LOCAL SPATIAL QUANTILE ESTIMATION OF MULTIVARIATE FUNCTIONAL-COEFFICIENT REGRESSION MODELS

by

Yetong Zhou

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2020

Approved by:

Dr. Jiancheng Jiang

Dr. Eliana Christou

Dr. Weihua Zhou

Dr. Hwan C. Lin

ABSTRACT

YETONG ZHOU. Local Spatial Quantile Estimation of Multivariate Functional-coefficient Regression Models. (Under the direction of DR. JIANCHENG JIANG)

Quantile regression (QR) has been widely studied in statistics and econometrics. However, there is no much work on nonlinear QR for vector time series. Therefore, we propose a local spatial QR method to estimate the functional-coefficient matrices of multivariate time series. We propose a local spatial quantile regression estimator (LSQR) using spatial QR and local smoothing. To improve the performance, we propose a weighted composite LSQR estimator (WCLSQR) which uses the idea of weighted composite QR. We establish the asymptotic normality of the proposed estimators, which is further used to select an optimal bandwidth and optimal weights for the estimation. Furthermore, to achieve computational efficiency, we propose a smoothed spatial QR which simplifies and accelerates the minimization problem in the spatial QR. Based on the smoothed spatial QR, we propose the smoothed LSQR and WCLSQR estimators using the same techniques as LSQR and WCLSQR. By establishing the asymptotic normality of the proposed estimators, we show that the estimators using the smoothed spatial QR can achieve comparable performance with a proper choice of the smoothing parameter while consuming much less computing resources. Simulation study of the proposed estimators demonstrates good finite sample performance and computational efficiency. Real-world applications are also demonstrated.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my advisor Dr. Jiancheng Jiang, who has offered patient guidance and support consistently through my graduate life. The completion of my dissertation would not have been possible without his constant help. It was a great fortune to learn from his extensive knowledge and tireless commitment to research, which inspired and motivated me a lot during my research and will benefit me in my future career.

I would also like to extend my gratitude to to the rest of my committee members, Dr. Eliana Christou, Dr. Weihua Zhou, and Dr. Hwan Lin, for their insightful comments and valuable advice on my research work.

Thanks also to Dr. Shaozhong Deng and Dr. Mohammad A. Kazemi for their guidance and help as graduate coordinators.

Last but not least, I'm deeply grateful to my parents, Xilin Du and Liya Zhou for making me who I am and for their selfless love and support. Special thanks to Debbie, who brought me great joy and accompanied me throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Quantile regression (QR) was introduced by Koenker and Bassett (1978) and has become an important statistical tool for estimation and inference. Compared to mean regression, QR portrays the stochastic relationship between random variables better and with more accuracy than mean regression (Chaudhuri, Doksum, Samariv, 1997; Koenker, 2005) and provides more robust and consequently more efficient estimates than mean regression when the error is non-normal (Koenker and Bassett, 1978; Koenker and Zhao, 1996). These advantages have stimulated a tremendous amount of works on QR in statistics and econometrics.

There exists a rich literature on QR in the analysis of time series, examples include, but are not limited to Koul and Saleh (1995), Davis and Dunsmuir (1997), Jiang, Zhao and Hui (2001), Peng and Yao (2003). However, most of these works are devoted to univariate cases. For nonlinear vector time series, the maximum likelihood or least squares estimation have been extensively studied, see, for example, Bollerslev (1990), Engle and Kroner (1995), Chen and Tsay (1993), Pan and Yao (2008), and references therein. However, there is no much work on nonlinear vector time series for QR.

The first major difficulty in the analysis of vector time series with QR is the definition of a multivariate quantile. However, a solution is given by Chaudhuri (1996) and Koltchinskii (1997), who proposed a compelling form of multivariate quantiles based on the L_1 norm, namely the spatial quantiles. The spatial quantile provides an appealing multivariate extension of univariate quantiles and generates a useful volume functional based on spatial central regions of increasing size. As stressed in Serfling (2004), it has some appealing features: the equivariance and outlyingness with respect to shift and the orthogonal and homogeneous scale transformations. This gives us a tool and motivates us to work on the spatial QR for vector time series data.

For vector time series data, one should generally use multivariate models. Although univariate models for each time series may be employed, they are not able to capture the relationship among different time series and may not be efficient. Since nonlinear features widely exist in each time series (Tong and Lim, 1980; Tong, 1983; Chen and Tsay, 1993; Yao and Tong, 1995; Tsay, 1998; Fan and Yao, 2003), it is important to model the nonlinear features using non-parametric vector time series models, which requires little prior information on the model structure and may provide an insight into further parametric fitting. However, a fully non-parametric method suffers from the "curse of dimensionality" in multivariate cases when the dimension is high. In this dissertation, we consider the multivariate functional-coefficient model (1.1) proposed by Jiang (2014), for modeling nonlinear vector time series data. This model allows us to apply non-parametric techniques to explore the nonlinear effect and avoids the "curse of dimensionality". The model is given by

$$\mathbf{y}_t = \mathbf{c}(z_{t-d}) + \sum_{i=1}^{p} \boldsymbol{\phi}_i(z_{t-d})\mathbf{y}_{t-i} + \sum_{j=1}^{q} \boldsymbol{\beta}_j(z_{t-d})x_{t-j} + \boldsymbol{\epsilon}_t, \tag{1.1}$$

where $\mathbf{y}_t = (y_{1t}, \cdots, y_{kt})'$ is a k-dimensional time series, $\mathbf{x}_t = (x_{1t}, \cdots, x_{vt})'$ is v-dimensional exogenous variable, z_t is the threshold variable, $\mathbf{c}(\cdot)$ is $k \times 1$ functional vector, $\boldsymbol{\phi}_i(\cdot)$'s are $k \times k$ functional matrices, and $\boldsymbol{\beta}_i(\cdot)$'s are $k \times v$ functional matrices, for $i = 1, \ldots, p$. The innovations $\boldsymbol{\epsilon}_t = \boldsymbol{\sigma}(z_{t-d})\mathbf{u}_t$ and $\mathbf{u}_t$ is a sequence of serially uncorrelated random vectors with possible infinite variance and unknown distribution. The threshold variable z_t is assumed to be stationary and has a continuous distribution.

Model (1.1) is a generalization of the threshold model in Tsay (1998) and functional coefficient models in Chen and Tsay (1993), Hastie and Tibshirani (1993), Fan and Zhang (1999), Cai, Fan and Yao (2000), and Huang and Shen (2004). Without specifying the error distribution, model (1.1) can be estimated by local least squares method, see Jiang (2014). In this dissertation we focus on quantile estimation of the model, which is robust and efficient. In particular, when there is no exogenous variable ($q = 0$), we allow for infinite variance of the innovation $\mathbf{u_t}$. The proposed local spatial QR estimators admit no close form, so it is challenging to establish asymptotic properties for the proposed methodology. Besides, due to the non-differentiability of the spatial quantile loss function and the complexity of our model, minimization in the spatial QR is expensive and difficult. We further propose a smoothed spatial QR which simplifies and accelerates the computation. Efforts have been made to solve the associated difficulties.

The remainder of this dissertation is organized as follows. In Chapter 2, we introduce the proposed local spatial QR (LSQR) and its weighted composite version. Then, we establish the asymptotic normality of the proposed estimators, where optimal weights and bandwidth selection are considered based on theoretical results. In Chapter 3, we propose the smoothed spatial QR. By establishing the sampling properties of the proposed method, we show that the estimators using the smoothed spatial QR can achieve comparable performance. The choice of optimal parameters are also discussed. In Chapter 4, we conduct simulations to evaluate the performance of the proposed methodology. In Chapter 5, real-world applications are demonstrated to illustrate the application of our proposed estimation procedure. Concluding remarks are presented in Chapter 6. Proofs of the main results are given in the appendix.

CHAPTER 2: LOCAL SPATIAL QUANTILE REGRESSION (LSQR)

In this chapter, we introduce the local spatial quantile estimation for model (1.1). To do so, we first review the spatial quantile and then extend it using the idea of local linear smoothing.

2.1 Review of Spatial QR

Since the seminal work of Koenker and Bassett (1978), the univariate QR has been a very useful tool in statistics and econometrics. For multivariate models, a variety of ad hoc notions of multivariate quantiles have been formulated, but there is no definitive multivariate generalization. Here, we focus on the spatial quantiles, introduced by Chaudhuri (1996) and Koltchinskii (1997) which pose several appealing features.

According to Chaudhuri (1996) and Koltchinskii (1997), given a sample $\{\mathbf{z}_i\}_{i=1}^n$ of a random vector $\mathbf{z}$ in R^k, the $\mathbf{u}$-th spatial quantiles are defined as

$$\hat{\boldsymbol{\alpha}}(\mathbf{u}) = arg\ \min_{\boldsymbol{\alpha}\in R^k} \sum_{i=1}^{n}\{||\mathbf{z}_i - \boldsymbol{\alpha}|| + \mathbf{u}^T(\mathbf{z}_i - \boldsymbol{\alpha})\}, \tag{2.1}$$

where $\mathbf{u} \in \mathcal{B}^k = \{\mathbf{u}|\ ||\mathbf{u}|| < 1, \mathbf{u} \in R^k\}$ and $||\cdot||$ is the Euclidean norm. For $k = 1$, the solution to (2.1) reduces to the sample τ-th quantile ($\tau = (1+u)/2$) based on the real-valued observations z_i's. Let $Q_u(\mathbf{t}) = ||\mathbf{t}|| + < \mathbf{u}, \mathbf{t} >$, where $< \cdot, \cdot >$ is the Euclidean inner product. Then, (2.1) can be rewritten as

$$\hat{\boldsymbol{\alpha}}(\mathbf{u}) = arg\ \min_{\boldsymbol{\alpha}\in R^k} \sum_{i=1}^{n} Q_u(\mathbf{z}_i - \boldsymbol{\alpha}).$$

Define the $\mathbf{u}$-th quantile of the distribution of $\mathbf{z}$ as

$$\boldsymbol{\alpha}(\mathbf{u}) = arg \min_{\boldsymbol{\alpha}\in R^k} E[Q_u(\mathbf{z}-\boldsymbol{\alpha}) - Q_u(\mathbf{z})].$$

Chaudhuri (1996) showed that $\sqrt{n}(\hat{\boldsymbol{\alpha}}(\mathbf{u}) - \boldsymbol{\alpha}(\mathbf{u}))$ is asymptotically normal with mean 0.

Given the estimate $\hat{\boldsymbol{\alpha}}(\cdot)$, one can develop some multivariate descriptive statistics. For example, we can estimate the multivariate mean of $\mathbf{z}$ by the trimmed mean $\int_S \hat{\boldsymbol{\alpha}}(\mathbf{u})\ \mu(d\mathbf{u})$, where $\mu(\cdot)$ is an appropriate chosen probability measure on unit ball $\mathcal{B}^k$ and $S = \{\mathbf{u}|\mathbf{u} \in R^k, ||\mathbf{u}|| \le r\}$ for $r \in (0,1)$. The idea can be extended to multivariate regression settings. Consider the multivariate linear model,

$$\mathbf{y}_i = \boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, n, \tag{2.2}$$

where $\mathbf{y}_i$ is a $k \times 1$ vector, $\boldsymbol{\beta}$ is a $k \times v$ matrix of unknown parameters, and $\mathbf{x}_i$ is a $v \times 1$ vector of covariates without the intercept. It is straightforward to extend the above spatial quantile notion by defining the $\mathbf{u}$-th spatial regression quantiles as

$$(\hat{\boldsymbol{\beta}}(\mathbf{u}), \hat{\boldsymbol{\epsilon}}_u) = argmin_{\boldsymbol{\beta},\boldsymbol{\epsilon}} \sum_{i=1}^{n} Q_u(\mathbf{y}_i - \boldsymbol{\beta}\mathbf{x}_i - \boldsymbol{\epsilon}_u), \tag{2.3}$$

where $\boldsymbol{\epsilon}_u$ is the $\mathbf{u}$-th quantile of $\boldsymbol{\epsilon}$. Then, for any $\mathbf{u} \in \mathcal{B}^k$, $\hat{\boldsymbol{\beta}}(\mathbf{u})$ is a consistent estimate for $\boldsymbol{\beta}$. When $\mathbf{u} = 0$, it reduces to the spatial median regression in Bai, Chen, Miao and Rao (1990). For the univariate response case with $k = 1$, the above spatial QR is equivalent to the one introduced by Koenker and Bassett (1978). By using a transformation retransformation procedure as in Chaudhuri (1996), Chakraborty and Chaudhuri (1998), and Chakraborty (2003), an affine equivariant spatial QR estimation can be constructed.

