}_t$, following the idea in Su and Wang (2017), I consider the least squares problem based on $\hat{\boldsymbol{b}}_{\boldsymbol{f},i}(Z_t)$ as follow

$$\min_{\boldsymbol{f}_t \in \mathbb{R}^r} \sum_{i=1}^{N} [cc_{it} - \hat{\boldsymbol{b}}_{\boldsymbol{f},i}^T(Z_t)\boldsymbol{f}_t]^2,$$

for $1 \leq t \leq n$. The solution to the above problem is the consistent estimator of updated factors $\boldsymbol{f}_t$, which is

$$\hat{\boldsymbol{f}}_t = \left(\sum_{i=1}^{N} \hat{\boldsymbol{b}}_{\boldsymbol{f},i}(Z_t)\hat{\boldsymbol{b}}_{\boldsymbol{f},i}^T(Z_t) \right)^{-1} \left(\sum_{i=1}^{N} \hat{\boldsymbol{b}}_{\boldsymbol{f},i}(Z_t)cc_{it} \right). \tag{3.11}$$

If necessary, the above procedure should be repeated.

After obtaining the estimated $\hat{\boldsymbol{f}}_t$ in (3.3) and given $\hat{P}_t = (\hat{\boldsymbol{f}}_t^T, \boldsymbol{y}_t^T)^T$, the second step follows from estimating (3.3) by a local linear approach, although a general local polynomial method is also applicable. The local (polynomial) linear method has been widely used in nonparametric regression during the past two decades due to its attractive mathematical efficiency, bias reduction, and adaptation of edge effects; see, for example, Cai et al. (2000). More specifically, let $\boldsymbol{\Gamma}(\cdot) = (\gamma_0(\cdot), \Gamma_1(\cdot), \ldots, \Gamma_q(\cdot))$ and by assuming that each entry $\gamma_{k\iota\ell,P}(\cdot)$ of matrix $\Gamma_k(\cdot)$ has a continuous second derivative, $\boldsymbol{\Gamma}(Z_t)$ can be approximated by a linear function at grid point $z_0 \in \mathbb{R}$ as follows

$$\mathrm{vec}[\boldsymbol{\Gamma}(Z_t)] \approx \mathrm{vec}[\boldsymbol{\Gamma}(z_0)] + \mathrm{vec}[\boldsymbol{\Gamma}^{(1)}(z_0)](Z_t - z_0),$$

where $\mathrm{vec}(\cdot)$ stacks the elements of a $m \times n$ matrix as a $mn \times 1$ vector, $\approx$ denotes the first-order Taylor approximation and $\boldsymbol{\Gamma}^{(1)}(\cdot)$ is the first-order derivative of each element of $\boldsymbol{\Gamma}(\cdot)$. Thus, (3.3) is approximated by

$$\hat{P}_t \approx \hat{\boldsymbol{P}}_t^{*T} \boldsymbol{\theta}(z_0) + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\theta}(z_0) = \begin{pmatrix} \mathrm{vec}[\boldsymbol{\Gamma}(z_0)] \\ \mathrm{vec}[\boldsymbol{\Gamma}^{(1)}(z_0)] \end{pmatrix}$ and $\hat{\boldsymbol{P}}_t^* = \begin{pmatrix} \hat{\boldsymbol{P}}_t \\ (Z_t - z_0)\hat{\boldsymbol{P}}_t \end{pmatrix}$ with $\hat{\boldsymbol{P}}_t = \hat{\mathcal{P}}_t \otimes I_Q$ $\equiv (1, \hat{P}_{t-1}^T, \ldots, \hat{P}_{t-q}^T)^T \otimes I_Q$, which becomes a local linear model. Therefore, the locally weighted sum of squares is

$$\sum_{t=1}^{n}[\hat{P}_t - \hat{\boldsymbol{P}}_t^{*T}\boldsymbol{\theta}(z_0)]^T[\hat{P}_t - \hat{\boldsymbol{P}}_t^{*T}\boldsymbol{\theta}(z_0)]K_{h_2}(Z_t - z_0), \tag{3.12}$$

where $K_h(u) = K(u/h)/h$ and h_2 is the bandwidth used for this step, which is different from the bandwidth h_1 used in (3.6) and (3.8) for under-smoothing purpose. By minimizing (3.12) with respect to $\boldsymbol{\theta}(z_0)$, I obtain the local linear estimate of $\boldsymbol{\Gamma}(z_0)$, denoted by $\hat{\boldsymbol{\Gamma}}(z_0)$, consisting of the first $(qQQ + Q)$ elements of $\hat{\boldsymbol{\theta}}(z_0)$, and the local linear estimator of the derivative of $\boldsymbol{\Gamma}(z_0)$, denoted by $\hat{\boldsymbol{\Gamma}}^{(1)}(z_0)$, consisting of the last $(qQQ + Q)$ elements of $\hat{\boldsymbol{\theta}}(z_0)$. The estimator of $\boldsymbol{\theta}(z_0)$ is then sequentially given by

$$\hat{\boldsymbol{\theta}}(z_0) = \left(\sum_{t=1}^{n}\hat{\boldsymbol{P}}_t^{*}\hat{\boldsymbol{P}}_t^{*T}K_{h_2}(Z_t - z_0)\right)^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{P}}_t^{*}\hat{P}_tK_{h_2}(Z_t - z_0). \tag{3.13}$$

By moving z_0 over the whole range of the data $\{Z_t\}$, the estimated curve of $\boldsymbol{\theta}(z_0)$ can be obtained.

In practical implementations of (3.12), there are some practical issues that need to be addressed. First, to obtain $\hat{\boldsymbol{\theta}}(z_0)$, one indeed needs to run a weighted least squares regression. Second, the number of factors r and lags q are selected by minimizing some well known criteria such as the nonparametric Bayesian information criterion proposed in Li et al. (2020) or the nonparametric AIC in Cai and Tiwari (2000). Finally, given the selected $\hat{r}$ and $\hat{q}$, I choose the optimal bandwidth h based on some bandwidth selectors such as the modified multifold cross-validation criterion developed in Cai et al. (2000) or the nonparametric AIC type criterion in Cai and Tiwari (2000), which are attentive to the structure of stationary time series data.

Remark 3.2.3. *(Asymptotics) Notice that the asymptotic theory for $\hat{\boldsymbol{\Gamma}}(z_0)$ can be obtained by following the ideas in Cai et al. (2006) and Li et al. (2020) and it may not be the exactly same as that in Cai et al. (2000) because P_t contains vector of latent factors $\boldsymbol{f}_t$. It would* *be very interesting to investigate the asymptotic theory for $\hat{\mathbf{\Gamma}}(z_0)$ and sequentially for the impulse response functions, which is not a trivial task. It is conjectured that the asymptotic variance of $\hat{\mathbf{\Gamma}}(z_0)$ might have an additional term to account for variability of the estimated latent factors at the first step. I leave this theoretical justification as a future research topic.*

Remark 3.2.4. *(Determination of the number of factors) The number of factor $\hat{r}$ is determined by a BIC-type information criterion proposed in Su and Wang (2017). In particular, following normalization rule that $\hat{\boldsymbol{B}}_{\boldsymbol{f}}^T(Z_t)\hat{\boldsymbol{B}}_{\boldsymbol{f}}(Z_t)/N = I_r$ and $\boldsymbol{f}^{(t)T}\boldsymbol{f}^{(t)}/n$ is a diagonal matrix with descending diagonal elements, denote $\hat{\boldsymbol{f}}_t(r)$ and $\boldsymbol{b}_{\boldsymbol{f},i}(Z_t, r)$ as the local PCA estimators of the factors and factor loadings when using r factors in the factor model. Define*

$$V(r, \{\breve{\boldsymbol{B}}_{\boldsymbol{f}}^{(r)}(Z_t)\}) = \min_{\breve{\boldsymbol{f}}=\{\breve{\boldsymbol{f}}_t\}_{t=1}^n} (Nn)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{n} [cc_{it} - \breve{\boldsymbol{b}}_{\boldsymbol{f},i}^T(Z_t, r)\breve{\boldsymbol{f}}_t]^2,$$

where $\breve{\boldsymbol{B}}_{\boldsymbol{f}}^{(r)}(Z_t) = (\breve{\boldsymbol{b}}_{\boldsymbol{f},1}(Z_t, r), \ldots, \breve{\boldsymbol{b}}_{\boldsymbol{f},N}(Z_t, r))^T = (Nn)^{-1}cc^{(t)T}cc^{(t)}\hat{\boldsymbol{B}}_{\boldsymbol{f}}^{(r)}(Z_t)$, with $\hat{\boldsymbol{B}}_{\boldsymbol{f}}^{(r)}(Z_t) = (\hat{\boldsymbol{b}}_{\boldsymbol{f},1}(Z_t, r), \ldots, \hat{\boldsymbol{b}}_{\boldsymbol{f},N}(Z_t, r))^T$. Then, the number of factor r is determined by minimizing following BIC-type information criterion

$$BIC(r) = \log V(r, \{\breve{\boldsymbol{B}}_{\boldsymbol{f}}^{(r)}(Z_t)\}) + \rho_{Nn} r,$$

where ρ_{Nn} satisfies $\rho_{Nn} \to 0$ and $\rho_{Nn}(\min\{\sqrt{nh_1}, \sqrt{N}\})^2 \to \infty$. In practice, I choose $\rho_{Nn} = \frac{N+nh_1}{Nnh_1}\log\left(\frac{Nnh_1}{N+nh_1}\right)$ as suggested in Su and Wang (2017).

3.2.3 One-Step Ahead Prediction and a Bootstrap Prediction Interval (BPI)

The focus in this subsection is on presenting one-step ahead prediction and bootstrap prediction interval (BPI) below. Denote $\hat{P}_t = (\hat{\boldsymbol{f}}_t^T, \boldsymbol{y}_t^T)^T$, using the observations $(\boldsymbol{y}_{t+1}, Z_t, \boldsymbol{x}_{t+1})$

with $1 \le t \le n-1$ and Z_n, I obtain the one-step ahead prediction of $\hat{P}_{n+1} = (\hat{\boldsymbol{f}}^T_{n+1}, \boldsymbol{y}^T_{n+1})^T$:

$$\hat{P}_{n+1|n} = \hat{\boldsymbol{P}}^T_{n+1} \hat{\boldsymbol{\theta}}_{0,n-1}(Z_n), \tag{3.14}$$

where $\hat{\boldsymbol{P}}_{n+1} = (1, \hat{P}^T_n, \ldots, \hat{P}^T_{n+1-q})^T \otimes I_Q$ and $\hat{\boldsymbol{\theta}}_{0,n-1}(Z_n) = \text{vec}[\hat{\boldsymbol{\Gamma}}(Z_n)]$ is the local linear estimate using the sample $(\boldsymbol{y}_{t+1}, Z_t, \boldsymbol{x}_{t+1})$ with $1 \le t \le n-1$. Denote $\hat{\boldsymbol{y}}_{n+1|n}$ as the last m elements in $\hat{P}_{n+1|n}$. Similar to the procedure in Li et al. (2020), I present a wild bootstrap procedure to construct the prediction interval as follow:

1. Using the observations $(\boldsymbol{y}_{t+1}, Z_t, \boldsymbol{x}_{t+1})$ with $1 \le t \le n-1$, estimate the coefficient functions $\boldsymbol{\theta}_{0,n-1}(Z_t) = \text{vec}[\boldsymbol{\Gamma}(Z_t)]$ by the proposed two-step estimation procedure, and denote the resulting estimates by $\hat{\boldsymbol{\theta}}_{0,n-1}(Z_t) = \text{vec}[\hat{\boldsymbol{\Gamma}}(Z_t)]$.

2. Generate bootstrap sample: $\hat{P}^*_{t+1} = \hat{\boldsymbol{P}}^T_{B,t+1} \hat{\boldsymbol{\theta}}_{0,n-1}(Z_t) + \boldsymbol{\varepsilon}^*_{t+1}$, where $\hat{\boldsymbol{P}}_{B,t+1} = (1, \hat{P}^{*T}_t, \ldots, \hat{P}^{*T}_{t+1-q})^T \otimes I_Q$ and $\boldsymbol{\varepsilon}^*_{t+1} = \tilde{\boldsymbol{\varepsilon}}_{t+1} \cdot \eta_{t+1}$. Here, $\{\eta_{t+1}\}$ is a sequence of i.i.d. random variables drawn from a standard normal distribution and $\tilde{\boldsymbol{\varepsilon}}_{t+1} = \hat{\boldsymbol{\varepsilon}}_{t+1} - \bar{\hat{\boldsymbol{\varepsilon}}}_{t+1}$, where $\bar{\hat{\boldsymbol{\varepsilon}}}_{t+1} = (n-1)^{-1} \sum_{t=1}^{n-1} \hat{\boldsymbol{\varepsilon}}_{t+1}$ and $\hat{\boldsymbol{\varepsilon}}_{t+1} = \hat{P}_{t+1} - \hat{\boldsymbol{P}}^T_{t+1} \hat{\boldsymbol{\theta}}_{0,n-1}(Z_t)$.

3. For $1 \le t \le n-1$, using the bootstrap sample $\hat{P}^*_{t+1}$ and $\hat{\boldsymbol{P}}_{B,t+1}$ generated in Step 2 to re-estimate the coefficient functions at Z_t, and denote the resulting estimators as $\boldsymbol{\theta}^*_{0,n-1}(Z_t)$. Construct the one-step ahead forecast: $\hat{P}_{n+1|n}(1) = \hat{\boldsymbol{P}}^T_{B,n+1} \boldsymbol{\theta}^*_{0,n-1}(Z_t)$.

4. Repeat Steps 2 and 3 for B times and obtain B bootstrap one-step ahead predicted values $\hat{P}_{n+1|n}(b)$ for $b = 1, \ldots, B$. Denote $\hat{\boldsymbol{y}}_{n+1|n}(b)$ as the last m elements in $\hat{P}_{n+1|n}(b)$ and $\widehat{Var}^*(\hat{\boldsymbol{y}}_{n+1|n})$ as the sample variance of $\{\hat{\boldsymbol{y}}_{n+1|n}(b) : b = 1, \ldots, B\}$.

5. For each $b = 1, \ldots, B$, use the sequence $\hat{\boldsymbol{y}}_{n+1|n}(b)$ and $\widehat{Var}^*(\hat{\boldsymbol{y}}_{n+1|n})$ to compute $s^*_n(b) = [\hat{\boldsymbol{y}}_{n+1|n}(b) - \hat{\boldsymbol{y}}_{n+1|n}] / \sqrt{\widehat{Var}^*(\hat{\boldsymbol{y}}_{n+1|n})}$.

6. Compute a sequence of absolute value of $s^*_n(b)$ as $\{|s^*_n(b)| : b = 1, \ldots, B\}$ and construct a $100(1-\alpha)\%$ symmetric percentile-t bootstrap interval for $\hat{\boldsymbol{y}}_{n+1|n}$:

$$\left[\hat{\boldsymbol{y}}_{n+1|n} - q^*_{1-\alpha} \sqrt{\widehat{Var}^*(\hat{\boldsymbol{y}}_{n+1|n})},\ \hat{\boldsymbol{y}}_{n+1|n} + q^*_{1-\alpha} \sqrt{\widehat{Var}^*(\hat{\boldsymbol{y}}_{n+1|n})} \right],$$

where $q^*_{1-\alpha}$ is the empirical $1-\alpha$ quantile of $\{|s^*_n(b)| : b = 1, \ldots, B\}$.

3.3 A Monte Carlo Simulation Study for Forecasting

In this section, I provide a simulation example to exam the performance of one-step ahead forecasting of proposed predictive VAR model. 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, I use a plug-in method as in Sheather and Jones (1991) 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, I 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 under-smoothing assumption. At step two, I choose optimal bandwidth $\hat{h}_2$ by a modified multifold cross-validation criterion developed in Cai et al. (2000). Finally, the Epanechnikov kernel $K(x) = 0.75(1-x^2)I(|x| \leq 1)$ is used.

In this example, the data are generated from:

$$P_{t+1} = \gamma_0(Z_t) + \Gamma_1(Z_t)P_t + \boldsymbol{\varepsilon}_{t+1},$$

and

$$x_{it} = \boldsymbol{b}^T_{\boldsymbol{f},i}(Z_t)\boldsymbol{f}_t + \boldsymbol{b}^T_{\boldsymbol{y},i}(Z_t)\boldsymbol{y}_t + u_{it}.$$

for $1 \leq i \leq N$. Here, $P_t = (\boldsymbol{f}^T_t, \boldsymbol{y}^T_t)^T$, with $\boldsymbol{f}_t = (f_{1t}, f_{2t})^T$ and $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})^T$. In addition, $\boldsymbol{\varepsilon}_t = (\varepsilon_{11,t}, \varepsilon_{12,t}, \varepsilon_{21,t}, \varepsilon_{22,t})^T$, with each component being mutually i.i.d. from $\mathcal{N}(0,1)$. Furthermore, u_{it} are mutually i.i.d. from $\mathcal{N}(0,1)$ and $Z_t = \Phi(Z^*_t)$, with $\Phi(\cdot)$ being the cumulative standard normal distribution function. The initial value Z^*_t is generated from an autoregressive process $Z^*_t = 0.15Z^*_{t-1} + \xi_t$, where ξ_t is generated from an i.i.d. standard normal distribution. For functional coefficient matrices, $\gamma_0(Z_t) = (\gamma_{10}(Z_t), \ldots, \gamma_{40}(Z_t))^T$, $\Gamma_1(Z_t) = (\gamma_{\iota\ell,P}(Z_t))_{4\times 4}$, where $\gamma_{10}(z) = 0.1\sin(-z)$, $\gamma_{20}(z) = 0.1\cos(-z)$, $\gamma_{30}(z) = \sin(-6z)$,

$\gamma_{40}(z) = \cos(-6z)$. In addition, for $\iota = 1$, $\gamma_{\iota 1,P}(z) = 0.1z$, $\gamma_{\iota 2,P}(z) = -0.1z + 0.1$, $\gamma_{\iota 3,P}(z) = 0.1\sin(z)$, $\gamma_{\iota 4,P}(z) = 0.1\cos(z)$. For $\iota = 2$, $\gamma_{\iota 1,P}(z) = -0.1z + 0.1$, $\gamma_{\iota 2,P}(z) = 0.1z$, $\gamma_{\iota 3,P}(z) = 0.1\cos(z)$, $\gamma_{\iota 4,P}(z) = 0.1\sin(z)$. For $\iota = 3$, $\gamma_{\iota 1,P}(z) = \sin(6z)$, $\gamma_{\iota 2,P}(z) = \cos(6z)$, $\gamma_{\iota 3,P}(z) = 0.2I(z \leq 0.5) - 0.2I(z > 0.5)$, $\gamma_{\iota 4,P}(z) = -\sin(6z)$. For $\iota = 4$, $\gamma_{\iota 1,P}(z) = \cos(6z)$, $\gamma_{\iota 2,P}(z) = \sin(6z)$, $\gamma_{\iota 3,P}(z) = 0.3I(z \leq 0.5) + 0.2I(z > 0.5)$, $\gamma_{\iota 4,P}(z) = -\cos(6z)$. As for functional factor loadings, $\boldsymbol{b}_{\boldsymbol{f},i}^T(Z_t) = (b_{\boldsymbol{f},1i}(Z_t), b_{\boldsymbol{f},2i}(Z_t))^T$ and $\boldsymbol{b}_{\boldsymbol{y},i}^T(Z_t) = (b_{\boldsymbol{y},1i}(Z_t), b_{\boldsymbol{y},2i}(Z_t))^T$, where:

$$
\begin{aligned}
b_{\boldsymbol{f},1i}(z) &= 2(1 + \exp(-6(z - \mu_{1i}))^{-1}, \\
b_{\boldsymbol{f},2i}(z) &= 2(1 + \exp(-5(z + 5(i/N) + 2)))^{-1}, \\
b_{\boldsymbol{y},1i}(z) &= (1 + \exp(-2(z - \mu_{2i}))^{-1}, \\
b_{\boldsymbol{y},2i}(z) &= (1 + \exp(-(z - 5(i/N) + 2)))^{-1},
\end{aligned}
$$

with μ_{1i} and μ_{2i} being mutually i.i.d. from $\mathcal{N}(0,1)$. Finally, let $n_0 = n - 1 - \lfloor 0.05n \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function. Given the generated samples $(P_{t+1}, Z_t, \boldsymbol{x}_{t+1})$ with $t = 1, \ldots, n_0 - 1$ and Z_{n_0}, the one-step ahead forecast $\hat{P}_{n_0+1|n_0}$ in this simulation study is constructed using (3.14). Denote $\hat{\boldsymbol{y}}_{n_0+1|n_0} = (\hat{y}_{1,n_0+1|n_0}, \hat{y}_{2,n_0+1|n_0})^T$ as the last two elements in $\hat{P}_{n_0+1|n_0}$.