2.2 Local Spatial QR

Let $\mathbf{X}_t^* = vec(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}, x_{t-1}, \ldots, x_{t-q})$, an $m \times 1$ vector with $m = pk + qv$, $\mathbf{X}_t = (1, \mathbf{X}_t^{*T})^T$, $\boldsymbol{\Phi}_2(z) = (\boldsymbol{\Phi}_1(z), \ldots, \boldsymbol{\Phi}_p(z), \boldsymbol{\beta}_1(z), \ldots, \boldsymbol{\beta}_q(z))$, and $\boldsymbol{\Phi}^*(z) = (\mathbf{c}(z), \boldsymbol{\Phi}_2(z))$. Then, model (1.1) can be written as

$$\mathbf{y}_t = \mathbf{\Phi}^*(z_{t-d})\mathbf{X}_t + \boldsymbol{\epsilon}_t, \tag{2.4}$$

where $\mathbf{\Phi}^*(\cdot)$ is a $k\times(m+1)$ matrix-valued function. Given z_{t-d}, we define the $\mathbf{u}$-th conditional quantile of $\boldsymbol{\epsilon}_t$ as

$$\mathbf{q}_u(z_{t-d}) = arg\ \min_{\mathbf{q}} E[Q_u(\boldsymbol{\epsilon}_t - \mathbf{q}) - Q_u(\boldsymbol{\epsilon}_t)|z_{t-d}].$$

Denote $\boldsymbol{\psi}_u(\mathbf{y}) = \partial Q_u(\mathbf{y})/\partial \mathbf{y} = \mathbf{y}/||\mathbf{y}|| + \mathbf{u}$ for $\mathbf{y} \neq 0$, then

$$E[\boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))|z_{t-d}] = 0.$$

As the first entry of $\mathbf{X}_t$ is 1, model (2.4) is equivalent to

$$\mathbf{y}_t = \mathbf{\Phi}(z_{t-d};\mathbf{u})(z_{t-d})\mathbf{X}_t + [\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})], \tag{2.5}$$

where $\mathbf{\Phi}(z;\mathbf{u}) = \mathbf{\Phi}^*(z) + (\mathbf{q}_u(z), 0, \ldots, 0)$. Also, since the conditional quantile $\mathbf{q}_u(z_{t-d})$ is not specified, the first column of $\mathbf{\Phi}^*(z)$, namely $\mathbf{z}(z)$, is not identifiable.

For any z_{t-d} in the neighborhood of z_0, using the Taylor expansion, we have

$$\mathbf{\Phi}(z_{t-d};\mathbf{u}) \approx \mathbf{\Phi}(z_0;\mathbf{u}) + \mathbf{\Phi}'(z_0;\mathbf{u})(z_{t-d} - z_0) \equiv \mathbf{A} + \mathbf{B}(z_{t-d} - z_0),$$

where $\mathbf{\Phi}'(z;\mathbf{u}) = \partial\mathbf{\Phi}(z;\mathbf{u})/\partial z$. In view of (2.3), applying the local linear approximation and the spatial QR for model (2.5), we minimize

$$\sum_{t=s'+1}^{n} Q_u(\mathbf{y}_t - [\mathbf{A} + \mathbf{B}(z_{t-d} - z_0)]\mathbf{X}_t)K(\frac{z_{t-d} - z_0}{h}) \tag{2.6}$$

over $\mathbf{A}$ and $\mathbf{B}$, where $s' = max(p, q)$. Let the resulting minimizers be $(\hat{\mathbf{A}}, \hat{\mathbf{B}})$. Then, $(\mathbf{\Phi}(z_0;\mathbf{u}), \mathbf{\Phi}'(z_0;\mathbf{u}))$ is estimated by $(\hat{\mathbf{A}}, \hat{\mathbf{B}})$, which is also denoted by $(\hat{\mathbf{\Phi}}(z_0;\mathbf{u}), \hat{\mathbf{\Phi}}'(z_0;\mathbf{u}))$ to emphasize dependence on $\mathbf{u}$ and z_0.

Let $\mathbf{\Phi}_1(z;\mathbf{u}) = \mathbf{c}(z) + \mathbf{q}_u(z)$. Then $\mathbf{\Phi}(z;\mathbf{u}) = [\mathbf{\Phi}_1(z;\mathbf{u}), \mathbf{\Phi}_2(z)]$. Partition $\hat{\mathbf{\Phi}}(z_0;\mathbf{u})$ into $[\hat{\mathbf{\Phi}}_1(z_0;\mathbf{u}), \hat{\mathbf{\Phi}}_2(z_0;\mathbf{u})]$, where $\hat{\mathbf{\Phi}}_1(z_0;\mathbf{u})$ is the first column of $\hat{\mathbf{\Phi}}(z_0;\mathbf{u})$. Then $[\hat{\mathbf{\Phi}}_1(z_0;\mathbf{u}), \hat{\mathbf{\Phi}}_2(z_0;\mathbf{u})]$ are the estimators of $[\mathbf{\Phi}_1(z_0;\mathbf{u}), \mathbf{\Phi}_2(z_0)]$, respectively.

2.3 Weighted Composite Local Spatial QR

Weighted composite quantile regression (WCQR) was initially studied by Koenker (1984) for classical linear models, and was later extended by Zou and Yuan (2008), Bradic, Fan and Jiang (2011), and Jiang, Jiang and Song (2012), using the penalized WCQR for model selection in the context of univariate parametric models. Here, we extend WCQR using the idea of local smoothing for the multivariate functional-coefficient model.

Consider J different quantiles, $\mathbf{u}_j \in \mathcal{B}^k,\ j = 1, \ldots, J$. For each $\mathbf{u}_j$ and any z_{t-d} in the neighborhood of z_0, we have

$$\mathbf{\Phi}_1(z_{t-d};\mathbf{u}_j) \approx \mathbf{\Phi}_1(z0;\mathbf{u}_j) + \mathbf{\Phi}_1'(z0;\mathbf{u}_j)(z_{t-d} - z_0) \equiv \mathbf{c}_{u_j} + \mathbf{d}_{u_j}(z_{t-d} - z_0)$$

$$\mathbf{\Phi}_2(z_{t-d}) \approx \mathbf{\Phi}_2(z_0) + \mathbf{\Phi}_2'(z_0)(z_{t-d} - z_0) \equiv \mathbf{A}_2 + \mathbf{B}_2(z_{t-d} - z_0)$$

Then, by equation (2.5),

$$\begin{aligned}\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}) &= \mathbf{y}_t - \boldsymbol{\phi}(z_{t-d})\mathbf{X}_t \\ &\approx \mathbf{y}_t - [\mathbf{c}_{u_j} + \mathbf{d}_{u_j}(z_{t-d} - z_0)] - [\mathbf{A}_2 + \mathbf{B}_2(z_{t-d} - z_0)]\mathbf{X}_t^*\end{aligned} \tag{2.7}$$

For simplicity of exposition, we introduce new notations $\boldsymbol{\theta} = [\boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{1J}, \boldsymbol{\theta}_2]$, where $\boldsymbol{\theta}_{1j} = [\mathbf{c}_{u_j}, h\mathbf{d}_{u_j}]$, and $\boldsymbol{\theta}_2 = [\mathbf{A}_2, h\mathbf{B}_2]$. Let $\mathbf{W}_{1t,h} = [1, h^{-1}(z_{t-d} - z_0)]^T$, and $\mathbf{W}_{2t,h} = \mathbf{W}_{1t,h} \otimes \mathbf{X}_t^*$, where $\otimes$ denotes the Kronecker product. It follows that

$$\begin{aligned}&\mathbf{y}_t - [\mathbf{c}_{u_j} + \mathbf{d}_{u_j}(z_{t-d} - z_0)] - [\mathbf{A}_2 + \mathbf{B}_2(z_{t-d} - z_0)]\mathbf{X}_t^* \\ &= \mathbf{y}_t - \boldsymbol{\theta}_{1j}\mathbf{W}_{1t,h} - \boldsymbol{\theta}_2\mathbf{W}_{2t,h} \\ &= \mathbf{y}_t - \boldsymbol{\theta}[\mathbf{e}_j \otimes \mathbf{W}_{1t,h}, \mathbf{W}_{2t,h}]^T,\end{aligned} \tag{2.8}$$

where $\mathbf{e}_j$ is a $J \times 1$ vector with the jth component as 1 and the rest as 0.

Denote (2.8) as $\boldsymbol{\xi}_t^j(\boldsymbol{\theta})$. Using the idea of WCQR, the weighted composite local spatial QR (WCLSQR) estimator $\hat{\boldsymbol{\theta}}$ can be obtained by minimizing

$$L_n(\boldsymbol{\theta};\boldsymbol{\omega}) \equiv \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} Q_{u_j}(\boldsymbol{\xi}_t^j(\boldsymbol{\theta}))K(\frac{z_{t-d}-z_0}{h}) \tag{2.9}$$

over $\boldsymbol{\theta}$, where $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_J)^T$ is a vector of positive weights. Denote the resulting minimizer by $\hat{\boldsymbol{\theta}} = [\hat{\mathbf{c}}_{u_1}, h\hat{\mathbf{d}}_{u_1}, \ldots, \hat{\mathbf{A}}_2, h\hat{\mathbf{B}}_2]$. Then $[\hat{\mathbf{A}}_2, h\hat{\mathbf{B}}_2]$ estimates $[\boldsymbol{\Phi_2}'(z_0), \boldsymbol{\Phi_2}(z_0)]$.

2.4 Sampling Properties

In this section, we establish asymptotic properties of the proposed estimators. Since the resulting estimators admit no close form, asymptotic normality is challenging to obtain. We will establish Bahadur's representation of the proposed estimators, which then lead to their asymptotic normality.

To facilitate presentations, we relegate conditions, supplementary lemmas and proofs of theorems to the appendices. To present the theorems, the following notations are needed.

Let $\hat{\boldsymbol{\gamma}} = \sqrt{nh}[\hat{\mathbf{A}} - \boldsymbol{\Phi}(z_0;\mathbf{u}), h(\hat{\mathbf{B}} - \boldsymbol{\Phi}'(z_0;\mathbf{u})]$ and $\hat{\boldsymbol{\zeta}}_2 = \sqrt{nh}[\hat{\mathbf{A}}_2 - \boldsymbol{\Phi}(z_0), h(\hat{\mathbf{B}}_2 - \boldsymbol{\Phi}'(z_0)]$. For $i = 0, 1, 2$, let $\mu_i = \int u^i K(u)\, du$ and $\nu_i = \int u^i K^2(u)\, du$. Also,

$$\mathbf{s} = \begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{bmatrix}, \mathbf{V} = \begin{bmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{bmatrix},$$

$\mathbf{M}(z_0) = E[\mathbf{X}_t\mathbf{X}_t^T|z_{t-d} = z_0]$, $\mathbf{N}_u(z_0) = E[\boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\{\boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\}^T|z_{t-d} = z_0]$, $\mathbf{D}_u(z_0) = E[\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))|z_{t-d} = z_0]$, $\mathbf{W}_{t,h} = [1, h^{-1}(z_{t-d} - z_0)]^T \otimes \mathbf{X}_t$, where $\boldsymbol{\psi}_u(\mathbf{y}) = \partial Q_u(\mathbf{y})/\partial \mathbf{y}$, $\boldsymbol{\Psi}_u(\mathbf{y}) = \partial^2 Q_u(\mathbf{y})/\partial \mathbf{y}\partial \mathbf{y}^T$.

Theorem 2.1. *Suppose conditions (A1)-(A6) hold. If $nh \to \infty$, $h \to 0$ and $nh^5 = O(1)$ as*

$n \to \infty$, then we have the following Bahadur representation:

$$vec(\hat{\boldsymbol{\gamma}}) - \sqrt{nh}\mathbf{B}_n(z_0;\mathbf{u}) = f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1}\mathbf{Z}_n + o_p(1),$$

where $\mathbf{B}_n(z_0;\mathbf{u}) = \frac{1}{2}h^2(\mathbf{S}^{-1}\mathbf{s}) \otimes vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))$, and

$$\mathbf{Z}_n = \frac{1}{\sqrt{nh}}\sum_{t=s'+1}^{n}[\mathbf{W}_{t,h} \otimes \boldsymbol{\psi}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h}).$$

Theorem 2.1 is also useful for making inference, including hypothesis testing, but this is out of the scope of the current study. The next theorem follows from Theorem 2.1.

Theorem 2.2. *Suppose conditions in Theorem 2.1 hold. Then*

$$\sqrt{nh}\left\{\begin{pmatrix} vec(\hat{\mathbf{A}} - \boldsymbol{\Phi}(z_0;\mathbf{u})) \\ vec\{h(\hat{\mathbf{B}} - \boldsymbol{\Phi}'(z_0;\mathbf{u}))\}\end{pmatrix} - \mathbf{B}_n(z_0;\mathbf{u})\right\} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Omega}(z_0)),$$

where $\boldsymbol{\Omega}(z_0) = f^{-1}(z_0)(\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}) \otimes \mathbf{M}^{-1}(z_0) \otimes (\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0))$.