To assess the forecast performance, I utilize the mean squared prediction error (MSPE) for $y_{1,t}$ and $y_{2,t}$, defined as

$$
\text{MSPE}(\hat{y}_1) = \frac{1}{\lfloor 0.05n \rfloor} \sum_{t=1}^{\lfloor 0.05n \rfloor} (y_{1,n_0+1+t} - \hat{y}_{1,n_0+1+t|n_0+t})^2,
$$

and

$$
\text{MSPE}(\hat{y}_2) = \frac{1}{\lfloor 0.05n \rfloor} \sum_{t=1}^{\lfloor 0.05n \rfloor} (y_{2,n_0+1+t} - \hat{y}_{2,n_0+1+t|n_0+t})^2.
$$

In this example, the sample sizes are $n = 200$, 400 and 800 and the dimensions of sample are $N = 50$, 100 and 500. For each setting, I replicate simulation 500 times and compute the median and standard deviation (in parentheses) of 500 MSPE values and the results are reported in Tables 3.1. One can see clearly from Tables 3.1 that both median and standard deviation of 500 MSPE values steadily decrease as the sample size n and dimension N increase for all setting.

Finally, I illustrate the finite sample performance for forecasting via evaluating the bootstrap prediction interval with the asymptotic bias ignored. To do this, I compute the average of empirical coverage rates (AECR) of 95% prediction interval of y_{1,n_0+1+t} and y_{2,n_0+1+t} without the asymptotic bias correction, defined as,

$$\mathrm{AECR}(\hat{y}_1) = \frac{1}{\lfloor 0.05n \rfloor B} \sum_{t=1}^{\lfloor 0.05n \rfloor} \sum_{b=1}^{B} I_b\{y_{1,n_0+1+t} \in \hat{y}_{1,n_0+1+t|n_0+t} \pm q^*_{1-\alpha} \times se(\hat{y}_1)\},$$

where $se(\hat{y}_1) = \left[\widehat{Var}^*(\hat{y}_{1,n_0+1+t|n_0+t})\right]^{1/2}$ and $q^*_{1-\alpha}$ are calculated by the wild bootstrap procedure proposed in Section 3.2.3, $I_b\{y_{1,n_0+1+t} \in \hat{y}_{1,n_0+1+t|n_0+t} \pm q^*_{1-\alpha} \times se(\hat{y}_1)\}$ is an indicator function which equals to 1 if y_{1,n_0+1+t} is covered by the interval $\hat{y}_{1,n_0+1+t|n_0+t} \pm q^*_{1-\alpha} \times se(\hat{y}_1)$ in the bth time of replication (equals to 0, otherwise), and the number of replication times B is 500. Similarly, AECR($\hat{y}_2$), $se(\hat{y}_2)$, and $I_b\{y_{2,n_0+1+t} \in \hat{y}_{2,n_0+1+t|n_0+t} \pm q^*_{1-\alpha} \times se(\hat{y}_2)\}$ can be defined in the same fashion. The simulation results are presented in Table 3.2 for all setting. From Table 3.2, one can see basically that as sample size n and dimension N become larger, AECRs of 95% prediction intervals are close to the nominal level 0.95. In general, the results of this simulated experiment demonstrate that the proposed procedure is reliable and works fairly well.

3.4 An Empirical Example of Forecasting

In this section, I examine the forecast performance of the proposed FC-FAVAR predictive model by forecasting the U.S. consumer price index (CPI) data. For the purpose of comparison, I also forecast the CPI data by using (1) the classical FAVAR with fixed factor loadings in Bernanke et al. (2005) and (2) a factor-augmented functional-coefficient (FA-FCM) forecasting model with fixed factor loadings proposed in Li et al. (2020). The data set $\boldsymbol{x}_t$ in (3.4) consists of a balanced panel of 122 monthly macroeconomic time series, which is called the FED-MD database as in McCracken and Ng (2016) and is available for download from https://research.stlouisfed.org/econ/mccracken/fred-databases/. The response variable y_t is defined as the first difference of logarithms of CPI data of U.S.. All data in this example are initially transformed to induce stationarity following the instruction in McCracken and Ng (2016) and normalized to have zero mean and unit variance. In this example, Z_t is taken to be y_{t-1} and the time span of all data is from August 1976 through September 2021, with sample size $n = 542$. Therefore, I consider FC-FAVAR predictive model as follow:

$$P_{t+1} = \gamma_0(Z_t) + \sum_{k=1}^{q} \Gamma_k(Z_t) P_{t-k+1} + \boldsymbol{\varepsilon}_{t+1},$$

and

$$x_{it} = \boldsymbol{b}_{\boldsymbol{f},i}^T(Z_t)\boldsymbol{f}_t + b_{\boldsymbol{y},i}(Z_t) y_t + u_{it}.$$

where $Z_t = y_{t-1}$, $P_t = (\boldsymbol{f}_t^T, y_t)^T$, with $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{rt})^T$. I set the number of factor $\hat{r} = 5$, which is selected through a BIC-type information criterion as in Su and Wang (2017). On the other hand, the number of lag $\hat{q}$ are determined by minimizing the same BIC criterion as in Li et al. (2020), which is 8.

Next, to measure the prediction accuracy, I computed the MSPE with the same definition as that in Section 3.3. The MPSE value for the out-sample prediction using the proposed FC-FAVAR model 0.6776. Meanwhile, the MSPE values for FA-FCM forecasting model with fixed factor loadings (5 factors and 8 lags) and the classical FAVAR with fixed factor loadings (5 factors and 8 lags) are 0.9015 and 1.0402, respectively. Finally, I obtain the 95% bootstrap prediction interval of out-sample forecasting for the last 25 observations (5% of total sample) based on the proposed FC-FAVAR model. The results are presented in Table 3.3, which shows that 24 of 25 predictive intervals contain the corresponding true values.

3.5 Conclusion

In this chapter, I investigate a new class of FC-FAVAR model, where both factor loadings of corresponding factor model and coefficients of this predictive VAR model vary with a smoothing variable. Different from the existing literature, the latent factors in the proposed model are estimated through a local PCA approach and coefficients functionals are obtained by using a local linear smoothing method. The simulation results of one-step ahead forecasting show that the performance of proposed estimation procedures is fairly well. The potential of proposed model is also demonstrated by an empirical application of forecasting CPI data of the U.S..

There are several issues still worth of further studies. First, the estimation approach of latent factors can be altered by using DP method proposed in Fan and Liao (2020) or LCCE method as in Cai et al. (2022), which may be helpful to improve the computational efficiency. Second, as is often witnessed in the empirical study, macroeconomic variables are usually need to be transformed to induce stationarity, which may be restrictive in some applications. Thus, it is of interest to further extend the FC-FAVAR model in this chapter by allowing some of regressors to be non-stationary. Third, to improve the performance of out-of-sample forecasting, it is also desirable to introduce mixed-frequency data in the proposed model, which involves new model set-up and different asymptotic theory. Aforementioned extensions are all interesting and challenging issues and I leave them as future research topics.

Table 3.1: MSPE of the one-step ahead forecast for $y_{1,t+1}$ and $y_{2,t+1}$

N	$n=200$		$n=400$		$n=800$	
	MSPE($\hat{y}_1$)	MSPE($\hat{y}_2$)	MSPE($\hat{y}_1$)	MSPE($\hat{y}_2$)	MSPE($\hat{y}_1$)	MSPE($\hat{y}_2$)
50	2.108 (1.377)	2.088 (1.406)	1.874 (0.790)	1.721 (0.679)	1.448 (0.458)	1.432 (0.374)
100	2.090 (1.066)	1.940 (1.451)	1.708 (0.718)	1.824 (0.750)	1.497 (0.387)	1.441 (0.391)
500	1.999 (2.123)	2.116 (1.505)	1.675 (0.685)	1.627 (0.705)	1.470 (0.317)	1.434 (0.418)

Table 3.2: AECRs of 95% bootstrap prediction intervals for the one-step ahead forecast for $y_{1,t+1}$ and $y_{2,t+1}$.

N	$n = 200$		$n = 400$		$n = 800$	
	AECR($\hat{y}_1$)	AECR($\hat{y}_2$)	AECR($\hat{y}_1$)	AECR($\hat{y}_2$)	AECR($\hat{y}_1$)	AECR($\hat{y}_2$)
50	0.875	0.863	0.911	0.911	0.951	0.948
100	0.869	0.880	0.924	0.920	0.943	0.948
500	0.871	0.864	0.921	0.930	0.948	0.943

Table 3.3: The post-sample bootstrap prediction intervals for the first difference of logarithms of CPI.

Observation number	Prediction interval	True value
518	(-1.507, 0.609)	-0.047
519	(-0.816, 0.689)	-0.325
520	(-1.127, 0.424)	-0.336
521	(-0.841, 0.312)	-0.405
522	(-0.893, 0.352)	-0.520
523	(-0.799, 0.375)	-1.938
524	(-3.614, 1.393)	-3.446
525	(-5.460, 2.442)	-1.099
526	(-3.336, 2.544)	0.645
527	(-1.611, 2.313)	0.702
528	(-1.515, 1.849)	0.344
529	(-2.709, 1.149)	-0.177
530	(-2.940, 1.955)	-0.719
531	(-3.475, 1.254)	-0.469
532	(-2.543, 1.081)	0.100
533	(-1.766, 0.943)	-0.151
534	(-1.292, 1.098)	0.457
535	(-1.275, 2.649)	1.087
536	(-1.515, 1.629)	1.094
537	(-2.007, 2.542)	1.284
538	(-1.857, 2.917)	1.831
539	(-2.167, 3.539)	0.508
540	(-2.637, 3.048)	0.133
541	(-1.732, 2.652)	0.374
542	(-1.699, 3.413)	1.798

Chapter 4

Solving the Price Puzzle Via A Functional Coefficient Factor-Augmented VAR Model

4.1 Introduction

In this chapter, I apply a class of FC-FAVAR models to an empirical study of macroeconomics. The detailed analysis results are reported in Section 4.3.3.

The motivation of this study arises from a debate over the issue that was found by various studies that a contractionary monetary policy is often followed by an increase of the price level, which is contrary to the standard economic theory, the so-called "price puzzle", see, e.g., Sims (1992), and Christiano, Eichenbaum and Evans (1999). Sims (1992) suggested that this puzzle results from the VAR analysis not fully capturing the information. In order to reduce the price puzzle, Sims (1992) considered adding commodity prices as an "information variable" in monetary VAR models because it contains information that helps the Federal Reserve forecast inflation, while Hanson (2004) questioned this explanation about the role for commodity prices in VAR models, finding that the ability that commodity prices have to resolve the price puzzle varies over the sample periods. Meanwhile, Bernanke et al. (2005) followed the idea in Sims (1992) and reduced the huge dimension of information set by using a FAVAR model. Other researches of solving the price puzzle include, to name a few, attributing the omission of output gap (or potential output) to the occurrence of price puzzle in Giordani (2004), referring cost channel as an alternative explanation for the price puzzle in Henzel, Hülsewig, Mayer and Wollmershäuser (2009), considering a Divisia M4 quantity of money aggregate as monetary police indicator rather than the Fed funds rate in a structural VAR model in Keating, Kelly and Valcarcel (2014), and among others. To the best of my knowledge, there is little literature to consider the relations between structural changes of economic variables and the existence of price puzzle. However, Hanson (2004) found that the price puzzle is more pronounced in specific sample periods. This observation indicates that the significance of the price puzzle may be related to the dynamic features of general economy. In addition, since the driving force for structural changes may be the institutional changes or the policy interventions, such as the changes of exchange rate systems and the U.S. quantitative easing policy, features about structural changes can apparently enrich the information set that the researchers and policy-makers care about and help correct abnormal results caused by the price puzzle. Thus, due to its ability of capturing features of structural changes, the proposed FC-FAVAR model may have the potential to reduce the price puzzle.

4.2 Generalized Impulse Response Function (GIRF)

The focus in this section is on presenting a generalized impulse responses with functional coefficients that will be used in the empirical study. As discussed in Potter (1994) and Koop, Pesaran and Potter (1996), nonlinear VAR models may produce impulse responses that are history and shock dependent (e.g., the impulse response functions are sensitive to initial conditions). To address these issues, Koop et al. (1996) constructed a generalized impulse response function (GIRF), which is defined as follows

$$\mathrm{GIRF}_P(k, \boldsymbol{\delta}_t, \boldsymbol{v}_{t-1}) = E(P_{t+k}|\boldsymbol{\delta}_t, \boldsymbol{v}_{t-1}) - E(P_{t+k}|\boldsymbol{v}_{t-1}),$$

where $k = 0, 1, \ldots, q$ are time lags after the impulse occurred, $\boldsymbol{\delta}_t$ is the $Q \times 1$ vector of shocks generating responses, and $\boldsymbol{v}_{t-1}$ is the $Q \times 1$ vector of "history" or initial values of $\mathbb{P}_t$ in (3.2).

Next, I present a GIRF generating procedure for FC-FAVAR model, which is similar to that used for parametric VAR model as in Koop et al. (1996). In particular, denote $\hat{\Omega} =$

$\frac{1}{n}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t^T$, where $\hat{\boldsymbol{\varepsilon}}_t$ is the estimated residuals. Let $\hat{\mathcal{P}}$ be the lower triangular matrix from the Cholesky decomposition such that $\hat{\Omega} = \hat{\mathcal{P}}\hat{\mathcal{P}}^T$. In addition, let $\hat{\boldsymbol{\omega}}_t$ be the corresponding structural shocks such that $\hat{\boldsymbol{\varepsilon}}_t = \hat{\mathcal{P}}\hat{\boldsymbol{\omega}}_t$. Then, the GIRF at given grid point z_0 is computed by following procedures:

1. Estimate FC-FAVAR model (3.3) at given grid point z_0 and denote the estimator as $\hat{\boldsymbol{\theta}}_0(z_0) = \text{vec}[\hat{\boldsymbol{\Gamma}}(z_0)]$. Pick a history $\boldsymbol{v}_{r,t-1}$ in a subsample of data; for example, a time interval in which an event of monetary policy occurred. The history is the actual value of the vector of lagged endogenous variables $\mathbb{P}_t$ at a particular date.

2. Pick a sequence of $Q \times 1$ vector of shocks $\{\boldsymbol{\omega}_{b,t+k}\}_{k=0}^{q}$ from $\{\hat{\boldsymbol{\omega}}_t\}_{t=1}^{n}$. All elements of $\boldsymbol{\omega}_{b,t+k}$ are drawn with replacement from the residuals $\{\hat{\boldsymbol{\omega}}_t\}_{t=1}^{n}$ as $Q \times 1$ vectors in an i.i.d. fashion.

3. Use $\hat{\boldsymbol{\theta}}_0(z_0)$, $\boldsymbol{v}_{r,t-1}$ and $\{\boldsymbol{\omega}_{b,t+k}\}_{k=0}^{q}$, simulate the evolution of P_{t+k} over $q+1$ periods. Denote the resulting baseline path $P_{b,t+k}(\boldsymbol{v}_{r,t-1}, \boldsymbol{\omega}_{b,t+k})$, for $k = 0, \ldots, q$.

4. Denote the shock to the ιth element of P_t occurs in period 0 ($k = 0$) as $\omega_{b,\iota 0}$. Substitute $\omega_{b,\iota 0}$ for the $\iota 0$th element of $\boldsymbol{\omega}_{b,t+k}$ and simulate the evolution P_{t+k} over $q+1$ periods. Denote the resulting path $P_{b,t+k}(\omega_{b,\iota 0}, \boldsymbol{v}_{r,t-1}, \boldsymbol{\omega}_{b,t+k})$, for $k = 0, \ldots, q$

5. Repeat steps 2 to 4 $B = 500$ times to obtain $\{P_{b,t+k}(\boldsymbol{v}_{r,t-1}, \boldsymbol{\omega}_{b,t+k})\}_{b=1}^{B}$ and $\{P_{b,t+k}(\omega_{b,\iota 0}, \boldsymbol{v}_{r,t-1}, \boldsymbol{\omega}_{b,t+k})\}_{b=1}^{B}$. Compute $\bar{X}_{r,t+k}(\omega_{b,\iota 0}) = \frac{1}{B}\sum_{b=1}^{B}[P_{b,t+k}(\omega_{b,\iota 0}, \boldsymbol{v}_{r,t-1}, \boldsymbol{\omega}_{b,t+k}) - P_{b,t+k}(\boldsymbol{v}_{r,t-1}, \boldsymbol{\omega}_{b,t+k})]$

6. Repeat steps 1 to 5 R times to obtain $\{\bar{X}_{r,t+k}(\omega_{b,\iota 0})\}_{r=1}^{R}$. Compute $\bar{X}_{t+k}(\omega_{b,\iota 0}) = \frac{1}{R}\sum_{r=1}^{R}\bar{X}_{r,t+k}(\omega_{b,\iota 0})$ for the average impulse response function at grid point z_0. Here, R is just the size of data of the selected subsample in step 1.