It is straightforward from Theorem 2.2 that

$$\sqrt{nh}[vec(\hat{\Phi}(z_0;\mathbf{u}) - \Phi(z_0;\mathbf{u})) - \mathbf{b}_n(z_0;\mathbf{u})] \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\sigma}^2(z_0;\mathbf{u})),$$

where $\mathbf{b}_n(z_0;\mathbf{u}) = \frac{1}{2}h^2\frac{\mu_2^2-\mu_1\mu_3}{\mu_0\mu_2-\mu_1^2}vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))$ and

$$\boldsymbol{\sigma}^2(z_0;\mathbf{u}) = f^{-1}(z_0)\frac{\mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2}{(\mu_0\mu_2-\mu_1^2)^2}\mathbf{M}^{-1}(z_0) \otimes [\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0)].$$

The asymptotic bias $\mathbf{b}_n(z_0;\mathbf{u})$ is the same as that of the least squares estimator for $\boldsymbol{\Phi}(z_0)$ (see Jiang 2014). This property is also shared by the univariate local M-estimation in Fan and Jiang (2000).

Remark. From Theorem 2.2, the asymptotic mean square error is

$$AMSE(vec(\hat{\mathbf{\Phi}}(z_0;\mathbf{u}))) = \frac{1}{4}h^4(\frac{\mu_2^2-\mu_1\mu_3}{\mu_0\mu_2-\mu_1^2})^2||vec(\mathbf{\Phi}''(z_0;\mathbf{u}))||^2 + \frac{1}{nh}tr(\boldsymbol{\sigma}^2(z_0;\mathbf{u})),$$

where $tr(\cdot)$ is the trace function. When $K(\cdot)$ is chosen as a symmetrical kernel, $\mu_0 = 1$, $\mu_1 = 0$ and $\nu_1 = 0$. It follows that

$$\begin{aligned} AMSE(vec(\hat{\mathbf{\Phi}}(z_0;\mathbf{u}))) = &\frac{1}{4}h^4\mu_2^2||vec(\mathbf{\Phi}''(z_0;\mathbf{u}))||^2 \\ &+ \frac{1}{nh}\nu_0 f^{-1}(z_0)tr(\mathbf{M}^{-1}(z_0)\otimes[\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0)]), \end{aligned} \quad (2.10)$$

and the pointwise optimal bandwidth minimizing (2.10) is given by

$$h_{opt} = n^{-1/5}\left\{\frac{\nu_0 tr(\mathbf{M}^{-1}(z_0)\otimes[\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0)])}{\mu_2^2||vec(\mathbf{\Phi}''(z_0;\mathbf{u}))||^2 f(z_0)}\right\}^{1/5}.$$

Using the above formula, we employ a bandwidth selection procedure by multi-fold cross-validation and the average mean squared error criterion introduced in Jiang (2014).

- *Step 1.* Choose two integers, the fold size m and the fold number Q, such that $n > mQ$. A common choice is $m = [0.1n],\ Q = 4$.

- *Step 2.* Divide the data $\{\mathbf{X}_t, \mathbf{y}_t, z_{t-d}\}_{t=1}^n$ into $Q+1$ subsets following the time order. The first subset has $n - mQ$ observations and each of the rest has m observations.

- *Step 3.* For each of the rest Q subsets, compute the mean squared error $AMS_q(h_n) = m^{-1}\sum_{t=n-qm+1}^{n-qm+m}||\mathbf{y}_t - \hat{\mathbf{\Phi}}_q(z_{t-d})X_t||^2$, where $\hat{\mathbf{\Phi}}_q$ is estimated using the sample $\{\mathbf{X}_t, \mathbf{y}_t, z_{t-d}\}_{t=1}^{n-qm}$ with the bandwidth $h_n\{n/(n-qm)\}^{1/5}$. Choose h_n to minimize $AMS(h_n) = Q^{-1}\sum_{q=1}^{Q} AMS_q(h_n)$.

Usually boundary points have larger bandwidth than the interior points. Therefore we allow a variable bandwidth $h_n(z)$ depending on z_{t-d}. Let h_n depend on the density of z_{t-d} through $h_n = c\{\hat{f}(z)\}^{-1/5}n^{-1/5}$ and minimize $AMS(h_n)$ over c, where $\hat{f}(\cdot)$ is the kernel density estimate of $f_z(\cdot)$, given as $\hat{f}(z) = (nh_1)^{-1}\sum_{t=1}^{n} K((z_t - z)/h_1)$. Here, we take K as the Gaussian density kernel and set $h_1 = 1.06 s_z n^{-1/5}$ by the rule of thumb, where s_z is the sample standard deviation of $\{z_{t-d}\}_{t=1}^{n}$.

In the remainder of this section, we establish the Bahadur representation of the WCLSQR estimator, which is used to derive the asymptotic normality of the estimator.

Theorem 2.3. *Suppose the conditions (A1)-(A6) hold. For positive weights $\{\omega_j\}_{j=1}^{J}$, there is a unique minimizer $\hat{\boldsymbol{\theta}}$ of $L_n(\boldsymbol{\theta};\boldsymbol{\omega})$. If $nh \to \infty$, $h \to 0$, and $nh^5 = O(1)$ as $n \to \infty$, then we have the following Bahadur representation:*

$$vec(\hat{\boldsymbol{\zeta}}_2) - \sqrt{nh}\mathbf{B}_{n2}(z_0) = f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{D}(z_0;\boldsymbol{\omega}))^{-1}\mathbf{Z}_n(\boldsymbol{\omega}) + o_p(1),$$

where $\mathbf{B}_{n2}(z_0) = \frac{1}{2}h^2(\mathbf{S}^{-1}\mathbf{s}) \otimes vec(\boldsymbol{\Phi}_2''(z_0))$, $\mathbf{M}^*(z_0) = var[\mathbf{X}_t^*|z_{t-d} = z_0]$, $\mathbf{D}(z_0;\boldsymbol{\omega}) = \sum_{j=1}^{J}\omega_j \mathbf{D}_{u_j}(z_0)$, $\mathbf{Z}_n(\boldsymbol{\omega}) = \sum_{j=1}^{J} w_j[\mathbf{Z}_{n2j} - (\mathbf{I}_2 \otimes \boldsymbol{\mu}^*(z_0) \otimes \mathbf{I}_k)\mathbf{Z}_{n1j}]$, *with* $\boldsymbol{\mu}^*(z_0) = E(\mathbf{X}_t^*|z_{t-d} = z_0)$ *and* $\mathbf{Z}_{nij} = \frac{1}{\sqrt{nh}}\sum_{t=s'+1}^{n}[\mathbf{W}_{it,h} \otimes \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h})$.

Theorem 2.4. *Suppose conditions in Theorem 2.3 hold. Then*

$$\sqrt{nh}\left\{\begin{pmatrix} vec(\hat{\mathbf{A}}_2 - \boldsymbol{\Phi}(z_0)) \\ vec\{h(\hat{\mathbf{B}}_2 - \boldsymbol{\Phi}'(z_0))\} \end{pmatrix} - \mathbf{B}_{n2}(z_0)\right\} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Omega}_2(z_0;\boldsymbol{\omega}))$$

where $\boldsymbol{\Omega}_2(z_0;\boldsymbol{\omega}) = f^{-1}(z_0)(\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}) \otimes \mathbf{M}^{*-1}(z_0) \otimes (\mathbf{D}^{-1}(z_0;\boldsymbol{\omega})\mathbf{N}(z_0;\boldsymbol{\omega})\mathbf{D}^{-1}(z_0;\boldsymbol{\omega}))$, $\mathbf{N}(z_0;\boldsymbol{\omega}) = \sum_{j,l=1}^{J} w_j w_l \mathbf{N}_{u_j,u_l}$ *with* $\mathbf{N}_{u_j,u_l} = E[\boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\{\boldsymbol{\psi}_{u_l}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_l}(z_{t-d}))\}^T|z_{t-d} = z_0]$.

For the WCLSQR estimator, the asymptotic bias does not depend on $\boldsymbol{\omega}$; the asymptotic variance depends on $\boldsymbol{\omega}$ through $\boldsymbol{\Omega}_2(z_0;\boldsymbol{\omega})$ and it is invariant to the scale of $\boldsymbol{\omega}$. By minimizing the asymptotic variance, the optimal $\boldsymbol{\omega}$ is given by

$$\boldsymbol{\omega}_+ = arg\ \min_{\boldsymbol{\omega}}\ det\{\mathbf{D}^{-1}(z_0;\boldsymbol{\omega})\mathbf{N}(z_0;\boldsymbol{\omega})\mathbf{D}^{-1}(z_0;\boldsymbol{\omega})\}$$

subject to $\omega_j \geq 0$, $||\boldsymbol{\omega}|| = 1$. This optimization problem has no closed form solution, but it can be solved numerically.

CHAPTER 3: LSQR WITH SMOOTHED LOSS FUNCTIONS

3.1 Smooth Loss Functions for Spatial Quantiles

In this section, we introduce the smoothed spatial QR and then combine it with local linear smoothing to estimate model (1.1). It is a challenging task to run the spatial QR as it does not have a closed form solution and lacks differentiability at zero. The complexity of multivariate data adds to this difficulty. By smoothing the loss function of the spatial quantile, we propose the smoothed spatial QR which simplifies and accelerates the minimization problem in the spatial QR.

We first illustrate the idea with the simplest case. For univariant data $\{z_i\}_{i=1}^n$, the u-th quantile is defined as $\alpha(u) = arg \underset{\alpha}{min} \sum_{i=1}^n Q_u(z_i - \alpha)$, where $u \in [0,1]$ and

$$Q_u(t) = |t| + ut.$$

Noting that $|t|$ is non-differentiable at zero, we replace $|t|$ with a smooth function within a small neighborhood of zero, denoted as $[-\delta, \delta]$. Here smooth is defined as C^2-continuous, i.e. this function has continuous second derivatives. If the combined function is C^2-continuous at $\pm\delta$, then it is C^2-continuous for $t \in R$. With this way of smoothing, the resulting function, denoted as $Q_{u,\delta}(\cdot)$, is only different from $Q_u(\cdot)$ within $[-\delta, \delta]$, and δ is a controlling parameter. The choice of the function to substitute is flexible. As we naturally require the function to be symmetric and C^2-continuous δ, a convenient choice is $y(t) = at^4 + bt^2 + c$, where a, b and c are constants. Then the smoothed loss function $Q_{u,\delta}(\cdot)$ can be written explicitly as

$$Q_{u,\delta}(t) = \begin{cases} at^4 + bt^2 + c + ut, & t \in (-\delta, \delta) \\ Q_u(t), & otherwise. \end{cases}$$

Denote the first derivative of $Q(\cdot)$ as $\psi(\cdot)$ and the second derivative as $\Psi(\cdot)$. Given the condition that $Q_{u,\delta}(t)$ is C^2-continuous, a linear system of a, b and c can be constructed. After calculation, we obtain that $a = -1/(8\delta^3)$, $b = 3/(4\delta)$, $c = 3\delta/8$ respectively. Figure 3.1 shows the graphs of $Q_u(t)$, $Q_{u,\delta}(t)$ and their first and second derivatives on the same axes. With $\delta = 0.2$, it is seen that the check of $Q_u(t)$ at zero is smoothed within $[-0.2, 0.2]$; the jump of $\psi_u(t)$ and the non-existing point of $\Psi_u(t)$ disappear, which make the smoothed loss function $Q_{u,\delta}(t)$ C^2-continuous.

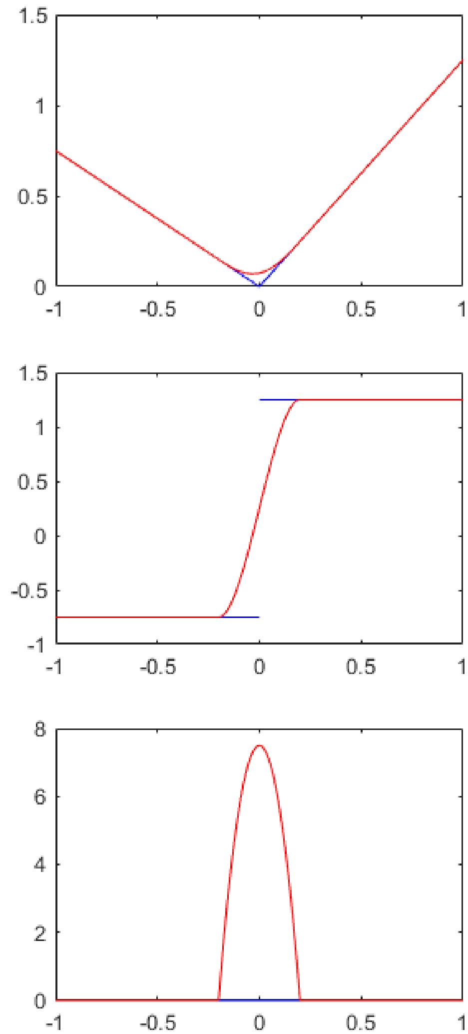

Figure 3.1: The three graphs show the original and smoothed spatial quantile loss function, their first and second derivatives, respectively, with the blue lines representing $Q_u(t)$, $\psi_u(t)$, $\Psi_u(t)$, and the red lines representing $Q_{u,\delta}(t)$, $\psi_{u,\delta}(t)$, $\Psi_{u,\delta}(t)$. Here $u = 0.25$, $\delta = 0.2$.

We extend this idea to the multivariate case. In the definition of the spatial quantile proposed by Chaudhuri (1996) and Koltchinskii (1997), the $\mathbf{u}$-th quantile of $\{\mathbf{z}_i\}_{i=1}^n$ in R^k is

$$\hat{\boldsymbol{\alpha}}(\mathbf{u}) = arg \min_{\boldsymbol{\alpha} \in R^k} \sum_{i=1}^{n} Q_u(\mathbf{z}_i - \boldsymbol{\alpha}),$$

where $Q_u(\mathbf{t}) = ||\mathbf{t}|| + < \mathbf{u}, \mathbf{t} >$ and $\mathbf{u} \in \mathcal{B}^k = \{\mathbf{u}|\mathbf{u} \in R^k, ||\mathbf{u}|| < 1\}$. With the same idea from univariate case, for $k > 2$, we look for a smooth function to replace $||\mathbf{t}||$ within $\mathcal{B}_\delta(\mathbf{0})$, where $\mathcal{B}_\delta(\mathbf{0})$ is a small ball centered at $\mathbf{0}$ with radius δ. As the isotropy and C^2-continuity are required, borrowing a similar form from the univariate case, we choose the substitute for $||\mathbf{t}||$ within $B_\delta(\mathbf{0})$ as the function in the form of $z(\mathbf{t}) = a(\mathbf{t}'\mathbf{t})^2 + b(\mathbf{t}'\mathbf{t}) + c$. The C^2-continuity of the resulting function is not that trivial as the univariate case. For the sake of simplicity, we introduce the following theorem and leave the detailed proof to Appendix C.

Theorem 3.1. *Define* $Q_{u,\delta}(\mathbf{t})$ *as*

$$Q_{u,\delta}(\mathbf{t}) = \begin{cases} a(\mathbf{t}'\mathbf{t})^2 + b(\mathbf{t}'\mathbf{t}) + c + \mathbf{u}'\mathbf{t}, & \mathbf{t} \in \mathcal{B}_\delta(\mathbf{0}) \\ Q_u(\mathbf{t}), & otherwise, \end{cases}$$

where $\mathbf{t} \in R^k$, $\mathbf{u} \in \mathcal{B}^k = \{\mathbf{u}|\mathbf{u} \in R^k, ||\mathbf{u}|| < 1\}$, δ *is a constant and* $\mathcal{B}_\delta(\mathbf{0}) = \{\mathbf{t}|\mathbf{t} \in R^k, ||\mathbf{t}|| < \delta\}$. *If* $a = -1/(8\delta^3)$, $b = 3/(4\delta)$, *and* $c = 3\delta/8$, *then* $Q_{u,\delta}(\mathbf{t})$ *is* C^2*-continuous for* $\mathbf{t} \in R^k$.

Denoting the gradient of $Q(\cdot)$ as $\boldsymbol{\psi}(\cdot)$, and the hessian matrix as $\boldsymbol{\Psi}(\cdot)$, Figure 3.2 shows the spatial quantile loss function $Q_u(\mathbf{t})$, the first component of its gradient $\boldsymbol{\psi}_u(\mathbf{t})$, the first entry of its hessian matrix $\boldsymbol{\Psi}_u(\mathbf{t})$ and their smoothed versions $Q_{u,\delta}(\mathbf{t})$, $\boldsymbol{\psi}_{u,\delta}(\mathbf{t})$, $\boldsymbol{\Psi}_{u,\delta}(\mathbf{t})$ for $k = 2$, $\mathbf{u} = [0, 0]'$ and $\delta = 0.25$. It is seen that in the three graphs on the right, the inverted cone at the bottom of $Q_u(\mathbf{t})$ is replaced by a smoothed surface within the sphere of radius 0.25; the jump along x-axis of $\boldsymbol{\psi}_u(\mathbf{t})$ and the non-existing point of $\boldsymbol{\Psi}_u(\mathbf{t})$ are also smoothed. With the smoothed loss function $Q_{u,\delta}(\cdot)$ at hand, we can derive the smoothed LSQR and WCLSQR estimators for model (2.5).

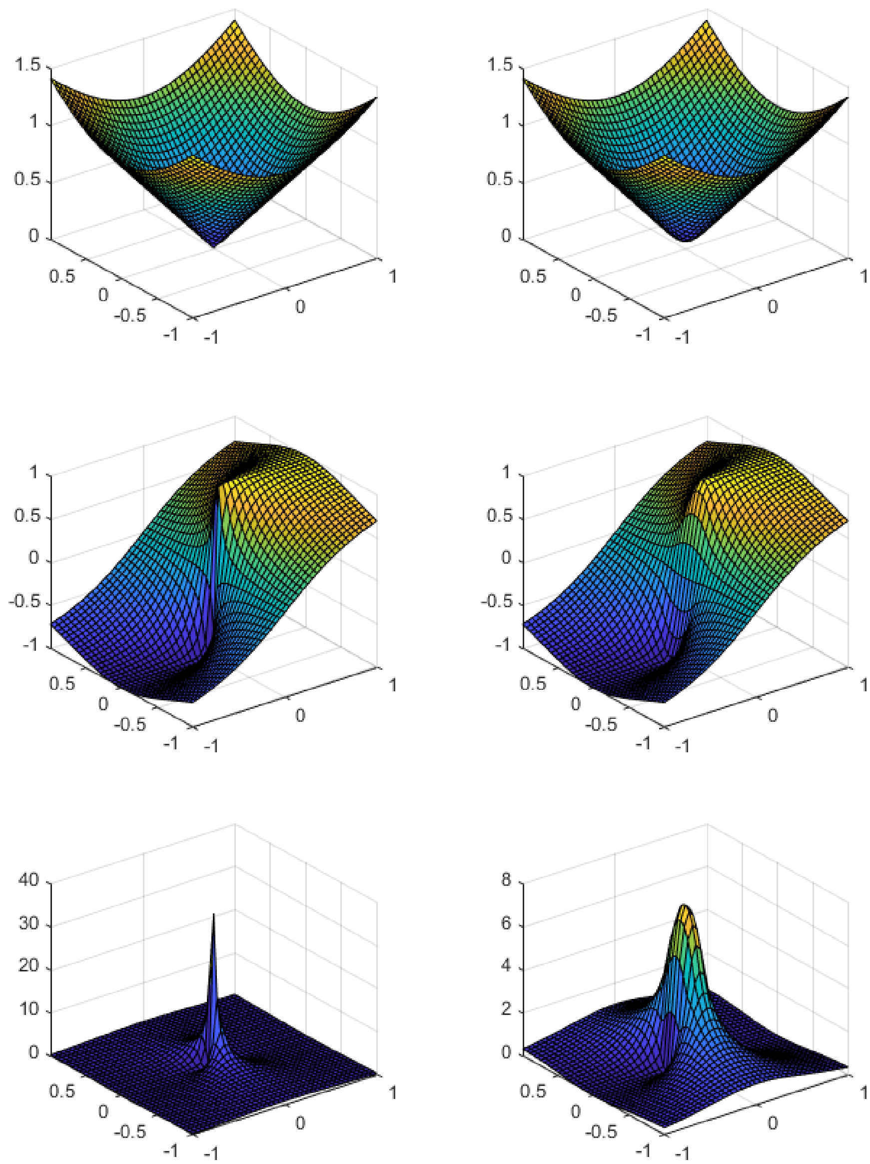

Figure 3.2: The left three graphs show the 2-D loss function $Q_u(\mathbf{t})$, the first component of its gradient $\boldsymbol{\psi}_u(\mathbf{t})$, and the first entry of its hessian matrix $\boldsymbol{\Psi}_u(\mathbf{t})$ respectively, while the right three graphs show their smoothed versions with $\delta = 0.25$.

3.2 Local Spatial Estimators with Smoothed Loss Functions

In this section, we derive the smoothed LSQR and WCLSQR estimators for model (2.5). Using the same notations from Chapter 2,

$$\mathbf{\Phi}(z_{t-d};\mathbf{u}) \approx \mathbf{\Phi}(z_0;\mathbf{u}) + \mathbf{\Phi}'(z_0;\mathbf{u})(z_{t-d}-z_0) \equiv \mathbf{A} + \mathbf{B}(z_{t-d}-z_0).$$

In view of (2.6), with the local linear approximation and the smoothed QR, we minimize

$$\sum_{t=s'+1}^{n} Q_{u,\delta}(\mathbf{y}_t - [\mathbf{A} + \mathbf{B}(z_{t-d}-z_0)]\mathbf{X}_t) K(\frac{z_{t-d}-z_0}{h}) \tag{3.1}$$

over $\mathbf{A}$ and $\mathbf{B}$. The resulting minimizers, denoted as $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ estimate $(\mathbf{\Phi}(z_0;\mathbf{u}), \mathbf{\Phi}'(z_0;\mathbf{u}))$. $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ is also denoted by $(\tilde{\mathbf{\Phi}}(z_0;\mathbf{u}), \tilde{\mathbf{\Phi}}'(z_0;\mathbf{u}))$ to emphasize dependence on $\mathbf{u}$ and z_0. Partition $\tilde{\mathbf{\Phi}}(z_0;\mathbf{u})$ into $[\tilde{\mathbf{\Phi}}_1(z_0;\mathbf{u}), \tilde{\mathbf{\Phi}}_2(z_0;\mathbf{u})]$, where $\tilde{\mathbf{\Phi}}_1(z_0;\mathbf{u})$ is the first column of $\tilde{\mathbf{\Phi}}(z_0;\mathbf{u})$. Then $[\tilde{\mathbf{\Phi}}_1(z_0;\mathbf{u}), \tilde{\mathbf{\Phi}}_2(z_0;\mathbf{u})]$ are the estimators of $[\mathbf{\Phi}_1(z_0;\mathbf{u}), \mathbf{\Phi}_2(z_0)]$, respectively. As will be shown in Theorem 3.3, $\tilde{\mathbf{\Phi}}(z_0;\mathbf{u})$ is a biased estimator of $\mathbf{\Phi}(z_0;\mathbf{u})$ while the bias can be controlled by the smoothing parameter δ.

Similarly, with $\boldsymbol{\theta}$ and $\boldsymbol{\xi}_t^j(\boldsymbol{\theta})$ defined in Chapter 2, the smoothed WCLSQR estimator is defined by minimizing

$$L_n(\boldsymbol{\theta};\boldsymbol{\omega}) \equiv \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} Q_{u_j,\delta}(\boldsymbol{\xi}_t^j(\boldsymbol{\theta})) K(\frac{z_{t-d}-z_0}{h}) \tag{3.2}$$

over $\boldsymbol{\theta}$, where $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_J)^T$ is a vector of positive weights. Denote the resulting minimizer by $\tilde{\boldsymbol{\theta}} = [\tilde{\mathbf{c}}_{u_1}, h\tilde{\mathbf{d}}_{u_1}, \ldots, \tilde{\mathbf{A}}_2, h\tilde{\mathbf{B}}_2]$. Then $[\tilde{\mathbf{A}}_2, h\tilde{\mathbf{B}}_2]$ estimates $[\mathbf{\Phi_2}'(z_0), \mathbf{\Phi_2}(z_0)]$. In the next section, we establish the asymptotic normality of the proposed estimators.

3.3 Sampling Properties

In this section, we will establish the Bahadur representations and the asymptotic normality of the proposed estimators. Detailed proofs of the following theorems are provided in

Appendix C.

The following additional notations are needed throughout the theorems and the proofs. Let $\tilde{\boldsymbol{\gamma}} = \sqrt{nh}[\tilde{\mathbf{A}} - \boldsymbol{\Phi}(z_0;\mathbf{u}), h(\tilde{\mathbf{B}} - \boldsymbol{\Phi}'(z_0;\mathbf{u})]$ and $\tilde{\boldsymbol{\zeta}} = \sqrt{nh}[\tilde{\mathbf{A}}_2 - \boldsymbol{\Phi}(z_0;\mathbf{u}), h(\tilde{\mathbf{B}}_2 - \boldsymbol{\Phi}'(z_0;\mathbf{u})]$. For $i = 0, 1, 2$, $\mu_i = \int u^i K(u)\, du$ and $\nu_i = \int u^i K^2(u)\, du$. Also, let $\mathbf{s}_0 = (\mu_0, \mu_1)^T$, $\boldsymbol{\mu}(z_0) = E[\mathbf{X}_t | z_{t-d} = z_0]$ and $\mathbf{n}_{u,\delta}(z_0) = E[\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) - \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) | z_{t-d} = z_0]$, where $\boldsymbol{\psi}_{u,\delta}(\mathbf{y}) = \partial Q_{u,\delta}(\mathbf{y})/\partial \mathbf{y}$, $\boldsymbol{\Psi}_{u,\delta}(\mathbf{y}) = \partial^2 Q_{u,\delta}(\mathbf{y})/\partial \mathbf{y} \partial \mathbf{y}^T$.