7. Denote $\bar{X}_{t+k}(\omega_{b,\iota 0}) = (\bar{\boldsymbol{f}}_{t+k}^T(\omega_{b,\iota 0}), \bar{\boldsymbol{y}}_{t+k}^T(\omega_{b,\iota 0}))^T$, where $\bar{\boldsymbol{f}}_{t+k}(\omega_{b,\iota 0})$ and $\bar{\boldsymbol{y}}_{t+k}(\omega_{b,\iota 0})$ are components of unobserved factors and observed variable, respectively, in $\bar{X}_{t+k}(\omega_{b,\iota 0})$. Then, the generalized impulse response functions of $\boldsymbol{x}_t$ at grid point z_0 is defined as

$$\mathrm{GIRF}\boldsymbol{x}_{t+k}(\omega_{b,\iota 0}) = \hat{\boldsymbol{B}}_{\boldsymbol{f}} \bar{\boldsymbol{f}}_{t+k}(\omega_{b,\iota 0}) + \hat{\boldsymbol{B}}_{\boldsymbol{y}} \bar{\boldsymbol{y}}_{t+k}(\omega_{b,\iota 0}) \tag{4.1}$$

4.3 Empirical Analysis

4.3.1 Literature Review on Price Puzzle

To clearly describe the common view of the cause of price puzzle, we first formalize the well-known reaction function of monetary policy that illustrates the relationship between the policy instrument variable and the data of economic activities. In particular, suppose that one element of $\boldsymbol{y}_t$ defined in Section 3.2.1 is the policy instrument of the monetary authority, denoted as $r_{f,t}$, then, the monetary policy reaction function is written as follows

$$r_{f,t} = \beta(\pi_{e,t} - \tilde{\pi}) + D(\boldsymbol{y}_t, \boldsymbol{f}_t) + \mu_t, \tag{4.2}$$

where $\pi_{e,t}$ is the expected future rate of inflation based on the information at time t and $\tilde{\pi}$ is the Fed's target inflation rate. In addition, μ_t is a exogenous policy shock which is an element of $\boldsymbol{\varepsilon}_t$ in (3.3) and $D(\boldsymbol{y}_t, \boldsymbol{f}_t)$ represents other observable or unobservable arguments of the reaction function (e.g., the output gap or lags of the policy instrument). Note that $r_{f,t}$ is selected to be the FFR in this chapter. In the impulse response analysis, an impulse is imposed on μ_t and then all variables in $\boldsymbol{x}_t$ can be affected by this impulse through (4.1). As pointed out in Sims (1992) and Hanson (2004), if there is a measurement error on $\pi_{e,t}$ such that

$$\pi_{e,t} = \pi_{m,t} + \pi_{\xi,t},$$

where $\pi_{m,t}$ represents the "measured" inflation expectations based only upon the information contained in the estimated model by the researcher, while $\pi_{\xi,t}$ captures information excluded from the estimation of $\pi_{e,t}$, then, (4.2) becomes the following misspecified model

$$r_{f,t} = \beta(\pi_{m,t} - \tilde{\pi}) + D(\boldsymbol{y}_t, \boldsymbol{f}_t) + \nu_t, \tag{4.3}$$

where $\nu_t = \beta\pi_{\xi,t} + \mu_t$.

Notice that the estimated policy shock ν_t is contaminated by a bias $\beta\pi_{\xi,t}$, where β is the degree to which the Fed reacts to inflationary pressures. Given the misspecified reaction function in (4.3), even the impact of a "true" policy shock μ_t upon price level is negative or zero followed by macroeconomic theory, the impulse response of price level to the estimated policy shock ν_t can be positive if $\beta\pi_{\xi,t}$ has positive impact on the price level. As a result, an empirical researcher would incorrectly infer that a contractionary policy shock had raised price level, which cause the price puzzle. For more discussion in detail, the reader is referred to the paper by Hanson (2004).

Therefore, the attempt of reducing price puzzle faces two challenges. First, the feature of $\pi_{\xi,t}$, which is associated with some omitted variables for forecasting inflation rate $\pi_{e,t}$, needs to be captured. One possible way to resolve the first challenge is to introduce more variables that could improve the forecast power of inflation rate into the VAR model. However, Bernanke et al. (2005) argued that the forecast equation of inflation rate $\pi_{e,t}$ involves the measurement of potential output and cost-push shock, which can not be directly observed by both the central bank and the econometrician. Under this circumstance, a factor-augmented VAR may demonstrate a strength of investigating models with unobserved variables. Second, the β needs to be estimated with correct specification. As documented in Hanson (2004), the magnitude of β is different substantially across regimes, which is obviously referred to as a nonlinear feature. For this reason, it is unnecessarily feasible to apply linear FAVAR model to studying the effect of monetary policy shock to macroeconomic variables. Thus, the proposed FC-FAVAR model is well-suitable to reduce price puzzle because it can not only capture nonlinearities in data, but also extract unobservable information from a huge dataset. It is worth mentioning that the aim of this empirical study is to demonstrate the usefulness of the proposed FC-FAVAR model in reducing the price puzzle compared to the classical FAVAR, rather than eradicating the existence of price puzzle. Further extensions may be realized through advocating alternative monetary instrument variables (e.g., Divisia index proposed in the seminal work of Barnett, 1980), which is out of the scope of this paper.

4.3.2 Data and Implementation

In this section, a class of FC-FAVAR model is applied to exploring the effects of innovations to monetary policy on large amounts of economic variables. To fully demonstrate the usefulness of the proposed FC-FAVAR models, I revisit one of issues that was discussed in Bernanke et al. (2005) and compare classical FAVAR models with FC-FAVAR models in the performance of reducing the price puzzle. For the purpose of making comparison, the factor model in this example is set as model (1.5), which is the same as that in Bernanke et al. (2005) and is nested by the factor model with dynamic loadings discussed in chapter 3. Thus, the FC-FAVAR model in this example is written as

$$\hat{P}_t = \gamma_0(Z_t) + \sum_{k=1}^{8} \Gamma_k(Z_t)\hat{P}_{t-k} + \boldsymbol{\varepsilon}_t, \tag{4.4}$$

where $\hat{P}_t = (\hat{\boldsymbol{f}}_t^T, y_t)^T$ and $\hat{\boldsymbol{f}}_t$ is estimated in the same way as in Bernanke et al. (2005). In particular, denote y_t as a series of Federal funds rate (FFR), while $\hat{C}(\boldsymbol{x}_t)$ is the vector of principal components estimated from the entire dataset $\boldsymbol{x}_t$. Since both y_t and $\hat{C}(\boldsymbol{x}_t)$ involve the series of FFR, it would be invalid to identify the effect of policy shocks when simply estimating a VAR in y_t and $\hat{C}(\boldsymbol{x}_t)$. Thus, the direct dependence of $\hat{C}(\boldsymbol{x}_t)$ on y_t should be removed. By following the procedures in Bernanke et al. (2005), I first regress $\hat{C}(\boldsymbol{x}_t)$ on y_t in the form of $\hat{C}(\boldsymbol{x}_t) = b_C\hat{C}(\hat{\boldsymbol{f}}_t) + b_{\text{FFR}}y_t + e_t$, where $\hat{C}(\hat{\boldsymbol{f}}_t)$ is an estimate of all the common components other than y_t. One way to obtain $\hat{C}(\hat{\boldsymbol{f}}_t)$ is to extract principal components from the subset of "slow-moving variables", which are not affected contemporaneously by

y_t. The reader is referred to Bernanke et al. (2005) for more discussion about "slow-moving variables". Next, I construct $\hat{\boldsymbol{f}}_t$ as $\hat{C}(\boldsymbol{x}_t) - \hat{b}_{\mathrm{FFR}} y_t$ and finally estimate the FC-FAVAR model (3.3) in y_t and $\hat{\boldsymbol{f}}_t$, with y_t ordered last. The number of factors r and lags q are selected by using a nonparametric BIC-type criterion proposed in Li et al. (2020) and the selection results are $\hat{r} = 6$ and $\hat{q} = 8$. In addition, $\Gamma_k(Z_t) = (\gamma_{k\iota\ell,P}(Z_t))_{7\times 7}$ has the same definition as in (3.2) and (3.3) and y_t is ordered last in each case.

In this example, my dataset $\boldsymbol{x}_t$ consists of a balanced panel of 100 monthly macroeconomic time series, which are updates of series used in Bernanke et al. (2005) and McCracken and Ng (2016). The data span the period from July 1962 through September 2021 and sample sizes are 711. These series are initially transformed to induce stationarity and were normalized to have zero mean and unit variance. The description of the series in the dataset and their transformation are described in Appendix C. In addition, the monetary policy indicator is chosen as FFR, which is denoted as y_t. Furthermore, $Z_t = \Pi_t/10$, where Π_t is the fifth-lagged spread between Moody BAA-rated corporate bond and FFR, denoted by SFYBAAC in what follows and in Appendix. Finally, by using the procedure presented in Section 4.2, I obtain the generalized impulse responses of all variables of $\boldsymbol{x}_t$, which is $\mathrm{GIRF}\boldsymbol{x}_{t+k}(\omega_{b,\iota 0})$ in (4.1).

This choice of smoothing variable is reasonable, because the SFYBAAC contains information of both the spread between Moody AAA-rated corporate bond and FFR, denoted by SFYAAAC in what follows and in Appendix C, and the spread between Moody BAA-rated corporate bond and Moody AAA-rated corporate, denoted by DEFAULT in what follows. Indeed, Bernanke (1990) suggested that DEFAULT should be used as a measure of the behavior of perceived default risk in the economy, which has an influence on the Federal Reserve Bank for making monetary policy decision. Furthermore, SFYAAAC can serve as a nice indicator of monetary policy changes. Thus, the temporal changes of SFYBAAC may indicate the shift of environment of decision making for monetary policy. For this regard, the reader is referred to the paper by Bernanke (1990) for more discussions. It is worth emphasizing that SFYBAAC is not the only choice for smoothing variable, of course, other variables of economic status may also be suitable to serve as the smoothing variable and this may be left in a future study.

Remark 4.3.1. *(Identification restrictions on factors and policy shocks). It is well known that the model in* (1.5) *and (3.2)-(3.3) can only be estimated after imposing identification restrictions on factors and policy shocks. To this end, let* $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$, $\boldsymbol{f} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)^T$ *and* $\hat{\boldsymbol{f}} = (\hat{\boldsymbol{f}}_1, \ldots, \hat{\boldsymbol{f}}_n)^T$, *I use the standard normalization implicit in the principal components in the same way as in Bernanke et al. (2005). That is, I restrict* $\hat{\boldsymbol{f}}^T \hat{\boldsymbol{f}}/n = I_r$. *Moreover, define the rotation matrix for factors as*

$$H = V^{-1}(\hat{\boldsymbol{f}}^T \boldsymbol{f}/n)(\boldsymbol{B}_{\boldsymbol{f}}^T \boldsymbol{B}_{\boldsymbol{f}}/N),$$

where V *is an* $r \times r$ *diagonal matrix with main diagonal elements as the* r *largest eigenvalues of* $\boldsymbol{x}\boldsymbol{x}^T/(nN)$, *in descending order. As discussed in Bernanke et al. (2005) and Yamamoto (2019), by assuming a recursive structure where all the factors in (3.3) respond with a lag to change in the monetary policy instrument, ordered last in* $\boldsymbol{y}_t$, *then, no further restrictions are required on factors in (3.3), and the identification of the policy shock can be achieved in (3.3) as if it were be a standard VAR.*

4.3.3 Empirical Results

The analysis in this section aims at comparing the results generated from a standard FAVAR model in Bernanke et al. (2005) to that from our proposed FC-FAVAR model in (3.3) based on the dataset $\boldsymbol{x}_t$ from a new time span. It is crucial to first show that the coefficients in (3.3) change over SFYBAAC in the empirical example. To this end, I only present estimation results of functional coefficients related to y_t for space saving, which are all components of the 7th column of $\Gamma_1(z_l)$ defined in (4.4) and $\{z_l = 0.05(l-1) - 0.3 : 1 \leq l \leq 13\}$ are grid points chosen from the interval $[-0.3, 0.3]$, which is the range of Z_t. The results are reported in Figure 4.1, in which the vertical axis of each panel measures the functional coefficient and the horizontal axis of each panel measures SFYBAAC. It is obvious that all functional coefficients in Figure 4.1 are not constant, but vary greatly over the interval $[-0.3, 0.3]$. This observation indicates that the dataset possesses features of nonlinearity. Therefore, adopting classical FAVAR model in this dataset may result in severe problem of misspecification and subsequently, the price puzzle.

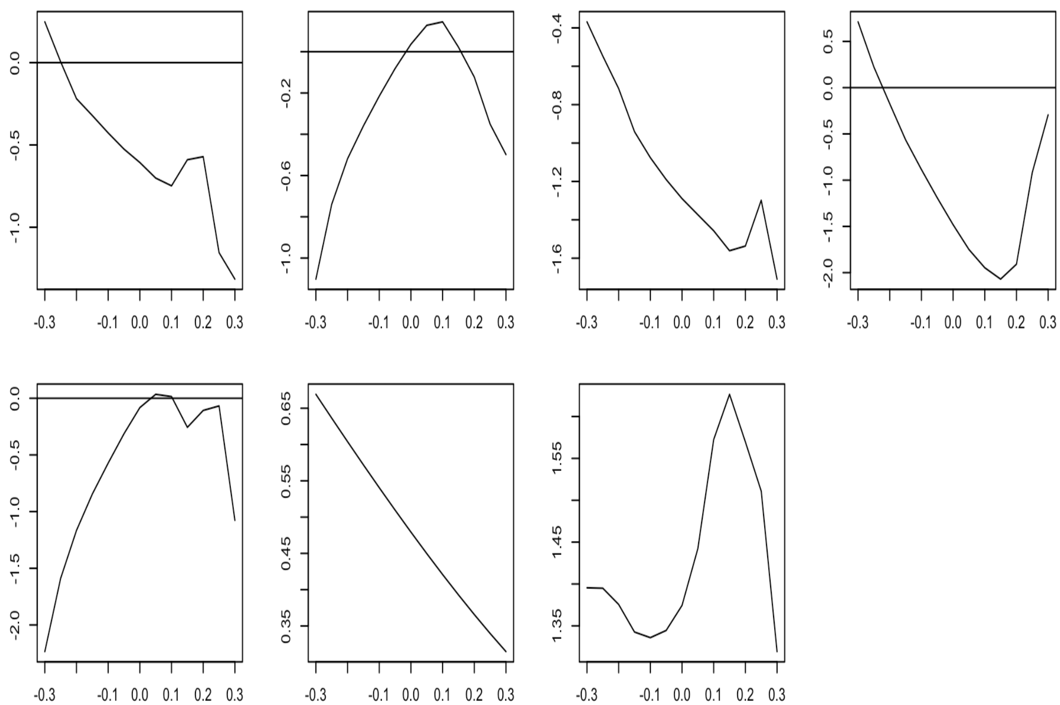

Figure 4.1: Entries of 7th column of functional coefficients matrix $\Gamma_1(\cdot)$ with respect to the changes of SFYBAAC estimated by the FC-FAVAR model with 6 factors and 8 lags.

Based on the analysis in Section 4.3.1, the reduction of price puzzle follows from the decrease of the time that the response of price level to policy shocks spends to become negative. Thus, the faster the curve of impulse response of CPI to monetary policy shocks becomes negative, the better the model performs in reducing price puzzle. In the next group of figures, I compare the results of generalized impulse response functions estimated by our FC-FAVAR model and by classical FAVAR in Bernanke et al. (2005). I choose five grid points as the data of SFYBAAC on five time points: 1971:01, 1980:03, 2006:08, 2011:08 and 2020:07, and then obtain generalized impulse responses at these grid points.

In October 1979, Fed Chairman Paul Volcker announced a shift of the instrument of policy from the federal funds rate to non-borrowed reserves. A number of researchers observed that this change of policy instrument significantly enlarged the β in monetary policy reaction function in (4.2), causing a regime-specific phenomenon on structural parameters; see, for instance, Taylor (1999), Clarida, Galì and Gertler (2000) and Hanson (2004). Therefore, the first and second time points 1971:01 and 1980:03 can act as nice proxies of the“pre-Volcker" and “post-Volcker" periods, respectively, to demonstrate my model's ability of capturing structural changes in data. In addition, the third, fourth and fifth time points 2006:08, 2011:08 and 2020:07 represent periods of “pre-financial crisis", “post-financial crisis" and COVID-19 pandemic, respectively. It is well-known that the FFR has been stuck at or near the zero lower bound (ZLB) since 2008 and during some periods of COVID-19 pandemic, which poses a criticism about the effectiveness of FFR as monetary policy indicator. Thus, the studies based on the fourth and fifth time points are suitable for checking the reliability of FC-FAVAR model under extreme conditions of economy.

In Figure 4.2, the vertical axis measures GIRF generated by procedures as in Koop et al. (1996) for classical FAVAR model, while in Figures 4.3 and 4.4, the vertical axis represents GIRF generated by procedures in Section 3.2.3 at given grid point z_0, and the horizontal axis in Figures 4.2-4.4 represent the time lag k. In addition, I standardize the monetary shock to correspond to a 25-basis-point innovation in the FFR. Figure 4.2 presents the resulting impulse response functions of FFR, industrial production and consumer price index of all items for the classical FAVAR proposed in Bernanke et al. (2005) for two sample periods: 1962:07-2001:08 (the top panel) and 1962:07-2021:09 (the bottom panel), respectively. The first period ends in August 2001 following Bernanke et al. (2005), and the second period extends the sample to September 2021. In both two sample periods, I employ 8 lags and the number of factors is 6. Observed that in the first time span, the response of all variables move in the same way as in Bernanke et al. (2005), while in the second time span, the response of CPI goes up to positive and fail to return to negative within 50 lags, which indicates that there is still a strong price puzzle in the classical FAVAR specification. These results are not surprising, because as the sample periods enlarged, the information about changes of general economy become significant and may eventually undermine the estimation results given by the linear FAVAR model.

In contrast, Figure 4.3 displays GIRFs on five time points obtained by the proposed FC-FAVAR model for the sample period ends in September 2021 and I use 8 lags for $\hat{q}$, and 6 factors for $\hat{r}$. It is interesting that after considering the changes of economic environment, the responses of CPI go down to negative within 40 lags at all five grid points, suggesting that the price puzzle is considerably reduced compared to the results of classical FAVAR. More specifically, the responses of CPI at time points 1971:01, 1980:03 and 2006:08 drop to negative within 30 lags, which show that FC-FAVAR model can nicely reduce price puzzle by correcting the measurement error discussed in Section 4.3.1. For the results on time points 2011:08 and 2020:07, even when FFR reaches to ZLB, the estimated responses of CPI still returns to negative within 40 lags. In this case, although there exists macroeconomic models with alternative policy instruments that work fairly well in correcting the abnormal of price level, the proposed FC-FAVAR model can reasonably reduce price puzzle without introducing new structures in the conventional macroeconomic model and replacing policy instruments. Of course, it is of great interest to use other variables as policy instruments instead of the FFR in FC-FAVAR model and we leave this as a future topic.

Finally, Figure 4.4 shows the GIRF of selected macroeconomic variables to monetary policy shocks generated by the procedure presented in Section 4.2 on time point 1971:01. The responses are generally of the expected sign and magnitude: following a contractionary monetary policy shock, prices go down to negative rapidly, money aggregates decline, and the dollar appreciates. The dividend yields initially jump above the steady state and finally go down. To sum up, these results seem to demonstrate that measures of the effects of monetary policy are consistent and sensible. Notice that I only display 12 responses of all 100 that could also be investigated technically. The results for the rest responses are available upon request.

4.4 Conclusion

In this paper, I investigate a functional coefficient FAVAR model with an application to resolving the price puzzle and coefficients functionals are estimated by using a two-stage kernel smoothing method. In addition, there is little literatures regarding the relationship between the existence of price puzzle and the structural changes in the economic environment. After considering the changes of specific state of economy, the proposed framework mitigates the issue of price puzzle and still allows to estimate responses of large amounts of economic variables to monetary policy shocks.

There are several issues still worth of further studies. First, it is interesting to introduce heteroscedasticity into model (3.3), although a functional coefficient model has an ability to capture partial heteroscedasticity as argued by Cai (2010), so that the dynamics of monetary policy shocks can also be captured. Second, the asymptotic properties of functional coefficients and impulse response functions need to be derived and this should not be hard given the similar precedents of theoretical work in Cai et al. (2000), Cai et al. (2006) and Li et al. (2020). Third, it is also desirable to extend model (3.3) by allowing some of regressors to be non-stationary, which is a commonly-faced scenario in the empirical studies of macroeconomics. I leave these important issues as future research topics.

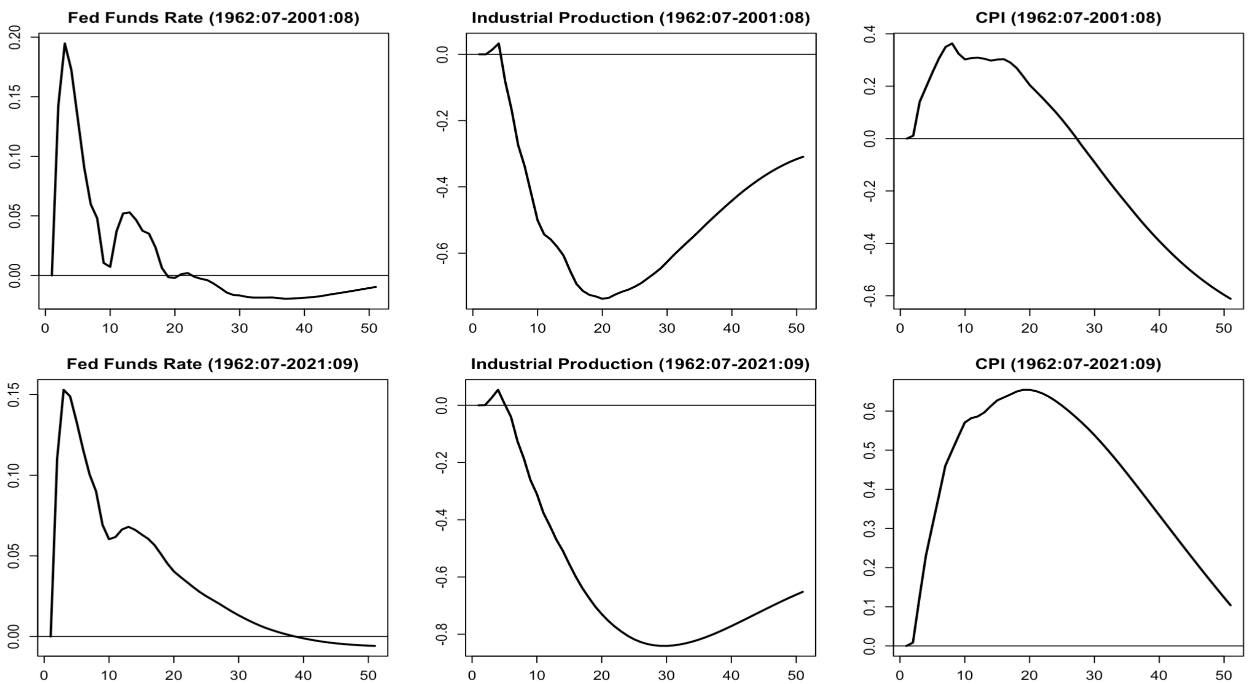

Figure 4.2: Generalized impulse response functions of FFR, industrial production and consumer price index of all items for the classical FAVAR model with 6 factors and 8 lags for two sample periods: 1962:07-2001:08 (the top panel) and 1962:07-2021:09 (the bottom panel).

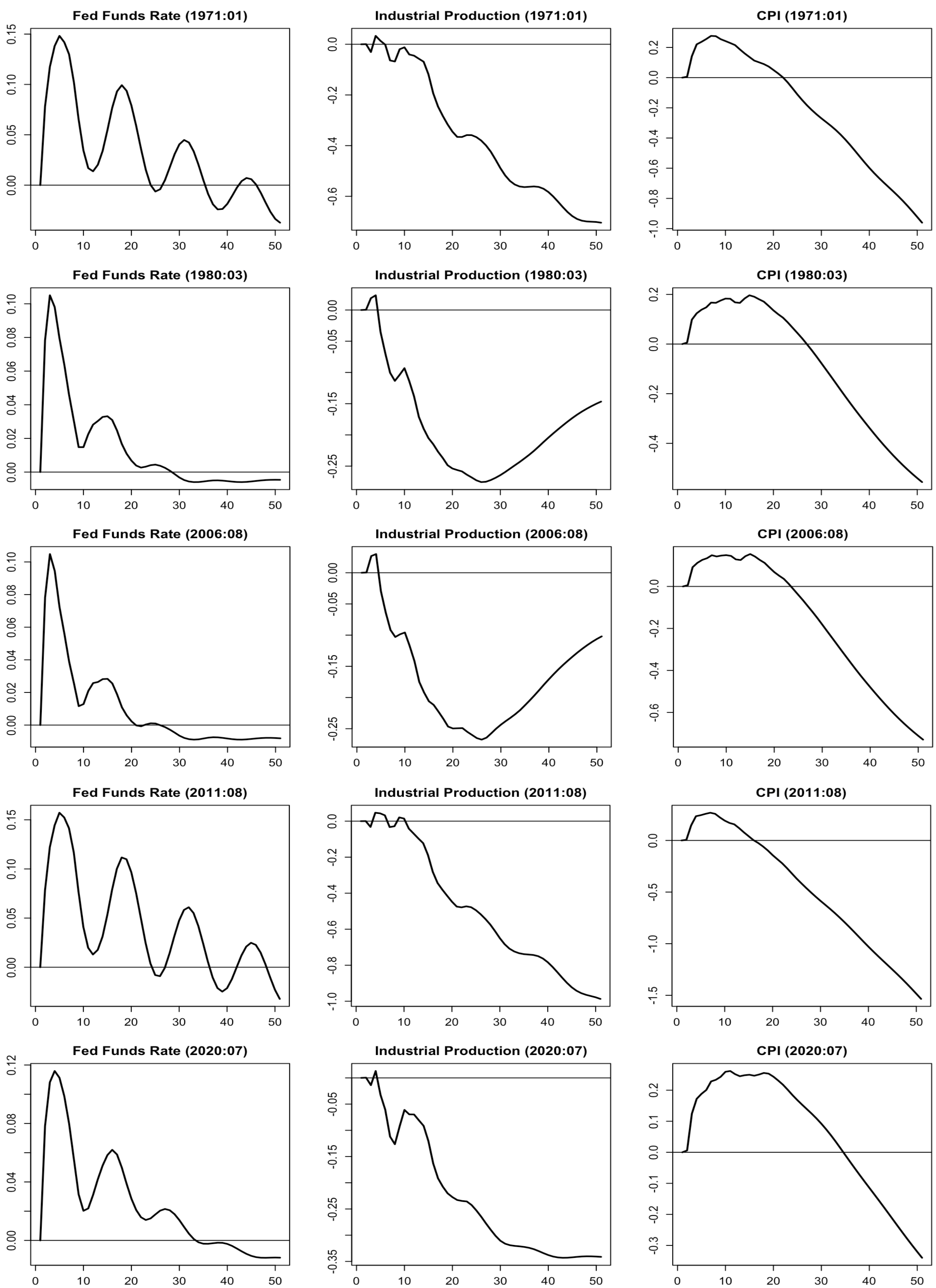

Figure 4.3: Generalized impulse response functions of FFR, industrial production and consumer price index of all items for the FC-FAVAR model with 6 factors and 8 lags on 1971:01 (the first row of panel), 1980:03 (the second row of panel), 2006:08 (the third row of panel), 2011:08 (the fourth row of panel) and 2020:07 (the fifth row of panel) .

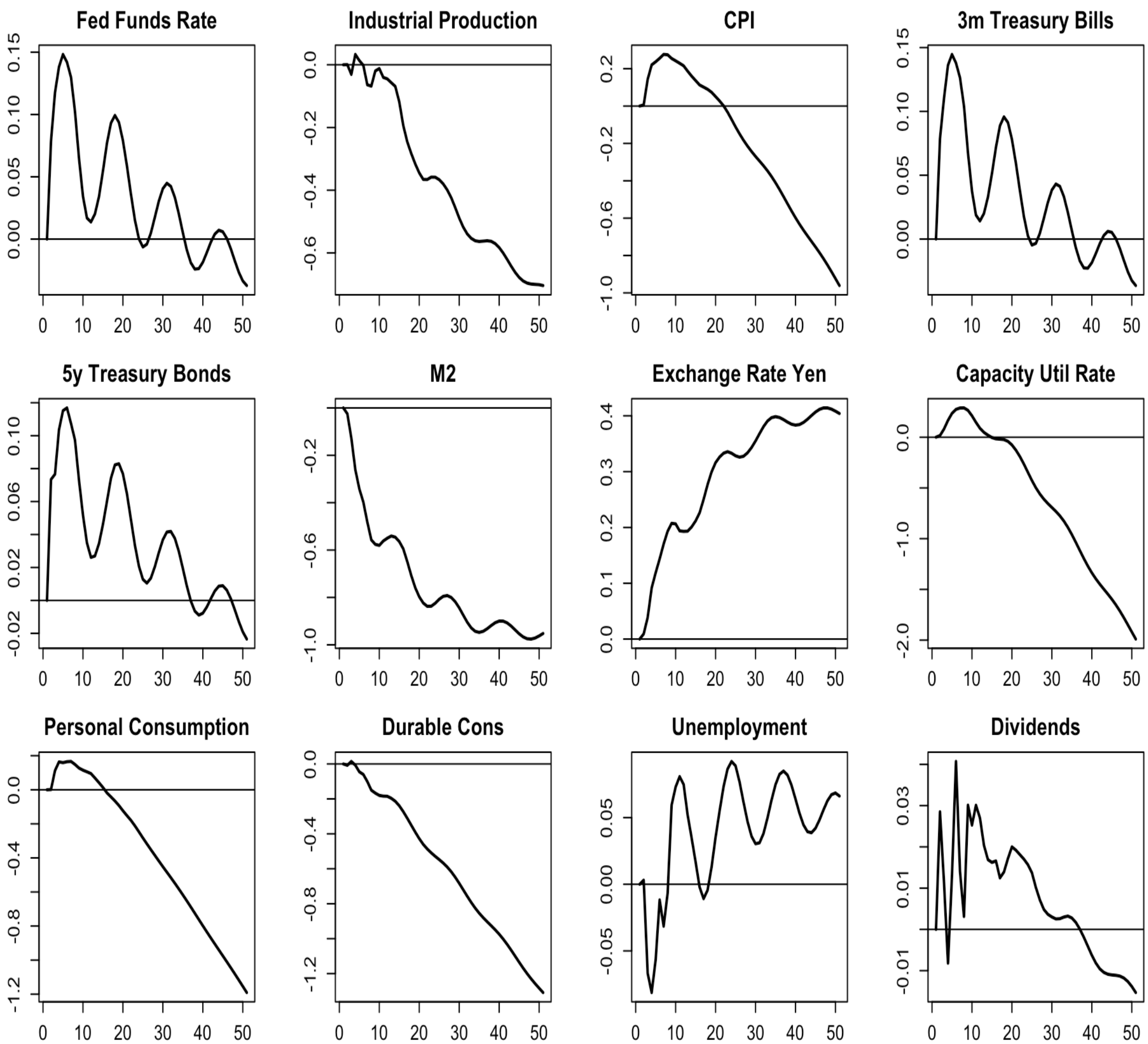

Figure 4.4: Generalized impulse responses of 12 variables generated from FC-FAVAR with 6 factors and 8 lags on time point 1971:01.

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.

Bai, J., Li, K. and Lu, L. (2016). Estimation and Inference of FAVAR Models. *Journal of Business & Economic Statistics*, 34(4), 620-641.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1), 191-221.

Bai, J. and Ng, S. (2006). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica*, 74(4), 1133-1150.

Barnett, W. (1980). Economic Monetary Aggregates: an Application of Aggregation and Index Number Theory. *Journal of Econometrics*, 14, 11-48.

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

Bernanke, B. (1990). On the Predictive Power of Interest Rates and Interest Rate Spreads. *NBER Working paper*.

Bernanke, B. and Boivin, J. (2003). Monetary Policy in a Data-rich Environment. *Journal of Monetary Economics*, 50(3), 525-546.

Bernanke, B., Boivin, J. and Eliasz, P. (2005). Measuring the Effects of Monetary Policy: a Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics*, 120(1), 387-422.

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

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. (2007). Trending Time-varying Coefficient Time Series Models with Serially Correlated Errors. *Journal of Econometrics*, 136(2), 163-188.

Cai, Z. (2010). Functional coefficient models for economic and financial data. In *Oxford Handbook of Functional Data Analysis (Eds: F. Ferraty and Y. Romain)*. Oxford University Press, Oxford, UK, pp.166-186.

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 Liu, X. (2022). A Nonparametric Dynamic Network via Multivariate Quantile Autoregressions. *Working Paper*, Department of Economics, University of Kansas.

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 Tiwari, R.C. (2000). Application of a local linear autoregressive model to BOD time series. *Environmetrics*, 11(3), 341-350.

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.

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.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, Factor Structure and Mean-variance Analysis in Large Asset Markets. *Econometrica*, 51(5), 1305-1324.

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

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

Christiano, L.J., Eichenbaum, M. and Evans, C.L. (1999). Monetary Policy Shocks: What Have We Learned and To What End? In: *Handbook of Macroeconomics*. Volume 1, part A, pp. 65-148.

Clarida, R., Galì, J. and Gertler, M. (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics*, 115 (1), 147-180.

Corradi, V. and Swanson, N.R. (2014). Testing for Structural Stability of Factor Augmented Forecasting Models. *Journal of Econometrics*, 182(1), 100-118.

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

Dufour, J.M. and Stevanović, D. (2013). Factor-Augmented VARMA Models with Macroeconomic Applications. *Journal of Business & Economic Statistics*, 31(4), 491-506.

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

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.

Fama, E. and French, K. (1992). The Cross-section of Expected Stock Returns. *Journal of Finance*, 47(2), 427-465.

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

Fan, J. and Liao, Y. (2020). Learning Latent Factors From Diversified Projections and Its Applications to Over-Estimated and Weak Factors. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2020.1831927.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large Covariance Estimation by Thresholding Principal Orthogonal Complements (With Discussions). *Journal of the Royal Statistical Society: Series B*, 75(4), 603-680.

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.

Giordani, P. (2004). An Alternative Explanation of the Price Puzzle. *Journal of Monetary Economics*, 51(6), 1271-1296.

Giraitis, L., Kapetanios, G. and Yates, T. (2014). Inference on Stochastic Time-varying Coefficient Models. *Journal of Econometrics*, 179(1), 46-65.

Giraitis, L., Kapetanios, G. and Yates, T. (2018). Inference on Multivariate Heteroskedastic Time-varying Random Coefficient Models. *Journal of Time Series Analysis*, 39(2), 129-149.

Gonçalves, S. and Perron, B. (2014). Bootstrapping Factor-augmented Regression Models. *Journal of Econometrics*, 182, 156-173.

Gourieroux, C. and Jasiak, J. (2008). Dynamic Quantile Models. *Journal of Econometrics*, 147(1), 198-205.

Guo, S., Wang, Y. and Yao, Q. (2016). High-Dimensional and Banded Vector Autoregressions. *Biometrika*, 103(4), 889-903.

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

Hansen, B.E. (2001). The New Econometrics of Structural Change: Dating Breaks in US Labour Productivity. *Journal of Economic Perspectives*, 15(4), 117-128.

Hanson, M.S. (2004). The "Price Puzzle" Reconsidered. *Journal of Monetary Economics*, 51(7), 1385-1413.

Härdle, W.K., Tsybakov, A. and Yang, L. (1998). Nonparametric Vector Autoregression. *Journal of Statistical Planning and Inference*, 68(2), 221-245.

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.

Henzel, S., Hülsewig, O., Mayer, E. and Wollmershäuser, T. (2009). The Price Puzzle Revisited: Can the Cost Channel Explain a Rise in Inflation After a Monetary Policy Shock? *Journal of Macroeconomics*, 31(2), 268-289.

Honda, T. (2000). Nonparametric Estimation of a Conditional Quantile for α-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.

Hsu, N.J., Hung, H.L. and Chang, Y.M. (2008). Subset Selection for Vector Autoregressive Processes Using Lasso. *Computational Statistics and Data Analysis*, 52(7), 3645-3657.

Keating, J.W., Kelly, L.J. and Valcarcel, V.J. (2014). Solving the Price Puzzle with an Alternative Indicator of Monetary Policy. *Economics Letters*, 124(2), 188-194.

Kilian, L. (1998). Small Sample Confidence Intervals for Impulse Response Functions. *Review of Economics and Statistics*, 80(2), 218-230.

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.

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

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.

Li, D., Tosasukul, J. and Zhang, W. (2020). Nonlinear Factor-Augmented Predictive Regression Models with Functional Coefficients. *Journal of Time Series Analysis*, 41(3), 367-386.

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.

McCracken, M.W. and Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, 34(4), 574-589.

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

Pelger, M. and Xiong, R. (2021). State-Varying Factor Models of Large Dimensions. *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2021.1927744.

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

Primiceri, G.E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies*, 72(3), 821-852.

Qu, Z. and Yoon, J. (2015). Nonparametric Estimation and Inference on Conditional Quantile Processes. *Journal of Econometrics*, 185(1), 1-19.

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

Sewell, D.K. and Chen, Y. (2015). Latent Space Models for Dynamic Networks. *Journal of the American Statistical Association*, 110(512), 1646-1657.