Theorem 3.2. *Suppose the conditions (A1) to (A6) holds. If $h \to 0$, $\delta \to 0$, $nh \to \infty$, $\delta^k nh \to \infty$, $nh^5 = O(1)$, and $\delta^{2k} nh = O(1)$ as $n \to \infty$, then we have the following Bahadur representation:*

$$vec(\tilde{\gamma}) - \sqrt{nh}\mathbf{B}_{n,\delta}(z_0;\mathbf{u}) = f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1}\mathbf{Z}_n + o_p(1),$$

where $\mathbf{B}_{n,\delta}(z_0;\mathbf{u}) = \mathbf{S}^{-1}\mathbf{s}_0 \otimes [\mathbf{M}(z_0)^{-1}\boldsymbol{\mu}(z_0) \otimes \mathbf{D}_u(z_0)^{-1}\mathbf{n}_{u,\delta}(z_0)](1 + o_p(1))$ $+\frac{h^2}{2}\mathbf{S}^{-1}\mathbf{s} \otimes vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))(1+o_p(1))$, $\mathbf{Z}_n(u) = \frac{1}{\sqrt{nh}}\sum_{t=s'+1}^{n}[\mathbf{W}_{t,h} \otimes \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h})$.

The above result is similar to the Bahadur representation of the LSQR estimator in Theorem 2.1 with an additional term for bias caused by the difference between the smoothed loss function and the original one. From Theorem 3.2, the asymptotic normality of the estimator is easy to get.

Theorem 3.3. *Suppose conditions in Theorem 3.2 hold. Then*

$$\sqrt{nh}\left\{\begin{pmatrix} vec(\tilde{\mathbf{A}} - \boldsymbol{\Phi}(z_0;\mathbf{u})) \\ vec(h(\tilde{\mathbf{B}} - \boldsymbol{\Phi}'(z_0;\mathbf{u}))) \end{pmatrix} - \mathbf{B}_{n,\delta}(z_0;\mathbf{u})\right\} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Omega}(z_0)),$$

where $\boldsymbol{\Omega}(z_0) = f^{-1}(z_0)(\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}) \otimes \mathbf{M}^{-1}(z_0) \otimes (\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0))$.

It is straightforward from Theorem 3.3 that

$$\sqrt{nh}[vec(\tilde{\boldsymbol{\Phi}}(z_0;\mathbf{u}) - \boldsymbol{\Phi}(z_0;\mathbf{u})) - \mathbf{b}_{n,\delta}(z_0;\mathbf{u})] \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\sigma}_2^2(z_0;\mathbf{u})),$$

where

$$\mathbf{b}_{n,\delta}(z_0;\mathbf{u}) = \mathbf{M}(z_0)^{-1}\boldsymbol{\mu}(z_0) \otimes \mathbf{D}_u(z_0)^{-1}\mathbf{n}_{u,\delta}(z_0) + \frac{1}{2}h^2\frac{\mu_2^2-\mu_1\mu_3}{\mu_0\mu_2-\mu_1^2}vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))$$

and

$$\boldsymbol{\sigma}_2^2(z_0;\mathbf{u}) = f^{-1}(z_0)\frac{\mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2}{(\mu_0\mu_2-\mu_1^2)^2}\mathbf{M}^{-1}(z_0) \otimes [\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0)].$$

Remark. Note that, as $\boldsymbol{\psi}_{u,\delta}(\cdot)$ is only different from $\boldsymbol{\psi}_u(\cdot)$ in a small ball centered at 0 with radius δ, then

$$\begin{aligned}||\mathbf{n}_{u,\delta}(z_0)||^2 &= ||E[\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) - \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))|z_{t-d} = z_0]||^2 \\ &\le c\delta^{2k}\ sup_{\mathbf{x}\in\mathcal{B}_\delta(0)}||\boldsymbol{\psi}_{u,\delta}(\mathbf{x}) - \boldsymbol{\psi}_u(\mathbf{x})||^2.\end{aligned}$$

Thus, $sup\,||\mathbf{n}_{u,\delta}(z_0)||^2 = c_\delta(z_0)\delta^{2k}$. From Theorem 3.3, we can calculate the supremum of the asymptotic mean square error

$$\begin{aligned}sup\,AMSE(vec(\tilde{A})) = 2c_\delta(z_0)\delta^{2k}||\mathbf{M}(z_0)^{-1}\boldsymbol{\mu}(z_0) \otimes \mathbf{D}_{u,\delta}(z_0)^{-1}\mathbf{j}_k||^2 \\ + \frac{1}{2}h^4(\frac{\mu_2^2-\mu_1\mu_3}{\mu_0\mu_2-\mu_1^2})^2||vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))||^2 + \frac{1}{nh}tr(\boldsymbol{\sigma}_2^2(z_0;\mathbf{u})),\end{aligned}$$

where $\mathbf{j}_k$ is a $k \times 1$ unit vector with the same direction as $\mathbf{n}_{u,\delta}(z_0)$ and $tr(\cdot)$ is the trace function. If $K(\cdot)$ is a symmetrical kernel, we have $\mu_0 = 1$, $\mu_1 = 0$ and $\nu_1 = 0$. Then

$$\begin{aligned}sup\,AMSE(vec(\tilde{A})) =&2c_\delta(z_0)\delta^{2k}||\mathbf{M}(z_0)^{-1}\boldsymbol{\mu}(z_0) \otimes \mathbf{D}_{u,\delta}(z_0)^{-1}\mathbf{j}_k||^2 \\ &+ \frac{1}{2}h^4\mu_2^2||vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))||^2 \\ &+ \frac{1}{nh}\nu_0 f^{-1}(z_0)tr(\mathbf{M}^{-1}(z_0) \otimes [\mathbf{D}_{u,\delta}^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_{u,\delta}^{-1}(z_0)]).\end{aligned} \tag{3.3}$$

To minimize (3.3), the optimal bandwidth is given as

$$h_{opt} = n^{-1/5} \left\{ \frac{\nu_0 tr(\mathbf{M}^{-1}(z_0) \otimes [\mathbf{D}_{u,\delta}^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_{u,\delta}^{-1}(z_0)])}{2\mu_2^2 ||vec(\mathbf{\Phi}''(z_0;\mathbf{u}))||^2 f(z_0)} \right\}^{1/5}.$$

The obtained optimal bandwidth has a similar form to the one derived in Section 2.4. Thus the bandwidth selection procedure introduced in Section 2.4 can be easily extended to smoothed LSQR. With $h = cn^{-1/5}$, one can choose δ that satisfies $\delta^k \ll n^{-2/5}$ to make the first term of (3.3) neglectable to the rest two terms. As $\delta^k nh \to \infty$ is required by Theorems 3.2 and 3.3, δ should also satisfy $\delta^k \gg n^{-4/5}$. With such choice of δ, the LSQR estimator with smoothed loss functions can achieve a comparable accuracy as LSQR.

Theorem 3.4. *Suppose the conditions (A1)-(A6) hold. If $h \to 0$, $\delta \to 0$, $nh \to \infty$, $\delta^k nh \to \infty$, $nh^5 = O(1)$, and $\delta^{2k} nh = O(1)$ as $n \to \infty$, then we have the following Bahadur representation:*

$$vec(\tilde{\boldsymbol{\zeta}}_2) - \sqrt{nh}\mathbf{B}_{n2}(z_0) = f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{D}(z_0;\boldsymbol{\omega}))^{-1}\mathbf{Z}_n(\boldsymbol{\omega}) + o_p(1),$$

where $\mathbf{B}_{n2}(z_0) = \frac{1}{2}h^2(\mathbf{S}^{-1}\mathbf{s}) \otimes vec(\mathbf{\Phi}_2''(z_0))$, $\mathbf{M}^*(z_0) = var[\mathbf{X}_t^*|z_{t-d} = z_0]$, $\mathbf{D}(z_0;\boldsymbol{\omega}) = \sum_{j=1}^J \omega_j \mathbf{D}_{u_j,\delta}(z_0)$, $\mathbf{Z}_n(\boldsymbol{\omega}) = \sum_{j=1}^J w_j[\mathbf{Z}_{n2j} - (\mathbf{I}_2 \otimes \boldsymbol{\mu}^*(z_0) \otimes \mathbf{I}_k)\mathbf{Z}_{n1j}]$, *with* $\boldsymbol{\mu}^*(z_0) = E(\mathbf{X}_t^*|z_{t-d} = z_0)$ *and* $\mathbf{Z}_{nij} = \frac{1}{\sqrt{nh}}\sum_{t=s'+1}^n [\mathbf{W}_{it,h} \otimes \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h})$.

Theorem 3.5. *Suppose conditions in Theorem 3.4 hold. Then*

$$\sqrt{nh}\left\{ \begin{pmatrix} vec(\tilde{\mathbf{A}}_2 - \mathbf{\Phi}(z_0)) \\ vec\{h(\tilde{\mathbf{B}}_2 - \mathbf{\Phi}'(z_0))\} \end{pmatrix} - \mathbf{B}_{n2}(z_0) \right\} \xrightarrow{d} \mathcal{N}(0, \mathbf{\Omega}_2(z_0;\boldsymbol{\omega})),$$

where $\mathbf{\Omega}_2(z_0;\boldsymbol{\omega}) = f^{-1}(z_0)(\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}) \otimes \mathbf{M}^{*-1}(z_0) \otimes (\mathbf{D}^{-1}(z_0;\boldsymbol{\omega})\mathbf{N}(z_0;\boldsymbol{\omega})\mathbf{D}^{-1}(z_0;\boldsymbol{\omega}))$, $\mathbf{N}(z_0;\boldsymbol{\omega}) = \sum_{j,l=1}^J w_j w_l \mathbf{N}_{u_j,u_l}$ *with* $\mathbf{N}_{u_j,u_l} = E[\boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\{\boldsymbol{\psi}_{u_l}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_l}(z_{t-d}))\}^T|z_{t-d} = z_0]$.

Since the coefficients of $\{\mathbf{y}_{t-i}\}$, $\{\mathbf{x}_{t-j}\}$ of model 1.1 are not related to quantiles, then $\tilde{\zeta}_2$

has the same asymptotic properties as $\hat{\zeta}_2$. Also, the method to select optimal weights for WCLSQR can be extended to smoothed WCLSQR without modifications.

CHAPTER 4: SIMULATIONS

In this section, we conduct numerical simulations to evaluate the performance of the proposed methodology. We consider the following two-dimensional EXPAR model:

$$\mathbf{y}_t = \mathbf{\Phi}_1(z_{t-1})\mathbf{y}_{t-1} + \mathbf{\Phi}_2(z_{t-1})\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t,$$

where $\mathbf{y_t} = (y_{t,1}, y_{t,2})'$, $\boldsymbol{\epsilon}_t = (\epsilon_{t,1}, \epsilon_{t,2})'$, z_t is a uniform process on $[0,1]$. The coefficient matrices are given by

$$\mathbf{\Phi}_k(z) = \begin{bmatrix} \Phi_{k,11}(z) & \Phi_{k,12}(z) \\ \Phi_{k,21}(z) & \Phi_{k,22}(z) \end{bmatrix}$$

for $k = 1, 2$, where

$$\Phi_{k,11}(z) = \Phi_{k,22}(z) = 0.01 + (0.3 + 0.9z)\exp(-3.9z^2)$$

and

$$\Phi_{k,12}(z) = \Phi_{k,21}(z) = -0.04 + (-0.7 + 4.3z)\exp(-6.9z^2).$$

For $\boldsymbol{\epsilon}_t$, We consider the following three types of errors:

(i) $\boldsymbol{\epsilon}_t$ follows a bivariate normal distribution with mean $\boldsymbol{\mu} = (0,0)'$ and covariance matrix $\Sigma = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}$.

(ii) $\boldsymbol{\epsilon}_t$ follows a bivariate $t-$distribution with 3 degrees of freedom and covariance matrix $\Sigma = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}$.

(iii) 95% of data points follow a a bivariate normal distribution with mean $\boldsymbol{\mu} = (0,0)'$ and covariance matrix $\Sigma = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$. The rest 5% follow normal distribution with mean $\boldsymbol{\mu} = (0,0)'$ and covariance matrix $\Sigma = \begin{bmatrix} 25.0 & 0.0 \\ 0.0 & 25.0 \end{bmatrix}$

Example 1. For comparison, we test the proposed LSQR estimator, WCLSQR estimator and the least squares local linear smoother (Jiang 2014) on three different types of innovations. For each type, we run 500 simulations of sample size 500.