Sheather, S.J. and Jones, M.C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B*, (53)3, 683-690.

Sims, C.A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.

Sims, C.A. (1992). Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy. *European Economic Review*, 36(1), 975-1011.

Sims, C.A. and Zha, T. (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, 96(1), 54-81.

Stock, J.H. and Watson, M.W. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business & Economic Statistics*, 14(1), 11-30.

Stock, J.H. and Watson, M.W. (2002). Forecasting using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460), 1167-1179.

Stock, J.H. and Watson, M.W. (2009). Forecasting in Dynamic Factor Models subject to Structural Instability. In: Castle, J., Shephard, N. (Eds.), The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry. Oxford University Press, pp. 173–205.

Stock, J.H. and Watson, M.W. (2016). Dynamic Factor Models, Factor-augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In: *Handbook of Macroeconomics*. Volume 2, pp. 415-525.

Su, L. and Wang, X. (2017). On Time-Varying Factor Models: Estimation and Testing. *Journal of Econometrics*, 198(1), 84-101.

Taylor, J.B. (1999). A Historical Analysis of Monetary Policy Rules. In: Taylor, J.B. (Ed.), Monetary Policy Rules. University of Chicago Press, Chicago, pp. 319–341.

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.

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.

Yamamoto, Y. (2019). Bootstrap Inference for Impulse Response Functions in Factor-augmented Vector Autoregressions. *Journal of Applied Econometrics*, 34, 247-267.

Yan, Y. and Cheng, T. (2022). Factor-augmented Forecasting Regressions with Threshold Effects. *The Econometrics Journal*, 25(1), 134-154.

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

Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H. (2017). Network Vector Autoregression. *Annals of Statistics*, 45(3), 1096-1123.

Zhu, X., Wang, W., Wang, H. and Härdle, W.K. (2019). Network Quantile Autoregression. *Journal of Econometrics*, 212(1), 345-358.

Appendix: Mathematical Proofs of Theorem 2.2.1 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). □

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, τ is dropped from $\boldsymbol{\alpha}_\tau(\cdot)$ and write h_1 as h for simplicity. Let $a_n = (nh_1)^{-1/2}$, for $1 \leq \mathfrak{s} \neq t \leq n$ and for any fixed $Z_t \neq Z_s$, define $\boldsymbol{\vartheta}_0 = a_n^{-1}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}(Z_t))$ and $\hat{\boldsymbol{\vartheta}}_0 = a_n^{-1}(\hat{\boldsymbol{\delta}}_0 - \boldsymbol{\alpha}(Z_t))$. Of course, $\boldsymbol{\vartheta} = a_n^{-1} \boldsymbol{H}_1 \begin{pmatrix} \boldsymbol{\delta}_0 - \boldsymbol{\alpha}(Z_t) \\ \boldsymbol{\delta}_1 - \boldsymbol{\alpha}^{(1)}(Z_t) \end{pmatrix}$, $\hat{\boldsymbol{\vartheta}} = a_n^{-1} \boldsymbol{H}_1 \begin{pmatrix} \hat{\boldsymbol{\delta}}_0 - \boldsymbol{\alpha}(Z_t) \\ \hat{\boldsymbol{\delta}}_1 - \boldsymbol{\alpha}^{(1)}(Z_t) \end{pmatrix}$, where $\boldsymbol{H}_1 = \mathrm{diag}\{I_{\kappa m+1}, h_1 I_{\kappa m+1}\}$. In addition, let $\boldsymbol{W}_{\mathfrak{s}}^* = \begin{pmatrix} \boldsymbol{W}_{\mathfrak{s}} \\ z_{\mathfrak{s}h} \boldsymbol{W}_{\mathfrak{s}} \end{pmatrix}$, where $z_{\mathfrak{s}h} = (Z_{\mathfrak{s}} - Z_t)/h$. Also, define $Y_{\mathfrak{s}}^* = Y_{\mathfrak{s}} -$

$\boldsymbol{W}_{\mathfrak{s}}^T[\boldsymbol{\alpha}(Z_t) + \boldsymbol{\alpha}^{(1)}(Z_t)(Z_{\mathfrak{s}} - Z_t)]$. Therefore,

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \sum_{\mathfrak{s}=m+1\neq t}^{n} \rho_\tau(Y_{\mathfrak{s}}^* - a_n \boldsymbol{\vartheta}^T \boldsymbol{W}_{\mathfrak{s}}^*) K(z_{\mathfrak{s}h}) \equiv \arg\min_{\boldsymbol{\vartheta}} G(\boldsymbol{\vartheta}).$$

The derivative of $G(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$ (except at point $Y_{\mathfrak{s}}^* = a_n \boldsymbol{\vartheta}^T \boldsymbol{W}_{\mathfrak{s}}^*$) is given by

$$T_n(\boldsymbol{\vartheta}) = a_n \sum_{\mathfrak{s}=m+1\neq t}^{n} \psi_\tau(Y_{\mathfrak{s}}^* - a_n \boldsymbol{\vartheta}^T \boldsymbol{W}_{\mathfrak{s}}^*) \boldsymbol{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h}), \tag{A.1}$$

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

$$T_n(\boldsymbol{\zeta}) = a_n \sum_{\mathfrak{s}=m+1\neq t}^{n} \psi_\tau(Y_{\mathfrak{s}}^* - \boldsymbol{\zeta}^T \boldsymbol{W}_{\mathfrak{s}}^*) \boldsymbol{W}_{\mathfrak{s}}^* K(z_{\mathfrak{s}h}). \tag{A.2}$$

In particular, suppose that $\mathscr{D}$ is any compact subset of $\mathbb{R}$. To show the uniform consistency of $\hat{\boldsymbol{\alpha}}(\cdot)$ in Lemma A.2 later, for any $z \in \mathscr{D}$, define $\hat{\boldsymbol{\vartheta}}(z) = a_n^{-1} \boldsymbol{H}_1 \begin{pmatrix} \hat{\boldsymbol{\alpha}}(z) - \boldsymbol{\alpha}(z) \\ \hat{\boldsymbol{\alpha}}^{(1)}(z) - \boldsymbol{\alpha}^{(1)}(z) \end{pmatrix}$ and $\hat{\boldsymbol{\zeta}}(z) = a_n \hat{\boldsymbol{\vartheta}}(z)$. Let $\boldsymbol{W}_{\mathfrak{s}}(z) = \begin{pmatrix} \boldsymbol{W}_{\mathfrak{s}} \\ ((Z_{\mathfrak{s}} - z)/h) \boldsymbol{W}_{\mathfrak{s}} \end{pmatrix}$ and $Y_{\mathfrak{s}}(z) \equiv Y_{\mathfrak{s}} - \boldsymbol{W}_{\mathfrak{s}}^T[\boldsymbol{\alpha}(z) + \boldsymbol{\alpha}^{(1)}(z)(Z_{\mathfrak{s}} - z)]$.

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

Proof. Let $\boldsymbol{v} \in \mathbb{R}^{2(\kappa m+1)}$ be an arbitrary $2(\kappa m + 1)$-dimension vector that satisfy $\|\boldsymbol{v}\| = 1$, 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

$$P\left\{ \inf_{\|\boldsymbol{v}\|=1} \sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T \psi_\tau(Y_{\mathfrak{s}}(z) - (C(m/nh)^{1/2}\boldsymbol{v})^T \boldsymbol{W}_{\mathfrak{s}}(z)) \boldsymbol{W}_{\mathfrak{s}}(z) K((Z_{\mathfrak{s}} - z)/h) > 0 \right\} > 1 - \varepsilon \tag{A.3}$$

uniformly over $z \in \mathscr{D}$, then, the proof is finished. We first show that (A.3) holds for any fixed $z_0 \in \mathscr{D}$. To this end, define $z_{\mathfrak{s}0h} = (Z_{\mathfrak{s}} - z_0)/h$ and let $\xi_{\mathfrak{s}}(v) = \psi_\tau(Y_{\mathfrak{s}}(z_0) - v^T \boldsymbol{W}_{\mathfrak{s}}(z_0)) \boldsymbol{W}_{\mathfrak{s}}(z_0) K(z_{\mathfrak{s}0h}) - \psi_\tau(Y_{\mathfrak{s}}(z_0)) \boldsymbol{W}_{\mathfrak{s}}(z_0) K(z_{\mathfrak{s}0h})$, then,

$$
\begin{aligned}
&\sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T \psi_\tau(Y_{\mathfrak{s}}(z_0) - (C(m/nh)^{1/2}\boldsymbol{v})^T \boldsymbol{W}_{\mathfrak{s}}(z_0))\boldsymbol{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h}) \\
= &\sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T \psi_\tau(Y_{\mathfrak{s}}(z_0))\boldsymbol{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h}) + \sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T E\{\xi_{\mathfrak{s}}(C(m/nh)^{1/2}\boldsymbol{v})\} \\
&+ \sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T [\xi_{\mathfrak{s}}(C(m/nh)^{1/2}\boldsymbol{v}) - E\{\xi_{\mathfrak{s}}(C(m/nh)^{1/2}\boldsymbol{v})\}] = M_1 + M_2 + M_3.
\end{aligned}
$$

Following the proof in Xiao and Koenker (2009), we first analyze M_3. Covering the ball $\{\|v\| \le C(m/nh)^{1/2}\}$ with cubes $\mathcal{C} = \{\mathcal{C}_k\}$, where $\mathcal{C}_k$ is a cube with center v_k and side length $C(m/(nh)^5)^{1/2}$, so that $N(n) = \#(\mathcal{C}) = (2(nh)^2)^m$, and for $v \in \mathcal{C}_k$, $\|v - v_k\| \le C(m/(nh)^{5/2})$. Since $I(Y_{\mathfrak{s}}(z_0) < x)$ is nondecreasing in x, then,

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

where $b_{n\mathfrak{s}}(v_k) = I(Y_{\mathfrak{s}}(z_0) < v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0)) - I(Y_{\mathfrak{s}}(z_0) < v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0) + C(m/(nh)^{5/2})\|\boldsymbol{W}_{\mathfrak{s}}(z_0)\|)$ and $d_{n\mathfrak{s}}(v_k) = I(Y_{\mathfrak{s}}(z_0) < v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0) + C(m/(nh)^{5/2})\|\boldsymbol{W}_{\mathfrak{s}}(z_0)\|) - I(Y_{\mathfrak{s}}(z_0) < v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0) - C(m/(nh)^{5/2})\|\boldsymbol{W}_{\mathfrak{s}}(z_0)\|)$. 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 $\flat > 0$, $|\psi_\tau(Y_{\mathfrak{s}}(z_0)) - \psi_\tau(Y_{\mathfrak{s}}(z_0) - v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0))|^\flat = I(d_{3\mathfrak{s}} < Y_{\mathfrak{s}} \le d_{4\mathfrak{s}})$, 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 $c_{2\mathfrak{s}} = [\boldsymbol{\alpha}(z_0) + \boldsymbol{\alpha}^{(1)}(z_0)(Z_{\mathfrak{s}} - z_0)]^T \boldsymbol{W}_{\mathfrak{s}}$ and $c_{3\mathfrak{s}} = v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0)$. Therefore, by Assumption A4, there exists a $C > 0$ such that $E\{|\psi_\tau(Y_{\mathfrak{s}}(z_0)) - \psi_\tau(Y_{\mathfrak{s}}(z_0) - v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0))|^\flat | Z_{\mathfrak{s}}, \boldsymbol{W}_{\mathfrak{s}}\} = F_{Y|Z,\boldsymbol{W}}(d_{4\mathfrak{s}}) - F_{Y|Z,\boldsymbol{W}}(d_{3\mathfrak{s}}) \le C|v_k^T \boldsymbol{W}_{\mathfrak{s}}(z_0)| \le C(m/nh)^{1/2}\|\boldsymbol{W}_{\mathfrak{s}}(z_0)\|$, which implies that

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

by Assumption A7. Thus, we have

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

and

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

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

$$\begin{aligned}
& P\left[\max_{1\le k\le N(n)}\left|\frac{1}{\sqrt{nh}}\sum_{\mathfrak{s}=m+1}^{n}\{\boldsymbol{v}^T[\xi_{\mathfrak{s}}(v_k) - E(\xi_{\mathfrak{s}}(v_k))]\}\right| > \epsilon\right] \\
\le\ & N(n)\max_k P\left[\left|\frac{1}{\sqrt{nh}}\sum_{\mathfrak{s}=m+1}^{n}\{\boldsymbol{v}^T[\xi_{\mathfrak{s}}(v_k) - E(\xi_{\mathfrak{s}}(v_k))]\}\right| > \epsilon\right] \\
\le\ & N(n)\max_k P\left[\left|\sum_{\mathfrak{s}=m+1}^{n}\boldsymbol{v}^T\eta_{\mathfrak{s}}(v_k)\right| > \sqrt{nh}\epsilon, W_n^2 + S_n^2 \le L\right] \\
& + N(n)\max_k P\left[\left|\sum_{\mathfrak{s}=m+1}^{n}\boldsymbol{v}^T\eta_{\mathfrak{s}}(v_k)\right| > \sqrt{nh}\epsilon, W_n^2 + S_n^2 > L\right] \equiv J_1 + J_2. \qquad \text{(A.5)}
\end{aligned}$$

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

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

For J_2, because $P[W_n^2 + S_n^2 > L] \le P[W_n^2 > L] + P[S_n^2 > L]$ 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

$$
\begin{aligned}
M_2 \equiv & \sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T E\{\xi_{\mathfrak{s}}(C(m/nh)^{1/2}\boldsymbol{v})\} \\
= & \sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T E\{\psi_\tau(Y_{\mathfrak{s}}(z_0) - (C(m/nh)^{1/2}\boldsymbol{v})^T \boldsymbol{W}_{\mathfrak{s}}(z_0))\boldsymbol{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h}) \\
& - \psi_\tau(Y_{\mathfrak{s}}(z_0))\boldsymbol{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h})\} \\
= & \sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T E\{[F_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}0h}\boldsymbol{\alpha}^{(1)}(z_0)^T\boldsymbol{W}_{\mathfrak{s}}|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}}) \\
& - F_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}0h}\boldsymbol{\alpha}^{(1)}(z_0)^T\boldsymbol{W}_{\mathfrak{s}} \\
& + C(m/nh)^{1/2}\boldsymbol{v}^T\boldsymbol{W}_{\mathfrak{s}}(z_0)|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}})]\boldsymbol{W}_{\mathfrak{s}}(z_0)K(z_{\mathfrak{s}0h})\} \\
= & - C(m/nh)^{1/2}\sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T E\{f_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}0h}\boldsymbol{\alpha}^{(1)}(z_0)^T\boldsymbol{W}_{\mathfrak{s}} \\
& + \eta C(m/nh)^{1/2}\boldsymbol{v}^T\boldsymbol{W}_{\mathfrak{s}}(z_0)|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}})\boldsymbol{W}_{\mathfrak{s}}(z_0)\boldsymbol{W}_{\mathfrak{s}}^T(z_0)K(z_{\mathfrak{s}0h})\}\boldsymbol{v},
\end{aligned}
$$

where $\boldsymbol{W}_{\mathfrak{s}}(z_0)\boldsymbol{W}_{\mathfrak{s}}^T(z_0) = \begin{pmatrix} 1 & z_{\mathfrak{s}0h} \\ z_{\mathfrak{s}0h} & z_{\mathfrak{s}0h}^2 \end{pmatrix} \otimes \boldsymbol{W}_{\mathfrak{s}}\boldsymbol{W}_{\mathfrak{s}}^T$. Similar to the idea in Xu (2005),

$$
\begin{aligned}
& f_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}}) + h z_{\mathfrak{s}0h}\boldsymbol{\alpha}^{(1)}(z_0)^T\boldsymbol{W}_{\mathfrak{s}} + \eta C(m/nh)^{1/2}\boldsymbol{v}^T\boldsymbol{W}_{\mathfrak{s}}(z_0)|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}}) \\
= & f_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}})|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}}) + Chz_{\mathfrak{s}0h}\boldsymbol{\alpha}^{(1)}(z_0)^T\boldsymbol{W}_{\mathfrak{s}} + o_p(h),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
& \sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T E\{\xi_{\mathfrak{s}}(C(m/nh)^{1/2}\boldsymbol{v})\} \\
\approx & - C(m/nh)^{1/2}\sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T E[f_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}})|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}})\boldsymbol{W}_{\mathfrak{s}}(z_0)\boldsymbol{W}_{\mathfrak{s}}^T(z_0)K(z_{\mathfrak{s}0h})]\boldsymbol{v} \\
& - C(m/nh)^{1/2}h^2\sum_{\mathfrak{s}=m+1}^{n} \boldsymbol{v}^T\{E[|\boldsymbol{\alpha}^{(1)}(z_0)^T\boldsymbol{W}_{\mathfrak{s}}|\boldsymbol{W}_{\mathfrak{s}}(z_0)\boldsymbol{W}_{\mathfrak{s}}^T(z_0)K_h(z_{\mathfrak{s}0h})]\}\boldsymbol{v} \\
= & M_{21} + M_{22}.
\end{aligned}
$$

Again, by Assumption A7,

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

Hence,

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

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

$$\begin{aligned}
M_2 \approx& -C(m/nh)^{1/2}\sum_{\mathfrak{s}=m+1}^{n}\boldsymbol{v}^T E[f_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}})|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}})\boldsymbol{W}_{\mathfrak{s}}(z_0)\boldsymbol{W}_{\mathfrak{s}}^T(z_0)K(z_{\mathfrak{s}0h})]\boldsymbol{v}\\
=& -\boldsymbol{v}^T\begin{pmatrix} L_0 & L_1\\ L_1 & L_2\end{pmatrix}\boldsymbol{v},
\end{aligned}$$

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

$$\begin{aligned}
L_d =& -C(mn)^{1/2}h^{-1/2}E[f_{Y|Z,\boldsymbol{W}}(q_\tau(z_0,\boldsymbol{W}_{\mathfrak{s}})|Z_{\mathfrak{s}},\boldsymbol{W}_{\mathfrak{s}})z_{\mathfrak{s}0h}^d\boldsymbol{W}_{\mathfrak{s}}\boldsymbol{W}_{\mathfrak{s}}^T K(z_{\mathfrak{s}0h})]\\
=& -C(mn)^{1/2}h^{-1/2}E[\boldsymbol{D}^*(z_0)z_{\mathfrak{s}0h}^d K(\frac{Z_{\mathfrak{s}}-z_0}{h})]\\
=& -C(mnh)^{1/2}\int \boldsymbol{D}^*(z_0+hz)z^dK(z)f_z(z_0+hz)dz \approx -C(mnh)^{1/2}\mu_d f_z(z_0)\boldsymbol{D}^*(z_0).
\end{aligned}$$

In addition, for M_1, since

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

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

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

Therefore, by Assumptions A7 and A10, one has

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

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

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

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

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

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

Since $\mathscr{D}$ is compact, it can be covered by a finite number $T(n)$ of cubes $\mathscr{D}_j=\mathscr{D}_{n,j}$ 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

$$
\begin{aligned}
&\sup_{z\in\mathscr{D}}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}\boldsymbol{v}^T[\psi_\tau(Y_{\mathfrak{s}}(z)-v^T\boldsymbol{W}_{\mathfrak{s}}(z))]\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\\
\le&\sup_{z\in\mathscr{D}}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}E\{\boldsymbol{v}^T\psi_\tau(Y_{\mathfrak{s}}(z)-v^T\boldsymbol{W}_{\mathfrak{s}}(z))\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\}\\
&+\sup_{z\in\mathscr{D}}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}\Big[\boldsymbol{v}^T\psi_\tau(Y_{\mathfrak{s}}(z)-v^T\boldsymbol{W}_{\mathfrak{s}}(z))\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\\
&-E\{\boldsymbol{v}^T\psi_\tau(Y_{\mathfrak{s}}(z)-v^T\boldsymbol{W}_{\mathfrak{s}}(z))\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\}\Big]\equiv K^{(1)}+K^{(2)}.
\end{aligned}
$$

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

$$
\begin{aligned}
K^{(2)}\equiv&\sup_{z\in\mathscr{D}}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}\Big[\boldsymbol{v}^T\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}-E\{\boldsymbol{v}^T\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\}\Big]\\
\le&\max_{1\le j\le T(n)}\sup_{v\in\mathscr{B}}\bigg|\sum_{\mathfrak{s}=m+1}^{n}\boldsymbol{v}^T\Big[\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)K_{z_j,h}-E\{\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)K_{z_j,h}\}\Big]\bigg|\\
&+\max_{1\le j\le T(n)}\sup_{z\in\mathscr{D}_j}\sup_{v\in\mathscr{B}}\bigg|\sum_{\mathfrak{s}=m+1}^{n}\Big[\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}-\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)K_{z_j,h}]\\
&-E\{\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}-\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)K_{z_j,h}]\}\Big]\bigg|\equiv H^{(1)}+H^{(2)}.
\end{aligned}
$$

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,

$$
\begin{aligned}
H^{(2)}=&\max_{1\le j\le T(n)}\sup_{z\in\mathscr{D}_j}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}\bigg\{\bigg|\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)-\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)]K_{z_j,h}\\
&-E\{\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)-\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)]K_{z_j,h}\}\bigg|\bigg\}\\
&+\max_{1\le j\le T(n)}\sup_{z\in\mathscr{D}_j}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}\bigg\{\bigg|\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)-\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)]\\
&\times[K_{z,h}-K_{z_j,h}]-E\{\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z)-\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)]\\
&\times[K_{z,h}-K_{z_j,h}]\}\bigg|\bigg\}\\
&+\max_{1\le j\le T(n)}\sup_{z\in\mathscr{D}_j}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}\bigg\{\bigg|\boldsymbol{v}^T\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)[K_{z,h}-K_{z_j,h}]\\
&-E\{\boldsymbol{v}^T\psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)[K_{z,h}-K_{z_j,h}]\}\bigg|\bigg\}\equiv H^{(21)}+H^{(22)}+H^{(23)}.
\end{aligned}
$$

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

$$E\{|[\psi_{\tau,\mathfrak{s}}(z,v) - \psi_{\tau,\mathfrak{s}}(z_j,v)]|^{\flat}|Z_{\mathfrak{s}}, \boldsymbol{W}_{\mathfrak{s}}\} \le Cm^{1/2}l_n$$

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

$$\begin{aligned}
&E\{|\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z) - \psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j)]K_{z_j,h}|^{\delta}\}\\
=&E\{|\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z,v) - \psi_{\tau,\mathfrak{s}}(z_j,v)]\boldsymbol{W}_{\mathfrak{s}}(z)K_{z_j,h}|^{\delta}\}\\
&+ E\{|\boldsymbol{v}^T[\psi_{\tau,\mathfrak{s}}(z_j,v)](\boldsymbol{W}_{\mathfrak{s}}(z) - \boldsymbol{W}_{\mathfrak{s}}(z_j))K_{z_j,h}|^{\delta}\}\\
\le&E\{|\psi_{\tau,\mathfrak{s}}(z,v) - \psi_{\tau,\mathfrak{s}}(z_j,v)|^{\delta}|\boldsymbol{v}^T\boldsymbol{W}_{\mathfrak{s}}(z)|^{\delta}K^{\delta}_{z_j,h}\}\\
&+ E\{|\psi_{\tau,\mathfrak{s}}(z_j,v)|^{\delta}|\boldsymbol{v}^T(\boldsymbol{W}_{\mathfrak{s}}(z) - \boldsymbol{W}_{\mathfrak{s}}(z_j))|^{\delta}K^{\delta}_{z_j,h}\} \le Cl_n^{\delta}m^{\delta/2}h
\end{aligned}$$

by the boundedness of $\psi_{\tau,\mathfrak{s}}(z_j,v)$ uniformly over $v \in \mathscr{B}$. Define

$$\Delta\psi_{\tau,\mathfrak{s}}(z,z_j) = \psi_{\tau,\mathfrak{s}}(z,v)\boldsymbol{W}_{\mathfrak{s}}(z) - \psi_{\tau,\mathfrak{s}}(z_j,v)\boldsymbol{W}_{\mathfrak{s}}(z_j).$$

Thus, we have

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

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

and

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

Now, let $\chi_{\mathfrak{s}}(z_j) = \Delta\psi_{\tau,\mathfrak{s}}(z,z_j)K_{z_j,h} - E\{\Delta\psi_{\tau,\mathfrak{s}}(z,z_j)K_{z_j,h}\}$. Thus, the fact that

$$\Delta\psi_{\tau,\mathfrak{s}}(z,z_j)K_{z_j,h} - E\{\Delta\psi_{\tau,\mathfrak{s}}(z,z_j)K_{z_j,h}\}$$

is a martingale difference sequence implies that

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

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

$$
\begin{aligned}
K^{(1)}\equiv&\sup_{z\in\mathscr{D}}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}E\{\boldsymbol{v}^T\psi_\tau(Y_{\mathfrak{s}}(z)-v^T\boldsymbol{W}_{\mathfrak{s}}(z))\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\}\\
\le&\sup_{z\in\mathscr{D}}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}E\{\boldsymbol{v}^T[\psi_\tau(Y_{\mathfrak{s}}(z)-v^T\boldsymbol{W}_{\mathfrak{s}}(z))-\psi_\tau(Y_{\mathfrak{s}}(z))]\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\}\\
&+\sup_{z\in\mathscr{D}}\sup_{v\in\mathscr{B}}\sum_{\mathfrak{s}=m+1}^{n}E\{\boldsymbol{v}^T\psi_\tau(Y_{\mathfrak{s}}(z))\boldsymbol{W}_{\mathfrak{s}}(z)K_{z,h}\}\equiv K^{(11)}+K^{(12)}.
\end{aligned}
$$

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 $z\in\mathscr{D}$ and $v\in\mathscr{B}$. Therefore, the proof of Lemma A.2 is finished. □

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

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

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

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

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

$$M_{n\mathfrak{s}}^{(1)}(\boldsymbol{\zeta}) = [\psi_\tau(Y_{n\mathfrak{s}}^*) - \psi_\tau(Y_{\mathfrak{s}}^*)]\boldsymbol{W}_{\mathfrak{s}} K(z_{\mathfrak{s}h})$$

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

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

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

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

which implies by Assumption A7 that

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

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 $\boldsymbol{\zeta}_k$ and side length $C(m/nh)^{1/2}$, so that $N(n) = \#(\mathcal{C}) = (2(nh)^2)^m$, and for $\boldsymbol{\zeta} \in \mathcal{C}_k$, $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}_k\| \le C(m/(nh)^{5/2})$. Since $I(Y_{\mathfrak{s}}^* < x)$ is nondecreasing in x, then,

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

where $b_{n\mathfrak{s}}^{(1a_1)}(\boldsymbol{\zeta}_k) = I(Y_{\mathfrak{s}}^* < \boldsymbol{\zeta}_k^T \boldsymbol{W}_{\mathfrak{s}}) - I(Y_{\mathfrak{s}}^* < \boldsymbol{\zeta}_k^T \boldsymbol{W}_{\mathfrak{s}} + C(m/(nh)^{5/2})\|\boldsymbol{W}_{\mathfrak{s}}\|)$ and
$d_{nt}^{(1a_1)}(\boldsymbol{\zeta}_k) = I(Y_{\mathfrak{s}}^* < \boldsymbol{\zeta}_k^T \boldsymbol{W}_{\mathfrak{s}} + C(m/(nh)^{5/2})\|\boldsymbol{W}_{\mathfrak{s}}\|) - I(Y_{\mathfrak{s}}^* < \boldsymbol{\zeta}_k^T \boldsymbol{W}_{\mathfrak{s}} - C(m/(nh)^{5/2})\|\boldsymbol{W}_{\mathfrak{s}}\|)$.

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

$$\begin{aligned}
Q_n^2 &= \sum_{\mathfrak{s}=m+1\neq t}^{n} E\{M_{n\mathfrak{s}}^{(1a_1)}(\boldsymbol{\zeta}_k) - E(M_{n\mathfrak{s}}^{(1a_1)}(\boldsymbol{\zeta}_k))\}^2 \\
&\le \sum_{\mathfrak{s}=m+1\neq t}^{n} E[M_{n\mathfrak{s}}^{(1a_1)}(\boldsymbol{\zeta}_k)]^2 = O((mnh)^{1/2} m^{3/2})
\end{aligned}$$

and

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

Also, notice that $\varphi_{\mathfrak{s}}(\boldsymbol{\zeta}_k) = \{M_{n\mathfrak{s}}^{(1)}(\boldsymbol{\zeta}_k) - E(M_{n\mathfrak{s}}^{(1)}(\boldsymbol{\zeta}_k))\}$ is a martingale difference sequence. Therefore, let $L = (mnh)^{1/2}$, we have

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

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

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

For K_{12}, because

$$
P\left[Q_n^2+R_n^2>L\right]\le P\left[Q_n^2>L\right]+P\left[R_n^2>L\right]
$$

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

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

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

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

Proof. First, notice that

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

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

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

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

Hence, it follows that

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

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

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

$$= -\frac{1}{h}\int \boldsymbol{D}^*(z)\sqrt{nh}\boldsymbol{\zeta}(\frac{z-Z_t}{h})^d K(\frac{z-Z_t}{h}) f_z(z)dz$$

$$- CE\{|\boldsymbol{\alpha}^{(1)}(Z_t)^T\boldsymbol{W}_{\mathfrak{s}}|z_{\mathfrak{s}h}^d \boldsymbol{W}_{\mathfrak{s}}\boldsymbol{W}_{\mathfrak{s}}^T\sqrt{nh}\boldsymbol{\zeta}K(z_{\mathfrak{s}h})\} + o(1)$$

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

$$- -CE\{|\boldsymbol{\alpha}^{(1)}(Z_t)^T\boldsymbol{W}_{\mathfrak{s}}|z_{\mathfrak{s}h}^d \boldsymbol{W}_{\mathfrak{s}}\boldsymbol{W}_{\mathfrak{s}}^T\sqrt{nh}\boldsymbol{\zeta}K(z_{\mathfrak{s}h})\} + o(1),$$

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

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

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

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

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

and

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

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

$$Var[T_n(0)] \to \tau(1-\tau)f_z(Z_t)\boldsymbol{D}_1(Z_t).$$

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,

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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,*

$$\hat{\boldsymbol{\vartheta}} \equiv \sqrt{nh_1}\hat{\boldsymbol{\zeta}} = \frac{1}{\sqrt{nh_1}f_z(Z_t)}(\boldsymbol{D}_1^*(Z_t)^{-1})\sum_{\mathfrak{s}=m+1\neq t}^{n}\psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{W}_{\mathfrak{s}}^*K(z_{\mathfrak{s}h_1}) + o_p(1),$$

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

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 $D = f_z(Z_t)\boldsymbol{D}_1^*(Z_t)$. 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 Δ_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 $V_n(\Delta_n, \boldsymbol{\vartheta}_n) + V_n(\boldsymbol{\vartheta}_n) + D\Delta_n - A_n = o_p(1)$. This completes the proof. $\square$

To show Lemmas A.8 and A.9 later, τ is dropped from $\boldsymbol{g}_\tau(z_0)$ and h_2 is written as h for simplicity. For the notational convenience again, define $b_n = (nh_2)^{-1/2}$, let $\boldsymbol{\theta}_0 = b_n^{-1}(\Theta_0 - \boldsymbol{g}(z_0))$ and $\boldsymbol{\theta}_1 = hb_n^{-1}(\Theta_1 - \boldsymbol{g}^{(1)}(z_0))$. Then, $\boldsymbol{\theta} = b_n^{-1}\boldsymbol{H}_2\begin{pmatrix}\Theta_0 - \boldsymbol{g}(z_0) \\ \Theta_1 - \boldsymbol{g}^{(1)}(z_0)\end{pmatrix}$, where

$\boldsymbol{H}_2 = \text{diag}\{I_{2\kappa+1}, h_2 I_{2\kappa+1}\}$. For convenience of analysis, we rewrite $\hat{\boldsymbol{X}}_t \equiv \boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}}_0) \equiv \boldsymbol{X}_t(\boldsymbol{\alpha}(Z_t) + (nh_1)^{-1/2}\hat{\boldsymbol{\vartheta}}_0)$ because it contains $\hat{q}_{\tau,t} = \boldsymbol{W}_t^T\hat{\boldsymbol{\delta}}_0$. Similarly, $\boldsymbol{X}_t(\boldsymbol{\vartheta}_0) \equiv \boldsymbol{X}_t(\boldsymbol{\alpha}(Z_t) + (nh_1)^{-1/2}\boldsymbol{\vartheta}_0)$, $\boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0) \equiv \boldsymbol{X}_t^*(\boldsymbol{\alpha}(Z_t) + (nh_1)^{-1/2}\boldsymbol{\vartheta}_0)$ and $\hat{\boldsymbol{X}}_t^* \equiv \boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}_0) \equiv \boldsymbol{X}_t^*(\boldsymbol{\alpha}(Z_t) + (nh_1)^{-1/2}\hat{\boldsymbol{\vartheta}}_0)$, where $\boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0) = \begin{pmatrix} \boldsymbol{X}_t(\boldsymbol{\vartheta}_0) \\ z_{th}\boldsymbol{X}_t(\boldsymbol{\vartheta}_0) \end{pmatrix}$ and $\boldsymbol{X}_t^*(\hat{\boldsymbol{\vartheta}}_0) = \begin{pmatrix} \boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}}_0) \\ z_{th}\boldsymbol{X}_t(\hat{\boldsymbol{\vartheta}}_0) \end{pmatrix}$ and $z_{th} = (Z_t - z_0)/h$. Of course, $\boldsymbol{X}_t^*(0) \equiv \boldsymbol{X}_t^* = \begin{pmatrix} \boldsymbol{X}_t \\ z_{th}\boldsymbol{X}_t \end{pmatrix}$. Hence, $\partial \boldsymbol{X}_t(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\vartheta}_0 = a_n\boldsymbol{\Pi}_t$, where $\boldsymbol{\Pi}_t^T = (0^T_{1\times(\kappa m+1)}, \boldsymbol{W}_t, \ldots, \boldsymbol{W}_t, 0^T_{\kappa\times(\kappa m+1)})$ has the same definition as that in the main article. Next, denote $v_t^*(\boldsymbol{\vartheta}_0) = Y_t - \boldsymbol{X}_t^T(\boldsymbol{\vartheta}_0)[\boldsymbol{g}(z_0) + \boldsymbol{g}^{(1)}(z_0)(Z_t - z_0)]$, $v_t^*(0) = Y_t - \boldsymbol{X}_t^T[\boldsymbol{g}(z_0) + \boldsymbol{g}^{(1)}(z_0)(Z_t - z_0)]$ and $v_{nt}^* = v_{nt}^*(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0) = v_t^*(\boldsymbol{\vartheta}_0) - b_n\boldsymbol{\theta}^T\boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0)$. In addition, define $\Gamma^*(Z_t) = E[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0, \boldsymbol{X}_t))\boldsymbol{X}_t^*\boldsymbol{g}_\tau(z_0)^T\boldsymbol{\Pi}_t|Z_t]$ and $\Gamma(Z_t) = E[f_{Y|Z,\boldsymbol{X}}(q_\tau(z_0, \boldsymbol{X}_t))\boldsymbol{X}_t\ \boldsymbol{g}_\tau(z_0)^T\boldsymbol{\Pi}_t|Z_t]$. Again, let $A_m = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \le M\}$ and $B_m = \{\boldsymbol{\vartheta}_0 : \|\boldsymbol{\vartheta}_0\| \le \mathfrak{L}\}$ for some $0 < M < \infty$ and for some $0 < \mathfrak{L} < \infty$, Therefore,

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

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

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

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

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

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

where $V_{nt}(\boldsymbol{\theta},\boldsymbol{\vartheta}_0) = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] \boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0) K(z_{th}) = \left(V_{nt}^{(1)T}, V_{nt}^{(2)T}\right)^T$, $U_{nt}(\boldsymbol{\theta},\boldsymbol{\vartheta}_0) = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] (\boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0) - \boldsymbol{X}_t^*) K(z_{th}) = \left(U_{nt}^{(1)T}, U_{nt}^{(2)T}\right)^T$,

$$
W_{nt}(\boldsymbol{\theta},\boldsymbol{\vartheta}_0) = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))] \boldsymbol{X}_t^* K(z_{th}) = \left(W_{nt}^{(1)T}, W_{nt}^{(2)T}\right)^T,
$$

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

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

$V_{nt}^{(2)} = [\psi_\tau(v_{nt}^*) - \psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] \boldsymbol{X}_t(\boldsymbol{\vartheta}_0) z_{th} K(z_{th})$, $U_{nt}^{(1)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] (\boldsymbol{X}_t(\boldsymbol{\vartheta}_0) - \boldsymbol{X}_t) K(z_{th})$, and

$$
U_{nt}^{(2)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0))] (\boldsymbol{X}_t(\boldsymbol{\vartheta}_0) - \boldsymbol{X}_t) z_{th} K(z_{th}).
$$

In addition, $W_{nt}^{(1)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))] \boldsymbol{X}_t K(z_{th})$,

$$
W_{nt}^{(2)} = [\psi_\tau(v_t^*(\boldsymbol{\vartheta}_0)) - \psi_\tau(v_t^*(0))] \boldsymbol{X}_t z_{th} K(z_{th}),
$$

$R_{nt}^{(1)} = a_n \Gamma(Z_t) \boldsymbol{\vartheta}_0 K(z_{th})$ and $R_{nt}^{(2)} = a_n \Gamma(Z_t) \boldsymbol{\vartheta}_0 z_{th} K(z_{th})$. Thus,