The other settings for the simulations are as follows. For LSQR, $\mathbf{u}$ is set as $(0,0)'$, equivalent to the median for univariate data. For WCLSQR, seven quantiles $\mathbf{U} = [\mathbf{u}_i]_{i=1}^{7}$ are selected along a line as $[-0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75] \cdot [sin(\pi/3), cos(\pi/3)]'$. Optimal bandwidth $h(z)$ and weights $w(z)$ are selected locally at z using the procedure introduced in Section 2.4.

Tables 4.1-4.3 displays the bias and standard deviations of the estimator for $\Phi_{k,11}(z)$ at $z = 0.5$. As expected, for standard normal error, the least square estimator still has the smallest standard deviation, while for t-distribution and mixed normal distribution with outliers, both LSQR and WCLSQR estimators outperform the least squares estimator.

Table 4.1: Summary results of the least square, LSQR, WCLSQR estimates on data (i).

	bias	std
Least Square	-1.02E-02	7.38E-02
LSQR	-1.29E-02	8.63E-02
WCLSQR	-1.14E-02	7.85E-02

Table 4.2: Summary results of the least squares, LSQR, WCLSQR estimates on data (ii).

	bias	std
Least Square	-1.37E-03	7.01E-02
LSQR	2.91E-06	5.72E-02
WCLSQR	6.35E-04	5.69E-02

Table 4.3: Summary results of the least squares, LSQR, WCLSQR estimates on data (iii).

	bias	std
Least Square	-8.97E-03	6.09E-02
LSQR	-5.62E-03	5.74E-02
WCLSQR	-6.89E-03	5.42E-02

Example 2. We first compare the proposed smoothed LSQR to LSQR with the fixed bandwidth h and the fixed smooth parameter δ. We run 500 simulations of sample size $n = 1000$ on data (i). Here $\mathbf{u}$ is set as $(0,0)'$, h is fixed as $n^{-1/5} = 0.2512$ and $\delta = n^{-3/10} = 0.1259$ for illustration purpose. For $z \in [0,1]$, as boundary points usually require larger bandwidth, we only run the test on the interior points from 0.3 to 0.7.

Denote $(\hat{\Phi}_{11}, \hat{\Phi}_{12})$ as the LSQR estimators of (Φ_{11}, Φ_{12}), and $(\tilde{\Phi}_{11}, \tilde{\Phi}_{12})$ as the smoothed LSQR estimators of (Φ_{11}, Φ_{12}). Tables 4.4-4.6 display the bias and standard deviations of the two estimation methods and their difference. We observe that for $k = 2$, with δ satisfying $n^{-4/5} \ll \delta^k \ll n^{-2/5}$, the difference between the bias of the smoothed LSQR estimator and the LSQR estimator is rather small.

Table 4.4: Bias and standard deviation of smoothed LSQR estimates with $h = 0.2512$, $\delta = 0.1259$.

z_0	bias of $\tilde{\Phi}_{11}$	std of $\tilde{\Phi}_{11}$	bias of $\tilde{\Phi}_{12}$	std of $\tilde{\Phi}_{12}$
0.3	-1.89E-02	5.89E-02	-9.80E-02	5.40E-02
0.4	-1.22E-02	5.87E-02	-6.15E-02	5.43E-02
0.5	-4.50E-03	5.94E-02	-1.44E-02	5.80E-02
0.6	2.81E-03	5.82E-02	1.54E-02	5.84E-02
0.7	6.16E-03	5.52E-02	2.26E-02	5.69E-02

Table 4.5: Bias and standard deviation of LSQR estimates with $h = 0.2512$.

z_0	bias of $\hat{\Phi}_{11}$	std of $\hat{\Phi}_{11}$	bias of $\hat{\Phi}_{12}$	std of $\hat{\Phi}_{12}$
0.3	-1.89E-02	5.90E-02	-9.80E-02	5.43E-02
0.4	-1.22E-02	5.90E-02	-6.15E-02	5.44E-02
0.5	-4.55E-03	5.93E-02	-1.43E-02	5.82E-02
0.6	2.91E-03	5.83E-02	1.57E-02	5.87E-02
0.7	6.20E-03	5.53E-02	2.28E-02	5.71E-02

Table 4.6: Comparison of smoothed LSQR and LSQR estimates.

z_0	Δbias of Φ_{11}	Δstd of Φ_{11}	Δbias of Φ_{12}	Δstd of Φ_{12}
0.3	-3.50E-05	-1.13E-04	-4.18E-07	-2.86E-04
0.4	5.46E-06	-2.79E-04	-2.10E-05	-6.92E-05
0.5	-4.08E-05	6.90E-05	4.82E-05	-2.11E-04
0.6	-9.72E-05	-1.54E-04	-2.27E-04	-3.42E-04
0.7	-4.11E-05	-1.13E-04	-2.01E-04	-2.10E-04

We also compare the smoothed WCLSQR estimator and the WCLSQR estimator using the optimal bandwidth and weights selection procedure. Simulation of a sample size $n = 1000$

on data (i) is replicated 500 times. The spatial quantiles $\mathbf{U}$ are the same as in Example 1. As illustrated in Section 3.3, the choice of δ does not affect the optimal bandwidth selection. For $z_0 = 0, 0.1, \ldots, 1$, after the optimal bandwidth $h(z_0)$ is selected, we set $\delta = h(z_0)n^{-1/10}$. Denote $(\hat{\Phi}_{11}(\boldsymbol{\omega}), \hat{\Phi}_{12}(\boldsymbol{\omega}))$ as the WCLSQR estimators of (Φ_{11}, Φ_{12}), and $(\tilde{\Phi}_{11}(\boldsymbol{\omega}), \tilde{\Phi}_{12}(\boldsymbol{\omega}))$ as the smoothed WCLSQR estimators of (Φ_{11}, Φ_{12}). Tables 4.7-4.9 report the simulation results. It is seen that the smoothed WCLSQR achieves similar performance as the WCLSQR, while for 500 groups of simulation, the smoothed WCLSQR consumes 11 hours 50 minutes, much less than the WCLSQR, which consumes 22 hrs 09 minutes.

Table 4.7: Bias and standard deviation of smoothed WCLSQR estimates with optimal bandwidths and weights.

z_0	bias of $\tilde{\Phi}_{11}(\boldsymbol{\omega})$	std of $\tilde{\Phi}_{11}(\boldsymbol{\omega})$	bias of $\tilde{\Phi}_{12}(\boldsymbol{\omega})$	std of $\tilde{\Phi}_{12}(\boldsymbol{\omega})$
0	2.59E-02	1.91E-01	3.75E-02	1.84E-01
0.1	-5.15E-03	8.59E-02	-1.22E-02	8.50E-02
0.2	-8.45E-03	8.24E-02	-4.17E-02	8.75E-02
0.3	-4.39E-03	8.58E-02	-4.09E-02	8.51E-02
0.4	-1.52E-03	7.76E-02	-2.73E-02	7.91E-02
0.5	-3.53E-03	7.27E-02	-7.37E-03	7.24E-02
0.6	-2.35E-03	7.10E-02	9.40E-03	7.79E-02
0.7	6.95E-03	7.33E-02	2.32E-02	8.36E-02
0.8	8.64E-03	6.65E-02	1.90E-02	6.80E-02
0.9	2.89E-03	7.83E-02	8.18E-03	7.47E-02
1	-1.41E-02	1.82E-01	-1.35E-02	1.87E-01

Table 4.8: Bias and standard deviation of WCLSQR estimates with optimal bandwidths and weights.

z_0	bias of $\tilde{\Phi}_{11}(\boldsymbol{\omega})$	std of $\tilde{\Phi}_{11}(\boldsymbol{\omega})$	bias of $\tilde{\Phi}_{12}(\boldsymbol{\omega})$	std of $\tilde{\Phi}_{12}(\boldsymbol{\omega})$
0	2.29E-02	1.96E-01	3.27E-02	1.89E-01
0.1	-4.64E-03	8.50E-02	-1.23E-02	8.32E-02
0.2	-9.44E-03	8.02E-02	-3.94E-02	8.64E-02
0.3	-5.13E-03	8.18E-02	-4.24E-02	7.86E-02
0.4	-2.01E-03	7.67E-02	-2.58E-02	7.95E-02
0.5	-3.05E-03	7.28E-02	-8.21E-03	7.66E-02
0.6	-2.44E-03	7.04E-02	8.22E-03	7.73E-02
0.7	7.46E-03	7.51E-02	2.05E-02	8.19E-02
0.8	8.83E-03	6.87E-02	1.82E-02	7.02E-02
0.9	4.43E-03	8.04E-02	8.77E-03	7.72E-02
1	-1.62E-02	1.92E-01	-1.50E-02	1.87E-01

Table 4.9: Comparison of smoothed WCLSQR and WCLSQR estimates.

z_0	Δbias of Φ_{11}	Δstd of Φ_{11}	Δbias of Φ_{12}	Δstd of Φ_{12}
0	2.97E-03	-5.25E-03	4.78E-03	-4.42E-03
0.1	5.18E-04	8.17E-04	-1.35E-04	1.76E-03
0.2	-9.84E-04	2.22E-03	2.25E-03	1.16E-03
0.3	-7.39E-04	4.00E-03	-1.46E-03	6.50E-03
0.4	-4.97E-04	8.35E-04	1.48E-03	-3.73E-04
0.5	4.77E-04	-1.40E-04	-8.43E-04	-4.17E-03
0.6	-8.62E-05	6.02E-04	1.18E-03	6.54E-04
0.7	-5.08E-04	-1.79E-03	2.64E-03	1.62E-03
0.8	-1.93E-04	-2.22E-03	8.04E-04	-2.17E-03
0.9	-1.54E-03	-2.14E-03	-5.90E-04	-2.54E-03
1	-2.11E-03	-9.19E-03	-1.54E-03	5.06E-04

CHAPTER 5: REAL EXAMPLES

5.1 Iceland River Flows

In this section, we study the vector time series consisting of two daily river flow series of Iceland using the proposed methodology. The data was analyzed in Tong, Thanoon and Gudmundsson (1985) as two individual time series using the threshold model and Tsay (1998) using the threshold multivariate model. Modeling the dynamics of the river flows can be quite complicated as they involve many factors, such as evaporation, transpiration, underground sources, and melting snow. However, for use of simulation and prediction, it is worth exploring a relatively simpler model to identify the relationship between the river flow and some meteorological variables that are easy to acquire. Following Tong, Thanoon and Gudmundsson (1985) and Tsay (1998), we consider the exogenous variables as lagged values of daily precipitation, measured in millimeters (mm), denoted as $\{x_{t-j}\}_{j=1}^{q}$, and the lagged value of temperature, measure in degrees Celsius (°C), denoted as z_{t-d}. Two daily river flow series, measured in m^3s^{-1}, are from the Jökulsá Eystri River and Vatnsdalsá River from 1972 to 1974, denoted as $\mathbf{y}_t = (y_{1t}, y_{2t})'$. The data are downloaded from the R package: Time Series Data Library. In Figure 5.1, the time plot of two river flow series shows strong evidence of nonlinear features, such as sharp rises and slow declines. Thus, it is reasonable to employ a nonlinear model. Tsay (1998) selected the contemporaneous value of daily temperature z_t as the threshold variable. As an extension to this, we assume that the coefficients of lagged values of the river flows and exogenous variables depend on z_{t-d}, and we consider the following model

$$\mathbf{y}_t = \mathbf{c}(z_{t-d}) + \sum_{i=1}^{p} \boldsymbol{\phi}_i(z_{t-d})\mathbf{y}_{t-i} + \sum_{i=1}^{q} \boldsymbol{\beta}_i(z_{t-d})x_{t-i} + \boldsymbol{\epsilon}_t. \tag{5.1}$$

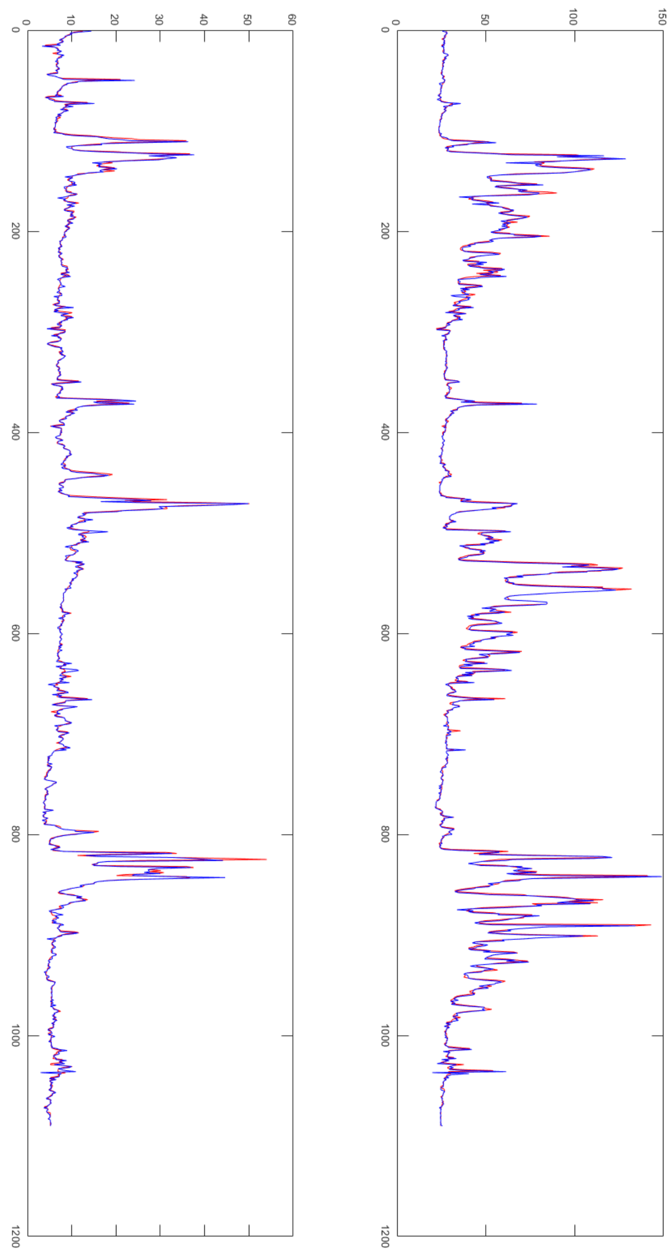

Figure 5.1: The right figure shows the time plot of the daily river flow from the Jökulsá Eystri River, with the red line representing actual values and blue lines representing its median estimates. The left figure shows the time plot of the daily river flow from Vatnsdalsá River.