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

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

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

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

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

with $d_n \to \infty$ 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

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

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 $p_{2t} = [\boldsymbol{g}_\tau(z_0) + \boldsymbol{g}_\tau^{(1)}(z_0)(Z_t - z_0)]^T \boldsymbol{X}_t(\boldsymbol{\vartheta}_0)$ and $p_{3t} = \frac{1}{\sqrt{nh}}\boldsymbol{\theta}^T \boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0)$. Therefore, by Assumption A4, there exists a $C > 0$ such that

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

which implies by Assumption A9 that

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

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 $\|\boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0)\| \leq C\|\boldsymbol{X}_t^*\|$. Then,

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

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

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

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 $d_n h \to 0$. 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

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

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

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

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

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

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

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

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 $(j_1\hbar M, \ldots,$ $j_{2(2\kappa+1)}\hbar M)$ and $(i_1\Bbbk\mathfrak{L}, \ldots, i_{2(2\kappa+1)}\Bbbk\mathfrak{L})$, where $j_k = 0, \pm1, \ldots, \pm[1/\hbar] + 1$, $i_k = 0, \pm1, \ldots, \pm[1/\Bbbk] + 1$, $[\cdot]$ denotes taking integer part of $\cdot$, and $\hbar$ and $\Bbbk$ are fixed positive small numbers. Denote $D(\boldsymbol{\theta})$ and $D(\boldsymbol{\vartheta}_0)$ the lower vertex of cubes that contain $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}_0$, respectively. Then,

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

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

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

$$2Cb_n \sup_{\boldsymbol{\vartheta}_0 \in B_m, \boldsymbol{\theta} \in A_m} \| \sum_{t=1}^{n} [I_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}} - EI_{\{|v_{nt}^*(D(\boldsymbol{\theta}), D(\boldsymbol{\vartheta}_0))| < \frac{C\max\{\hbar,\Bbbk\}}{\sqrt{nh}}\}}]$$

$$\times \boldsymbol{X}_t^*(D(\boldsymbol{\vartheta}_0))K(z_{th})\| = o_p(1)$$

by following the same steps as in (A.10). Let $\max\{\hbar, \Bbbk\} \to 0$. Then, it follows that the first term of H_3 is $o_p(1)$. As for the second term of H_3, in the same way as in (A.11),

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

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

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

It is easy to see that by following the same deduction as in (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,

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

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

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

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

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

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

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

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

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

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

In the same way as in analyzing (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. □

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

Proof. Notice that

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

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,

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

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

R_{11} as follows

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

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

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

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

$$\begin{aligned}
R_{13} \equiv& b_n \sum_{t=1}^{n} E\{[\psi_\tau(v_{nt}^*(0, \boldsymbol{\vartheta}_0)) - \psi_\tau(v_{nt}^*(0, 0))]\boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \\
=& b_n \sum_{t=1}^{n} E\{[F_{Y|Z,\boldsymbol{X}}(q_\tau(z_0, \boldsymbol{X}_t) + hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)^T\boldsymbol{X}_t|Z_t, \boldsymbol{X}_t) \\
&- F_{Y|Z,\boldsymbol{X}}(q_\tau(z_0, \boldsymbol{X}_t(\boldsymbol{\vartheta}_0)) + hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)^T\boldsymbol{X}_t(\boldsymbol{\vartheta}_0)|Z_t, \boldsymbol{X}_t)]\boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0)K(z_{th})\} \\
=& -b_n \sum_{t=1}^{n} E\{[f_{Y|Z,\boldsymbol{X}}(\tilde{\boldsymbol{X}}_t^T(\boldsymbol{g}_\tau(z_0) + hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0))|Z_t, \boldsymbol{X}_t)] \\
&\times \boldsymbol{X}_t^*(\boldsymbol{\vartheta}_0)(\boldsymbol{X}_t(\boldsymbol{\vartheta}_0) - \boldsymbol{X}_t)^T[\boldsymbol{g}_\tau(z_0) + hz_{th}\boldsymbol{g}_\tau^{(1)}(z_0)]K(z_{th})\} \\
=& -b_n \sum_{t=1}^{n} E\{\Gamma^*(Z_t)a_n\boldsymbol{\vartheta}_0K(z_{th})\} + o(h)
\end{aligned}$$

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

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

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

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

and

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

where $H_1(z_0) = E[\boldsymbol{X}_1\boldsymbol{W}_1^T|Z_1 = z_0](\boldsymbol{D}^*(z_0))^{-1}\Gamma^T(z_0) + \Gamma(z_0)(\boldsymbol{D}^*(z_0))^{-1}E[\boldsymbol{W}_1\boldsymbol{X}_1^T|Z_1 = z_0]$ *and* $H_2(z_0) = \Gamma(z_0)(\boldsymbol{D}^*(z_0))^{-1}\boldsymbol{D}(z_0)(\boldsymbol{D}^*(z_0))^{-1}\Gamma^T(z_0)$*. Then,*

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

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

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

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

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

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

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

As for $E[B_1B_1^T]$, we have

$$
\begin{aligned}
E[B_1B_1^T] = E\Bigg(\Big\{&\psi_\tau^2(v_1^*(0))\boldsymbol{X}_1^*\boldsymbol{X}_1^{*T} - [\psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\boldsymbol{X}_1^*\boldsymbol{W}_1^T(\boldsymbol{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1)\\
&+ \psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\Gamma^*(Z_1)(\boldsymbol{D}^*(Z_1))^{-1}\boldsymbol{W}_1\boldsymbol{X}_1^{*T}]\\
&+ \psi_\tau^2(Y_1^*)\Gamma^*(Z_1)(\boldsymbol{D}^*(Z_1))^{-1}\boldsymbol{W}_1\boldsymbol{W}_1^T(\boldsymbol{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1)\Big\}K^2(z_{1h_2})\Bigg)\\
=&E\{\psi_\tau^2(v_1^*(0))\boldsymbol{X}_1^*\boldsymbol{X}_1^{*T}K^2(z_{1h_2})\}\\
&- E\{[\psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\boldsymbol{X}_1^*\boldsymbol{W}_1^T(\boldsymbol{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1)\\
&+ \psi_\tau(v_1^*(0))\psi_\tau(Y_1^*)\Gamma^*(Z_1)(\boldsymbol{D}^*(Z_1))^{-1}\boldsymbol{W}_1\boldsymbol{X}_1^{*T}]K^2(z_{1h_2})\}\\
&+ E\{\psi_\tau^2(Y_1^*)\Gamma^*(Z_1)(\boldsymbol{D}^*(Z_1))^{-1}\boldsymbol{W}_1\boldsymbol{W}_1^T(\boldsymbol{D}^*(Z_1))^{-1}\Gamma^{*T}(Z_1)K^2(z_{1h_2})\}\\
\equiv&P^{(1)} + P^{(2)} + P^{(3)}.
\end{aligned}
$$

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

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

Similarly,

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

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

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

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

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

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

Therefore,

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

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

$$Var\left\{\frac{1}{\sqrt{nh_2}}\sum_{t=1}^{n}B_t\right\} = \frac{1}{h}[Var(B_1) + 2\sum_{\ell=1}^{n-1}(1-\frac{\ell}{n})Cov(B_1, B_{\ell+1})]$$

$$\leq \frac{1}{h}Var(B_1) + \frac{2}{h}\sum_{\ell=1}^{e_n-1}|Cov(B_1, B_{\ell+1})| + \frac{2}{h}\sum_{\ell=e_n}^{\infty}|Cov(B_1, B_{\ell+1})| \equiv G_1 + G_2 + G_3.$$

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

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

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. □

A.2 Proof of Theorem 2.2.1:

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 $\sup_{\|\Delta\|\leq M, \|\boldsymbol{\vartheta}_0\|\leq \mathfrak{L}}\|V_n(\Delta, \boldsymbol{\vartheta}_0) + V_n(\boldsymbol{\vartheta}_0) + D\Delta - A_n\| = o_p(1)$ 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,

$$
\begin{aligned}
&E[V_n(\hat{\boldsymbol{\vartheta}}_0)]\\
=&b_n\sum_{t=1}^{n}E\left\{\left[\Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1})\frac{f_z^{-1}(Z_t)}{nh_1}\sum_{\mathfrak{s}=m+1\neq t}^{n}\psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{W}_{\mathfrak{s}}K(z_{\mathfrak{s}h_1})\right]K(z_{th_2})\right\}\\
=&b_n\sum_{t=1}^{n}E\left\{\left[\Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1})\frac{f_z^{-1}(Z_t)}{nh_1}\sum_{\mathfrak{s}=m+1\neq t}^{n}\{\psi_\tau(Y_t^*)\boldsymbol{W}_t+\psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{W}_{\mathfrak{s}}\right.\right.\\
&\left.\left.-\psi_\tau(Y_t^*)\boldsymbol{W}_t\}K(z_{\mathfrak{s}h_1})\right]K(z_{th_2})\right\}\\
=&b_n\sum_{t=1}^{n}E\left\{\left[\psi_\tau(Y_t^*)\Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1})\boldsymbol{W}_t\frac{f_z^{-1}(Z_t)}{nh_1}\sum_{\mathfrak{s}=m+1\neq t}^{n}K(z_{\mathfrak{s}h_1})\right]K(z_{th_2})\right\}\\
&+b_n\sum_{t=1}^{n}E\left\{\left[\Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1})\frac{f_z^{-1}(Z_t)}{nh_1}\sum_{\mathfrak{s}=m+1\neq t}^{n}\{\psi_\tau(Y_{\mathfrak{s}}^*)\boldsymbol{W}_{\mathfrak{s}}-\psi_\tau(Y_t^*)\boldsymbol{W}_t\}\right.\right.\\
&\left.\left.\times K(z_{\mathfrak{s}h_1})\right]K(z_{th_2})\right\}\equiv T^{(1)}+T^{(2)}.
\end{aligned}
$$

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

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

by the fact that $f_z^{-1}(Z_t)(nh_1)^{-1}\sum_{\mathfrak{s}=m+1\neq t}^{n}K(z_{\mathfrak{s}h_1})=1+o(1)$ 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 $Var[V_n(\hat{\boldsymbol{\vartheta}}_0)]=o(1)$. 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

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

where $\hat{\boldsymbol{\theta}}$ is the minimizer of $J(\boldsymbol{\theta})$. Finally, because $\psi_\tau(x)$ is an increasing function of x;

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

$$\begin{aligned}
\hat{\boldsymbol{\theta}} =& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n [\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - a_n \Gamma^*(Z_t)\hat{\boldsymbol{\vartheta}}_0] K(z_{th_2}) + o_p(1) \\
=& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \Bigg[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1}) \\
& \times \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{\mathfrak{s}=m+1 \neq t}^n \psi_\tau(Y_{\mathfrak{s}}^*) \boldsymbol{W}_{\mathfrak{s}} K(z_{\mathfrak{s}h_1}) \Bigg] K(z_{th_2}) + o_p(1) \\
=& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \Bigg[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1}) \\
& \times \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{\mathfrak{s}=m+1 \neq t}^n \{\psi_\tau(Y_t^*)\boldsymbol{W}_t + \psi_\tau(Y_{\mathfrak{s}}^*) \boldsymbol{W}_{\mathfrak{s}} - \psi_\tau(Y_t^*)\boldsymbol{W}_t\} K(z_{\mathfrak{s}h_1}) \Bigg] K(z_{th_2}) + o_p(1) \\
=& \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \Bigg[\psi_\tau(v_t^*(0))\boldsymbol{X}_t^* - \psi_\tau(Y_t^*)\Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1})\boldsymbol{W}_t \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{\mathfrak{s}=m+1 \neq t}^n K(z_{\mathfrak{s}h_1}) \Bigg] \\
& \times K(z_{th_2}) - \frac{(\Omega_1^*(z_0))^{-1}}{\sqrt{nh_2} f_z(z_0)} \sum_{t=1}^n \Bigg[\Gamma^*(Z_t)(\boldsymbol{D}^*(Z_t)^{-1}) \\
& \times \frac{f_z^{-1}(Z_t)}{nh_1} \sum_{\mathfrak{s}=m+1 \neq t}^n \{\psi_\tau(Y_{\mathfrak{s}}^*) \boldsymbol{W}_{\mathfrak{s}} - \psi_\tau(Y_t^*)\boldsymbol{W}_t\} K(z_{\mathfrak{s}h_1}) \Bigg] K(z_{th_2}) + o_p(1).
\end{aligned}$$

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

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

by the fact that $f_z^{-1}(Z_t)(nh_1)^{-1} \sum_{\mathfrak{s}=m+1 \neq t}^n K(z_{\mathfrak{s}h_1}) = 1 + o(1)$. Therefore, following the proof of Theorem 1 in Cai and Xu (2008), the theorem is proved. □

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

Proof. We first focus on $\hat{\Gamma}(z_0)$ in Section 2.2.4. Notice that

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

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

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

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

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

Therefore, by Lemma A.2,

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

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,

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

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 2.2.4. 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. □

Appendix: Mathematical Proof for Stationarity and α-Mixing of Model (2.1)

In this section, we show that the model (2.1) in the main article can generate a strictly stationary and α-mixing process. Throughout this section, $0_{a\times b}$ stands for a $(a \times b)$ matrix of zeros and I_a is a $(a\times a)$ identity matrix. Next, we define $\psi(\cdot) = \|\cdot\|$, where $\|\cdot\|$ is the Euclidean norm. For a random vector Z and random matrix A, we denote $\|Z\|_{\psi,2} = [E\|Z\|^2]^{1/2}$ and $\|A\|_{\psi,2} = \sup_{z\neq 0} \|Az\|_{\psi,2}/\|z\|$. In addition, for $1 \le i \le \kappa$, let $\mathcal{F}_{i,a}^b$ be the σ-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 α-mixing (strongly mixing) if the mixing coefficient $\alpha(t)$ defined by

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

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

To study the probabilistic properties of model (2.1) in the main article, $\mathbb{Y}_t$ and $\boldsymbol{q}_{\tau,t}$ in (2.1) 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 (2.1) in the main article, denote U_{it} ($1 \le i \le \kappa$, $1 \le t \le n$) 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

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

and

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

for some p and q, where Y_{it}, $\boldsymbol{q}_{\tau,t}$ and $\mathbb{Y}_t$ in (B.1) and (B.2) have the same definition as that in (2.1) and equation (B.2) is the same as (2.1) 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 $\boldsymbol{\gamma}_{i,s}(\cdot,\cdot) = (\gamma_{si1}(\cdot,\cdot),\ldots,\gamma_{si\kappa}(\cdot,\cdot))^T$ and $\boldsymbol{\beta}_{i,l}(\cdot,\cdot) = (\beta_{li1}(\cdot,\cdot),\ldots,\beta_{li\kappa}(\cdot,\cdot))^T$ 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 $(Z_t, \{\boldsymbol{q}_{\tau,t-s}\}_{s=1}^q, \{\mathbb{Y}_{t-l}\}_{l=1}^p)$ 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

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

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

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

Here, for $s = 1,\ldots,q$ and $l = 1,\ldots,p$, $\boldsymbol{\Gamma}_{\beta,U_t}(Z_t) = (\boldsymbol{\Gamma}_{\beta,1,U_t}(Z_t),\ldots,\boldsymbol{\Gamma}_{\beta,p,U_t}(Z_t))$, where $\boldsymbol{\Gamma}_{\beta,l,U_t}(Z_t) = (\beta_{lij}(U_{it}, Z_t))_{1\le i\le\kappa,1\le j\le\kappa}$ is a $\kappa\times\kappa$ matrix. In addition,

$$\boldsymbol{\Gamma}_{U_t}(Z_t) = (\boldsymbol{\Gamma}_{1,U_t}(Z_t),\ldots,\boldsymbol{\Gamma}_{q,U_t}(Z_t)),$$

where $\boldsymbol{\Gamma}_{s,U_t}(Z_t) = (\gamma_{sij}(U_{it}, Z_t))_{1\le i\le\kappa,1\le j\le\kappa}$ is a $\kappa\times\kappa$ matrix. Similarly,

$$\boldsymbol{\Gamma}_{\beta,\tau}(Z_t) = (\boldsymbol{\Gamma}_{\beta,1,\tau}(Z_t),\ldots,\boldsymbol{\Gamma}_{\beta,p,\tau}(Z_t)),$$

where

$$\boldsymbol{\Gamma}_{\beta,l,\tau}(Z_t) = (\beta_{lij,\tau}(Z_t))_{1\le i\le\kappa,1\le j\le\kappa}$$

is a $\kappa \times \kappa$ matrix. Also, $\mathbf{\Gamma}_\tau(Z_t) = (\mathbf{\Gamma}_{1,\tau}(Z_t), \ldots, \mathbf{\Gamma}_{q,\tau}(Z_t))$, where

$$\mathbf{\Gamma}_{s,\tau}(Z_t) = (\gamma_{sij,\tau}(Z_t))_{1\le i\le\kappa, 1\le j\le\kappa}$$

is a $\kappa\times\kappa$ matrix. Furthermore, $\boldsymbol{\mu}(Z_t) = (E_U^T(\boldsymbol{\gamma}_0(U_{it}, Z_t)), 0, \ldots, 0, \boldsymbol{\gamma}_{0,\tau}^T(Z_t), 0, \ldots, 0)^T$, where $E_U(\boldsymbol{\gamma}_0(U_{it}, Z_t)) = (E_U(\gamma_{10}(U_{1t}, Z_t)), \ldots, E_U\ (\gamma_{\kappa 0}(U_{\kappa t}, Z_t)))^T$ and

$$\boldsymbol{\gamma}_{0,\tau}(Z_t) = (\gamma_{10,\tau}(Z_t), \ldots, \gamma_{\kappa 0,\tau}(Z_t))^T.$$

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, $\boldsymbol{D}_{U_t}(Z_t) = (\check{\gamma}_{10}(U_{1t}, Z_t), \ldots, \check{\gamma}_{\kappa 0}(U_{\kappa t}, Z_t), 0_{1\times\kappa(p+q-1)})^T$, 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, ..., ($\kappa p + \kappa$)-th rows of* (B.3) *are exactly the* (B.2) *and the model* (2.1) *in the main article, while the ith row of* (B.3) *with $i = 1, \ldots, \kappa$ is equation* (B.1). *Given these relations, one can conclude that $\mathbb{Y}_t$ and $\boldsymbol{q}_{\tau,t}$ jointly follow a VAR process of order 1 in* (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 $\lambda_{\max}(\boldsymbol{A}_{U_t})$ as the largest eigenvalue in absolute value of following matrix $\boldsymbol{A}_{U_t}$:

$$
\boldsymbol{A}_{U_t} = \begin{pmatrix}
\boldsymbol{\Gamma}_{\beta,1,U_t} & \boldsymbol{\Gamma}_{\beta,2,U_t} & \dots & \boldsymbol{\Gamma}_{\beta,p-1,U_t} & \boldsymbol{\Gamma}_{\beta,p,U_t} & \boldsymbol{\Gamma}_{1,U_t} & \boldsymbol{\Gamma}_{2,U_t} & \dots & \boldsymbol{\Gamma}_{q-1,U_t} & \boldsymbol{\Gamma}_{q,U_t} \\
I_\kappa & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\
0_{\kappa\times\kappa} & I_\kappa & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & I_\kappa & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\
\boldsymbol{\Gamma}_{\beta,1} & \boldsymbol{\Gamma}_{\beta,2} & \dots & \boldsymbol{\Gamma}_{\beta,p-1} & \boldsymbol{\Gamma}_{\beta,p} & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \dots & \boldsymbol{\Gamma}_{q-1} & \boldsymbol{\Gamma}_q \\
0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & I_\kappa & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\
0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & I_\kappa & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & 0_{\kappa\times\kappa} & \dots & I_\kappa & 0_{\kappa\times\kappa}
\end{pmatrix},
$$

where

$$
\boldsymbol{\Gamma}_{\beta,l,U_t} = \begin{pmatrix}
\beta_{l11}(U_{1t}) & \beta_{l12}(U_{1t}) & \dots & \beta_{l1\kappa}(U_{1t}) \\
\beta_{l21}(U_{2t}) & \beta_{l22}(U_{2t}) & \dots & \beta_{l2\kappa}(U_{2t}) \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{l\kappa1}(U_{\kappa t}) & \beta_{l\kappa2}(U_{\kappa t}) & \dots & \beta_{l\kappa\kappa}(U_{\kappa t})
\end{pmatrix},
$$

$$
\boldsymbol{\Gamma}_{s,U_t} = \begin{pmatrix}
\gamma_{s11}(U_{1t}) & \gamma_{s12}(U_{1t}) & \dots & \gamma_{s1\kappa}(U_{1t}) \\
\gamma_{s21}(U_{2t}) & \gamma_{s22}(U_{2t}) & \dots & \gamma_{s2\kappa}(U_{2t}) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{s\kappa1}(U_{\kappa t}) & \gamma_{s\kappa2}(U_{\kappa t}) & \dots & \gamma_{s\kappa\kappa}(U_{\kappa t})
\end{pmatrix},
$$

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

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

Assumption B.

B1: *Let $\{\mathbb{X}_t\}$ be a ϕ-irreducible and aperiodic Markov chain. For $i = 1, \ldots, \kappa$, $j = 1, \ldots, \kappa$, $l = 1, \ldots, p$ and $s = 1, \ldots, q$, each entry of $\mathbf{\Gamma}_{s,U_t}(Z_t)$ and $\mathbf{\Gamma}_{\beta,l,U_t}(Z_t)$ in* (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, $E\{[\lambda_{\max}(\boldsymbol{A}_{U_t})]^2\} < 1$.*

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

Remark B.2. *The ϕ-irreducibility and aperiodicity in Assumption B1 are key assumptions for deriving geometric ergodicity and subsequently, α-mixing property. The conditions that imply ϕ-irreducibility and aperiodicity of nonlinear time series have been studied extensively in literature. For example, Chan and Tong (1985) showed that under some mild conditions, a simple nonparametric autoregressive process is a ϕ-irreducible and aperiodic Markov chain. In addition, Pham (1986) obtained conditions for random coefficient autoregressive (RCA) models to be ϕ-irreducible. In this article, we simply impose the assumptions of ϕ-irreducibility and aperiodicity on $\{\mathbb{X}_t\}$, which are common settings among literature, see, for example, Chen and Tsay (1993). It is of particular interest to explore the conditions under which $\{\mathbb{X}_t\}$ is ϕ-irreducibility and aperiodicity and we leave this as a future topic. Moreover, the moment conditions $E\{[\lambda_{\max}(\boldsymbol{A}_{U_t})]^2\} < 1$ in Assumption B1 is used to bound the random matrices $\boldsymbol{A}_{U_t}(Z_t)$, which is similar to the condition in Carrasco and Chen (2002). We stress that we are not seeking to achieve the weakest possible regularity conditions for probabilistic properties of model* (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 α-mixing process.*

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

Lemma B.1. *Under Assumptions B1 and B2, for any $\mathbb{W} = (w_1, \ldots, w_{\kappa(p+q)})^T$, we have $\|\boldsymbol{A}_{U_t}(Z_t)\mathbb{W}\|_{\psi,2} \leq \|\boldsymbol{A}_{U_t}|\mathbb{W}|\|_{\psi,2}$. Here, $\boldsymbol{A}_{U_t}(Z_t)$ is defined in* (B.3)*, $\boldsymbol{A}_{U_t}$ is defined previously and $|\mathbb{W}| = (|w_1|, \ldots, |w_{\kappa(p+q)}|)^T$.*

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

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

and

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

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

Proof of Proposition B.1:

Proof. By Proposition 3 in Carrasco and Chen (2002) and Lemma 2 in Pham (1986), Assumption B1 implies $\|\boldsymbol{A}_{U_t}\|_{\psi,2} < 1$ for all $U_{it} \in [0,1]$. Then, we can find $0 < \delta < 1$ and $\varrho > 0$, such that $\|\prod_{\jmath=0}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}}\|_{\psi,2} < 1-\delta$. Consequently, by Assumption B2 and Lemma

B.1, for some constant $C > 0$,

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

where each element of $\boldsymbol{D}_{U_t} = (\check{\gamma}_{10}(U_{1t}), \ldots, \check{\gamma}_{\kappa 0}(U_{\kappa t}), 0_{1\times\kappa(p+q-1)})^T$ is defined in Assumption B2 and the first inequality follows from Jensen's inequality. Notice that

$$
E \left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] \right\|
$$

is bounded and by Assumption B2, $E\|\boldsymbol{D}_{U_t}\|$ is bounded, so that

$$
E \left\| \sum_{\jmath=1}^{\varrho} \left[\prod_{\imath=\jmath}^{\varrho-1} \boldsymbol{A}_{U_{t+\jmath}} \right] |\boldsymbol{D}_{U_{t+\jmath}}| \right\|
$$

is bounded and the bound does not depend on $\mathbb{X}$ and Z_t. Thus, we can find a sufficiently large $M > 0$ such that when $\|\mathbb{X}\| > M$,

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

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

Appendix: Some Assumptions and Descriptions of Dataset in Chapter 3

C.1 Descriptions of Dataset

All series are directly taken from the Federal Reserve Bank of St. Louis with a dataset proposed in McCracken and Ng (2016) and the format is as that in Bernanke et al. (2005): series number; series mnemonic; transformation code and series description as appearing in the database for the data span from 1960:02 to 2021:09. The transformation codes are 1-no transformation; 2-first difference; 4-logarithm; 5-first difference of logarithm. An asterisk $*$, next to the mnemonic, denotes a variable assumed to be slow-moving in the estimation.

Table C1: Description of data

	Real output and income	
1.	IPF*	5 INDUSTRIAL PRODUCTION: FINAL PRODUCTS (SA)
2.	IPC*	5 INDUSTRIAL PRODUCTION: CONSUMER GOODS (SA)
3.	IPCD*	5 INDUSTRIAL PRODUCTION: DURABLE CONS. GOODS (SA)
4.	IPCN*	5 INDUSTRIAL PRODUCTION: NONDURABLE CONS. GOODS (SA)

5.	IPE*	5 INDUSTRIAL PRODUCTION: BUSINESS EQUIPMENT (SA)
6.	IPM*	5 INDUSTRIAL PRODUCTION: MATERIALS (SA)
7.	IPMD*	5 INDUSTRIAL PRODUCTION: DURABLE GOODS MATERIALS (SA)
8.	IPMND*	5 INDUSTRIAL PRODUCTION: NONDUR. GOODS MATERIALS (SA)
9.	IPMFG*	5 INDUSTRIAL PRODUCTION: MANUFACTURING (SA)
10.	IPMIN*	5 INDUSTRIAL PRODUCTION: MINING (SA)
11.	IPUT*	5 INDUSTRIAL PRODUCTION: UTILITIES (SA)
12.	IP*	5 INDUSTRIAL PRODUCTION: TOTAL INDEX (SA)
13.	IPXMCA*	1 CAPACITY UTIL RATE: MANUFAC., TOTAL (SA)
14.	GMPYQ*	5 REAL PERSONAL INCOME (SAAR)
15.	GMYXPQ*	5 REAL PERSONAL INCOME EX TRANSFER RECEIPTS (SAAR)
	Employment and hours	
16.	LHEM*	5 CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS., SA)
17.	LHUR*	1 UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (SA)
18.	LHU680*	1 UNEMPLOY. BY DURATION: AVERAGE (MEAN) DURATION IN WEEKS (SA)
19.	LHU5*	1 UNEMPLOY. BY DURATION: PERS UNEMPL. LESS THAN 5 WKS (THOUS., SA)

20.	LHU14*	1 UNEMPLOY. BY DURATION: PERS UNEMPL. 5 TO 14 WKS (THOUS., SA)
21.	LHU15*	1 UNEMPLOY. BY DURATION: PERS UNEMPL. 15 WKS + (THOUS., SA)
22.	LHU26*	1 UNEMPLOY. BY DURATION: PERS UNEMPL. 15 TO 26 WKS (THOUS., SA)
23.	LPNAG*	5 EMPLOYEES ON NONAG. PAYROLLS: TOTAL (THOUS., SA)
24.	LPGD*	5 EMPLOYEES ON NONAG. PAYROLLS: GOODS-PRODUCING (THOUS., SA)
25.	LPMI*	5 EMPLOYEES ON NONAG. PAYROLLS: MINING (THOUS., SA)
26.	LPCC*	5 EMPLOYEES ON NONAG. PAYROLLS: CONTRACT CONSTRUC. (THOUS., SA)
27.	LPEM*	5 EMPLOYEES ON NONAG. PAYROLLS: MANUFACTURING (THOUS., SA)
28.	LPED*	5 EMPLOYEES ON NONAG. PAYROLLS: DURABLE GOODS (THOUS., SA)
29.	LPEN*	5 EMPLOYEES ON NONAG. PAYROLLS: NONDURABLE GOODS (THOUS., SA)
30.	LPSP*	5 EMPLOYEES ON NONAG. PAYROLLS: SERVICE-PRODUCING (THOUS., SA)
31.	LPTU*	5 EMPLOYEES ON NONAG. PAYROLLS: TRANS. & PUBLIC UTIL. (THOUS., SA)
32.	LPTW*	5 EMPLOYEES ON NONAG. PAYROLLS: WHOLESALE (THOUS., SA)
33.	LPTR*	5 EMPLOYEES ON NONAG. PAYROLLS: RETAIL (THOUS., SA)
34.	LPFR*	5 EMPLOYEES ON NONAG. PAYROLLS: FINANCE, INS. & REAL EST (THOUS., SA)
35.	LPGOV*	5 EMPLOYEES ON NONAG. PAYROLLS: GOVERNMENT (THOUS., SA)
36.	LPHRM*	1 AVG. WEEKLY HRS. OF PRODUCTION

		WKRS.: MANUFACTURING (SA)
37.	LPMOSA*	1 AVG. WEEKLY HRS. OF PROD. WKRS.: MFG., OVERTIME HRS. (SA)
38.	HWI*	2 HELP-WANTED INDEX FOR USA
39.	HWIURATIO*	2 RATIO OF HELP WANTED/NO. UNEMPLOYED
	Consumption	
40.	GMCQ*	5 PERSONAL CONSUMPTION EXPEND TOTAL (BIL, SAAR)
41.	GMCDQ*	5 PERSONAL CONSUMPTION EXPEND TOT. DUR. (BIL, SAAR)
42.	GMCNQ*	5 PERSONAL CONSUMPTION EXPEND NONDUR. (BIL, SAAR)
43.	GMCSQ*	5 PERSONAL CONSUMPTION EXPEND SERVICES (BIL, SAAR)
	Housing starts and sales	
44.	HOUST	4 HOUSING STARTS: TOTAL NEW PRIV
45.	HSNE	4 HOUSING STARTS: NORTHEAST (THOUS.U.) S.A.
46.	HSMW	4 HOUSING STARTS: MIDWEST (THOUS.U.) S.A.
47.	HSSOU	4 HOUSING STARTS: SOUTH (THOUS.U.) S.A.
48.	HSWST	4 HOUSING STARTS: WEST (THOUS.U.) S.A.
49.	HSBR	4 HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING (THOUS., SAAR)

	Real inventories and orders	
50.	AMDMNOx	5 NEW ORDERS FOR DURABLE GOODS
51.	AMDMUOx	5 UNFILLED ORDERS FOR DURABLE GOODS
52.	BUSINVx	5 TOTAL BUSINESS INVENTORIES
53.	ISRATIOx	2 TOTAL BUSINESS: INVENTORIES TO SALES RATIO
	Stock prices	
54.	FSPCOM	5 S&P'S COMMON STOCK PRICE INDEX: COMPOSITE
55.	FSPIN	5 S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS
56.	FSDXP	1 S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD
57.	FSDXE	1 S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO
	Exchange rates	
58.	EXRSW	5 FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U. S.$)
59.	EXRJAN	5 FOREIGN EXCHANGE RATE: JAPAN (YEN PER U. S.$)
60.	EXRUK	5 FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
61.	EXRCAN	5 FOREIGN EXCHANGE RATE: CANADA

		(CANADIAN $ PER U. S.$)
	Interest rates	
62.	FYFF	1 INTEREST RATE: FEDERAL FUNDS (EFFECTIVE)
63.	FYGM3	1 INTEREST RATE: U. S. TREASURY BILLS,SEC MKT,3-MO.
64.	FYGM6	1 INTEREST RATE: U. S. TREASURY BILLS,SEC MKT,6-MO.
65.	FYGT1	1 INTEREST RATE: U. S. TREASURY CONST MATUR., 1-YR.
66.	FYGT5	1 INTEREST RATE: U. S. TREASURY CONST MATUR., 5-YR.
67.	FYGT10	1 INTEREST RATE: U. S. TREASURY CONST MATUR., 10-YR.
68.	FYAAAC	1 BOND YIELD: MOODY'S AAA CORPORATE
69.	FYBAAC	1 BOND YIELD: MOODY'S BAA CORPORATE
70.	SFYGM3	1 Spread FYGM3-FYFF
71.	SFYGM6	1 Spread FYGM6-FYFF
72.	SFYGT1	1 Spread FYGT1-FYFF
73.	SFYGT5	1 Spread FYGT5-FYFF
74.	SFYGT10	1 Spread FYGT10-FYFF
75.	SFYAAAC	1 Spread FYAAAC-FYFF
76.	SFYBAAC	1 Spread FYBAAC-FYFF

	Money and credit quantity aggregates	
77.	FM1	5 MONEY STOCK: M1 (BIL$,SA)
78.	FM2	5 MONEY STOCK: M2 (BIL$,SA)
79.	FM3	5 MONEY STOCK: M3 (BIL$,SA)
80.	FMFBA	5 MONETARY BASE, ADJ FOR RESERVE (MIL$,SA)
81.	FMRRA	5 DEPOSITORY INST RESERVES: TOTAL, ADJ FOR RES. (MIL$, SA)
82.	FMRNBA	5 DEPOSITORY INST RESERVES: NONBOR., ADJ RES. (MIL$,SA)
83.	FCLNQ	5 COMMERCIAL & INDUST. LOANS
84.	CCINRV	5 TOTAL CONSUMER LOANS AND LEASES OUTSTANDING
	Price indexes	
85.	PWFSA*	5 PRODUCER PRICE INDEX: FINISHED GOODS (SA)
86.	PWFCSA*	5 PRODUCER PRICE INDEX: FINISHED CONSUMER GOODS (SA)
87.	PWIMSA*	5 PRODUCER PRICE INDEX: INTERMED MAT. SUP & COMPONENTS (SA)
88.	PWCMSA*	5 PRODUCER PRICE INDEX: CRUDE MATERIALS (SA)
89.	PUNEW*	5 CPI-U: ALL ITEMS (SA)

90.	PU83*	5 CPI-U: APPAREL & UPKEEP (SA)
91.	PU84*	5 CPI-U: TRANSPORTATION (SA)
92.	PU85*	5 CPI-U: MEDICAL CARE (SA)
93.	PUC*	5 CPI-U: COMMODITIES (SA)
94.	PUCD*	5 CPI-U: DURABLES (SA)
95.	PUS*	5 CPI-U: SERVICES (SA)
96.	PUXF*	5 CPI-U: ALL ITEMS LESS FOOD (SA)
97.	PUXHS*	5 CPI-U: ALL ITEMS LESS SHELTER (SA)
98.	PUXM*	5 CPI-U: ALL ITEMS LESS MIDICAL CARE (SA)
	Average hourly earnings	
99.	LEHCC*	5 AVG HR EARNINGS OF CONSTR WKRS: CONSTRUCTION
100.	LEHM*	5 AVG HR EARNINGS OF PROD WKRS: MANUFACTURING

C.2 Probabilistic Property: Strictly Stationary and α-mixing

Denote matrix $\mathbf{\Phi}$ as the same way as $\mathbf{\Phi}(Z_t)$ in (3.2). To show strictly stationary and α-mixing of process $\{\mathbb{P}_t\}$ in (3.2), the following assumptions are needed.

Assumption C.

C1: *Let* $\{\mathbb{P}_t\}$ *in* (3.2) *be a* ϕ*-irreducible and aperiodic Markov chain. For all* $1 \leq \iota \leq Q$ *and*

$1 \le \ell \le Q$, $\gamma_{k\iota\ell,P}(\cdot)$ in (3.2) is bounded such that $|\gamma_{k\iota\ell,P}(\cdot)| \le \gamma_{k\iota\ell,P}$ and the density function of $\varepsilon_{\ell,t}$ in (3.2) is positive every where on the real line $\mathbb{R}$ for all $1 \le \ell \le Q$. Furthermore, the roots of $I_Q - \Gamma_1 L - \cdots - \Gamma_q L^q = 0_{Q\times Q}$ all lie outside the unit circle.

C2: *Let $\boldsymbol{\varepsilon}_t$ in (3.3) be an i.i.d. process with $\boldsymbol{\varepsilon}_t = \Omega^{1/2}\boldsymbol{\eta}_t$, where $E(\boldsymbol{\eta}_t) = 0$, $\mathrm{var}(\boldsymbol{\eta}_t) = I_Q$, and $\Omega > 0$, $E(\|\boldsymbol{\eta}_t\|^4) < \infty$ and the elements of $\boldsymbol{\eta}_t$ are mutually independent.*

Assumption C makes the regularity conditions on P_t. It guarantees that P_t is strictly stationary and α-mixing, which is similar to that in Chen and Tsay (1993) and Cai et al. (2000). Finally, the following theorem is presented without proof, which might be derived in a similar way as in Cai and Liu (2022).

Theorem C.1. *Under Assumption C, if $\mathbb{P}_0$ is initialized from the invariant measure, then, $\{\mathbb{P}_t\}$ defined in (3.2) is a strictly stationary and α-mixing process.*