Using the average prediction error criterion (APE), we select the lagged order $p = 5$, $q = 3$ and $d = 4$, although the results indicate that model (5.1) is not that sensitive to the selection of the lagged order. Tsay (1998) also found that the autoregressive order, ranging from AR(4) to AR(22) is not a significant factor. This is reasonable as we only have a very limited number of exogenous variables, and some factors affecting the river flow may be contained in the past observations. It is seen in Figure 5.1 that the median estimates well capture the non-linearity of the data, which show that the temperature, as an important role affecting the snow melting, can explain much of the non-linearity. Figures 5.2 - 5.3 show the median estimates of the functional coefficients of $x_{t-1}, x_{t-2}, x_{t-3}, z_{t-4}$ with respect to y_{1t} and y_{2t}. By modeling the coefficients as functions of the temperature, it is evident that the relation between the river flows and other meteorological variables have patterns that vary with seasonal changes, as the two regime threshold model fails to reveal. It is also worth pointing out that, although the residuals have no strong serial correlations, residuals are relatively larger in the regime where $z_{t-4} > 0$, which indicate the possibility of improving the model by bringing in more meteorological variables.

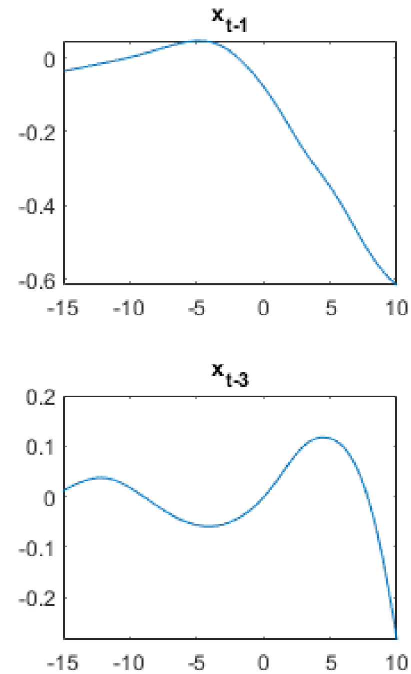

Figure 5.2: LSQR estimates of the coefficients of $x_{t-1}, x_{t-2}, x_{t-3}, z_{t-4}$ with respect to y_{1t}.

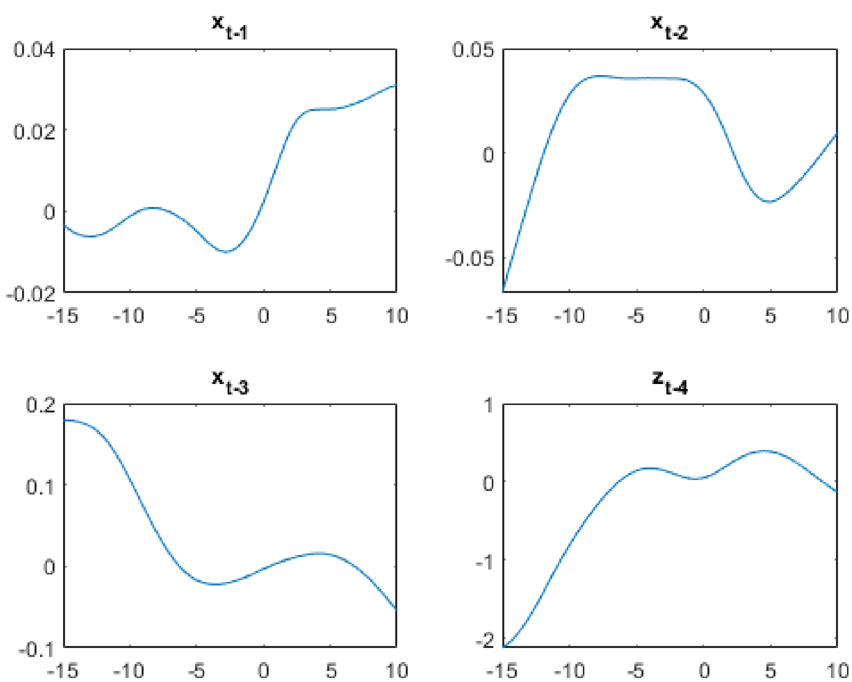

Figure 5.3: LSQR estimates of the coefficients of $x_{t-1}, x_{t-2}, x_{t-3}, z_{t-4}$ with respect to y_{2t}.

5.2 U.S. Interest Rates

The risk-free interest rates implied by the returns of bonds play a fundamental role in the financial modeling. In this example, we consider the series consisting of monthly yields of three-month U.S. treasury bills (T-bill) and three-year treasury notes (T-note), representing the short-term and intermediate-term interest rates. The data, consisting of 409 observations from January 1959 to February 1993, was studied by Tsay (1998) with the multivariate threshold model and Jiang (2014) with model (1.1) using the least square estimation.

Denote the series of treasury bills and treasury bonds by Y_{1t} and Y_{2t}, and let $s_{it} = log(Y_{it})$. We use the logarithmic returns as $\mathbf{y}_t = (y_{1t}, y_{2t})'$, where $y_{it} = s_{it} - s_{i,t-1}$. Let $x_t = s_{1t} - s_{2t}$ be the spread between the logarithmic interest rates, which indicates the status of the U.S. economy. Following Tsay (1998) and Jiang (2014), we use the three-month average of x_t as the threshold variable,

$$z_1 = x_1, \quad z_2 = (x_1 + x_2)/2, \quad z_t = (x_t + x_{t-1} + x_{t-2})/3.$$

We fit the data to the following model by LSQR with $u = (0, 0)'$,

$$\mathbf{y}_t = \mathbf{c}(z_{t-d}) + \gamma(z_{t-d})u_{t-1} + \sum_{i=1}^{7} \boldsymbol{\phi}_i(z_{t-d})\mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t, \tag{5.2}$$

where $u_{t-1} = \boldsymbol{\theta}'\mathbf{S}_t + 0.3984$, $\boldsymbol{\theta} = (1, -1.1151)'$ and $\mathbf{S}_t = (s_{1t}, s_{2t})'$. Here the co-integration is identified by Engle & Granger Test (1987). To save space, we only demonstrate the LSQR estimates of the functional coefficient matrix of the first autoregressive term $\boldsymbol{\phi}_1$ in Figure 5.4. The LSQR estimates show similar patterns as the least squares estimates in Figure 4, Jiang (2014), with narrower confidence intervals. It is seen in Figure 5.4 that the relation of the returns of T-bill and T-note evolve with the economic status: as z_{t-4} increases, indicating better economic status, the return of the T-bill is decreasing and the return of the T-note is increasing, and vice versa. This trend is expected as the market would favor the longer term securities under better economic situations. Figure 5.5 shows the LSQR estimates of the functional coefficient matrix $\boldsymbol{\phi}_1$ with high, low and median quantiles are the same, which validates the model set up.

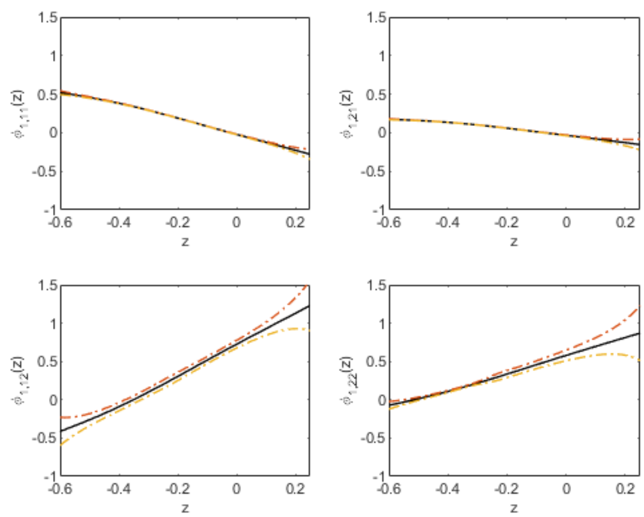

Figure 5.4: LSQR estimates of the functional coefficient matrix $\boldsymbol{\phi}_1$. In each graph, the solid line is the estimated function and the dash lines are the limits of the 95% pointwise confidence intervals.

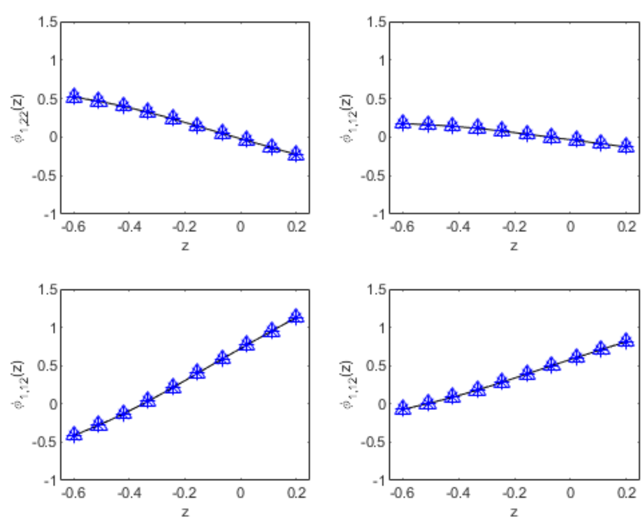

Figure 5.5: LSQR estimates of the functional coefficient matrix $\boldsymbol{\phi}_1$ with high, low and median quantiles. The black solid line is the median estimates; the blue star and triangle markers represent low and high quantile estimates.

CHAPTER 6: DISCUSSION

In this dissertation, we propose a local spatial QR method to estimate the functional-coefficient matrices of multivariate time series. We first propose the local spatial QR estimator by running spatial QR and local smoothing. Then we propose a weighted composite LSQR estimator using the idea of weighted composite QR for better performance. The asymptotic normality of the proposed estimators are established. We also consider the procedures to select the optimal bandwidth and the optimal weights for the estimation. Furthermore, to achieve computational efficiency, we propose a smoothed spatial QR which simplifies and accelerates the minimization problem in the spatial QR. Based on the smoothed spatial QR, we introduce the smoothed LSQR and WCLSQR estimators for the multivariate functional-coefficient model. By establishing the sampling properties of the proposed estimators, we show that the estimators using the smoothed spatial QR can achieve comparable performance with a proper choice of the smoothing parameter while consuming less computing resources. Simulation study of the proposed estimators demonstrates good finite sample performance and computational efficiency. We also analyze the Iceland river flow data and U.S. interest rate data to show the applicability of our method to real data.

Our future work may include the hypothesis testing on the significance of coefficients for LSQR and WCLSQR method. Moreover, a full procedure for model selection may be considered. Both topics have important applications and completes our method, yet are challenging due to the complexity of the multivariate functional-coefficient model.

REFERENCES

[1] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, p. 33, 1978.

[2] P. Chaudhuri, K. Doksum, and A. Samarov, "On average derivative quantile regression," *Annals of Statistics*, vol. 25, pp. 715–744, apr 1997.

[3] R. Koenker, *Quantile Regression.* Cambridge: Cambridge University Press, 2005.

[4] R. Koenker and Q. Zhao, "Conditional quantile estimation and inference for ARCH models," *Econometric Theory*, vol. 12, pp. 793–813, dec 1996.

[5] H. L. Koul and A. K. M. E. Saleh, "Autoregression quantiles and related rank-scores processes," *The Annals of Statistics*, vol. 23, no. 2, pp. 670–689, 1995.

[6] R. A. Davis and W. T. Dunsmuir, "Least absolute deviation estimation for regression with ARMA errors," *Journal of Theoretical Probability*, vol. 10, no. 2, pp. 481–497, 1997.

[7] J. Jiang, Q. Zhao, and Y. Van Hui, "Robust modelling of ARCH models," *Journal of Forecasting*, vol. 20, pp. 111–133, mar 2001.

[8] L. Peng and Q. Yao, "Least absolute deviations estimation for ARCH and GARCH models," *Biometrika*, vol. 90, pp. 967–975, dec 2003.

[9] T. Bollerslev, "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model," *The Review of Economics and Statistics*, vol. 72, p. 498, aug 1990.

[10] R. F. Engle and K. F. Kroner, "Multivariate simultaneous generalized ARCH," *Econometric Theory*, vol. 11, pp. 122–150, feb 1995.

[11] R. Chen and R. S. Tsay, "Functional-coefficient autoregressive models," *Journal of the American Statistical Association*, vol. 88, pp. 298–308, mar 1993.

[12] J. Pan and Q. Yao, "Modelling multiple time series via common factors," *Biometrika*, vol. 95, pp. 365–379, feb 2008.

[13] P. Chaudhuri, "On a geometric notion of quantiles for multivariate data," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 862–872, 1996.

[14] V. I. Koltchinskii, "M-estimation, convexity and quantiles," *The Annals of Statistics*, vol. 25, no. 2, pp. 435–477, 1997.

[15] R. Serfling, "Nonparametric multivariate descriptive measures based on spatial quantiles," *Journal of Statistical Planning and Inference*, vol. 123, no. 2, pp. 259–278, 2004.

[16] H. Tong and K. S. Lim, "Threshold autoregression, limit cycles and cyclical data," in *Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*, vol. 42, pp. 9–56, WileyRoyal Statistical Society, 2009.

[17] H. Tong, *Threshold models in non-linear time series analysis*, vol. 21 of *Lecture Notes in Statistics*. New York, NY: Springer New York, 1983.

[18] Q. Yao and H. Tong, "On initial-condition sensitivity and prediction in nonlinear stochastic systems," *Online*, no. IP 10.3, pp. 395–412, 1995.

[19] R. S. Tsay, "Testing and modeling multivariate threshold models," *Journal of the American Statistical Association*, vol. 93, p. 1188, sep 1998.

[20] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*. Berlin: Springer-Verlag, 2003.

[21] J. Jiang, "Multivariate functional-coefficient regression models for nonlinear vector time series data," *Biometrika*, vol. 101, pp. 689–702, sep 2014.

[22] T. Hastie and R. Tibshirani, "Varying-coefficient models (with discussion)," *Journal of the Royal Statistical Society. Series B*, vol. 55, pp. 757–796, 1993.

[23] J. Fan and W. Zhang, "Statistical estimation in varying coefficient models," *Annals of Statistics*, vol. 27, pp. 1491–1518, oct 1999.

[24] Z. Cai, J. Fan, and Q. Yao, "Functional-coefficient regression models for nonlinear time series," *Journal of the American Statistical Association*, vol. 95, p. 941, sep 2000.

[25] J. Z. Huang and H. Shen, "Functional coefficient regression models for non-linear time series: a polynomial spline approach," *Scandinavian Journal of Statistics*, vol. 31, pp. 515–534, dec 2004.

[26] Z. D. Bai, X. Chen, B. Miao, and C. Radhakrishna Rao, "Asymptotic theory of least distances estimate in multivariate linear models," *Statistics*, vol. 21, pp. 503–519, jan 1990.

[27] B. Chakraborty and P. Chaudhuri, "On an adaptive transformation-retransformation estimate of multivariate location," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 60, no. 1, pp. 145–157, 1998.

[28] B. Chakraborty, "On multivariate quantile regression," *Journal of Statistical Planning and Inference*, vol. 110, pp. 109–132, jan 2003.

[29] R. Koenker, "A note on L-estimates for linear models," *Statistics and Probability Letters*, vol. 2, pp. 323–325, dec 1984.

[30] H. Zou and M. Yuan, "Composite quantile regression and the oracle model selection theory," *Annals of Statistics*, vol. 36, no. 3, pp. 1108–1126, 2008.

[31] J. Bradic, J. Fan, and J. Jiang, "Regularization for Cox's proportional hazards model with NP-dimensionality," *Annals of Statistics*, vol. 39, no. 6, pp. 3092–3120, 2011.

[32] X. Jiang, J. Jiang, and X. Song, "Oracle model selection for nonlinear models based on weighted composite quantile regression," *Statistica Sinica*, vol. 22, no. 4, pp. 1479–1506, 2012.

[33] J. Fan and J. Jiang, "Variable bandwidth and one-step local M-estimator," *Science in China Series A: Mathematics*, vol. 43, pp. 65–81, jan 2000.

[34] H. Tong, B. Thanoon, and G. Gudmundsson, "Threshold time series modeling of two Icelandic riverflow systems," *Journal of the American Water Resources Association*, vol. 21, pp. 651–662, aug 1985.

[35] R. F. Engle and C. W. J. Granger, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, vol. 55, p. 251, mar 1987.

[36] W. Niemiro, "Asymptotics for M-Estimators defined by convex minimization," *The Annals of Statistics*, vol. 20, pp. 1514–1533, sep 1992.

APPENDIX A: CONDITIONS

(A1) The marginal density $f(z)$ of the stationary process $\{z_t\}$ is bounded away from 0 and is continuous at $z = z_0$.

(A2) The functions $\mathbf{c}(\cdot)$, $\boldsymbol{\phi}_i(\cdot)$ and $\boldsymbol{\beta}_i(\cdot)$ have continuous second derivatives at $z = z_0$.

(A3) The kernel function $K(t)$ is continuous with bounded support $[-1, 1]$. Further, the functions $t^3K(t)$ and $t^3K'(t)$ are bounded and $\int t^4K(t)\ dt < \infty$.

(A4) Error term $\boldsymbol{\epsilon}_t$ has an absolutely continuous distribution $g(\mathbf{x})$ on R^k with $k \geq 2$.

(A5) $E[\mathbf{X}_t\mathbf{X}_t^T|z_{t-d} = z_0] < \infty$ and $E[||\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})||^2|z_{t-d} = z_0] < \infty$.

(A6) Matrix $\mathbf{M}(z)$ is continuous and invertible at $z = z_0$.

The above conditions are standard. Condition (A4) ensures that the conditional quantile function uniquely exists. It is worthwhile to point out that, if there is no AR part in model 1.1, Condition (A5) is satisfied even when $\mathbf{a}_t$ has infinite variance.

APPENDIX B: PROOFS of THEOREMS IN CHAPTER 2

Notations

For convenience, we use the following notations throughout the proofs. Let $\hat{\boldsymbol{\gamma}} = \sqrt{nh}[\hat{\mathbf{A}} - \boldsymbol{\Phi}(z_0;\mathbf{u}), h(\hat{\mathbf{B}} - \boldsymbol{\Phi}'(z_0;\mathbf{u})]$ and $\hat{\boldsymbol{\zeta}} = (\hat{\boldsymbol{\zeta}}_{11}, \ldots, \hat{\boldsymbol{\zeta}}_{1J}, \hat{\boldsymbol{\zeta}}_2)$, where $\hat{\boldsymbol{\zeta}}_{1j} = \sqrt{nh}[\hat{\mathbf{c}}_{u_j} - \boldsymbol{\Phi}_1(z_0;\mathbf{u}_j), h(\hat{\mathbf{d}}_{u_j} - \boldsymbol{\Phi}_1'(z_0;\mathbf{u}_j)]$ and $\hat{\boldsymbol{\zeta}}_2 = \sqrt{nh}[\hat{\mathbf{A}}_2 - \boldsymbol{\Phi}_2(z_0;\mathbf{u}), h(\hat{\mathbf{B}}_2 - \boldsymbol{\Phi}_2'(z_0;\mathbf{u})]$.

Define the remainders of the local linear approximation

$$\mathbf{R}(z_{t-d};\mathbf{u}) = \boldsymbol{\Phi}(z_{t-d};\mathbf{u}) - \boldsymbol{\Phi}(z_0;\mathbf{u}) - \boldsymbol{\Phi}'(z_0;\mathbf{u})(z_{t-d} - z_0),$$

$$\mathbf{R}_1(z_{t-d};\mathbf{u}_j) = \boldsymbol{\Phi}_1(z_{t-d};\mathbf{u}_j) - \boldsymbol{\Phi}_1(z_0;\mathbf{u}_j) - \boldsymbol{\Phi}_1'(z_0;\mathbf{u}_j)(z_{t-d} - z_0),$$

$$\mathbf{R}_2(z_{t-d}) = \boldsymbol{\Phi}_2(z_{t-d}) - \boldsymbol{\Phi}_2(z_0) - \boldsymbol{\Phi}_2'(z_0)(z_{t-d} - z_0).$$

Let $\boldsymbol{\eta}_t^u = \boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}) + \mathbf{R}(z_{t-d};\mathbf{u})\mathbf{X}_t$ and $\boldsymbol{\eta}_t^{u_j} = \boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}) + \mathbf{R}_1(z_{t-d};\mathbf{u}_j) + \mathbf{R}_2(z_{t-d})\mathbf{X}_t^*$.

Lemma B.1. *For any quadratic function* $g(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, *where* $\mathbf{A}$ *is a* $p \times p$ *positive definite matrix,* $\mathbf{b}$ *is a* $p \times 1$ *vector and* c *is a constant, we have*

(i) $g(\mathbf{x})$ *achieves the minimum value* $g(\mathbf{x}_0) = c - \frac{1}{4}\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}$ *at* $\mathbf{x}_0 = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$*;*

(ii) $g(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + g(\mathbf{x}_0)$.

Proof of Lemma B.1. Routine. □

Lemma B.2. *Let* $\boldsymbol{\xi}_n = \{vec(\boldsymbol{\gamma})\}^T \frac{1}{\sqrt{nh}} \sum_{t=s'+1}^{n} [\mathbf{W}_{t,h} \otimes \boldsymbol{\psi}_u(\boldsymbol{\eta}_t^u)] K(\frac{z_{t-d}-z_0}{h})$. *Suppose Condition A holds. Then*

$$\boldsymbol{\xi}_n = \{vec(\boldsymbol{\gamma})\}^T \boldsymbol{Z}_n(\mathbf{u}) + \frac{h^2}{2}\sqrt{nh} f(z_0)\{vec(\boldsymbol{\gamma})\}^T vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\} + o_p(\sqrt{nh^5}).$$

Proof of Lemma B.2. Rewrite $\boldsymbol{\xi}_n$ as

$$\begin{aligned}\boldsymbol{\xi}_n =& \frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T \sum_{t=s'+1}^{n} [\mathbf{W}_{t,h} \otimes \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})] K(\frac{z_{t-d} - z_0}{h}) \\ &+ \frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T \sum_{t=s'+1}^{n} [\mathbf{W}_{t,h} \otimes \{\boldsymbol{\psi}_u(\boldsymbol{\eta}_t^u) - \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\}] K(\frac{z_{t-d} - z_0}{h}) \qquad \text{(B.1)} \\ \equiv& \boldsymbol{\xi}_{n1} + \boldsymbol{\xi}_{n2}.\end{aligned}$$

By Taylor's expansion, $\mathbf{R}(z_{t-d}; \mathbf{u}) = \frac{1}{2}\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})(z_{t-d} - z_0)^2$, where ξ_{t-d} is between z_{t-d} and z_0 and independent of $\mathbf{u}$. Then

$$\begin{aligned}\boldsymbol{\xi}_{n2} =& \frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T \sum_{t=s'+1}^{n} [\mathbf{W}_{t,h} \otimes \{\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\frac{h^2}{2}\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})\mathbf{X}_t\}] K^{(2)}(\frac{z_{t-d} - z_0}{h}) \\ &+ \frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T \sum_{t=s'+1}^{n} [\mathbf{W}_{t,h} \otimes \{\boldsymbol{\chi}_{t,u}\frac{h^2}{2}\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})\mathbf{X}_t\}] K^{(2)}(\frac{z_{t-d} - z_0}{h}) \\ \equiv& \boldsymbol{\xi}_{n21} + \boldsymbol{\xi}_{n22},\end{aligned}$$

where $K^{(j)}(x) = x^j K(x)$ and $\boldsymbol{\chi}_{t,u} = \boldsymbol{\psi}_u(\boldsymbol{\eta}_t^u) - \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) - \boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\mathbf{R}(z_{t-d}; \mathbf{u})\mathbf{X}_t$. Applying the identity, $vec(\mathbf{a}\mathbf{b}^T) = \mathbf{b} \otimes \mathbf{a}$ for any column vectors $\mathbf{a}$ and $\mathbf{b}$, we obtain that

$$\begin{aligned}&\mathbf{W}_{t,h} \otimes \{\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})\mathbf{X}_t\} \\ =& vec\{\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})\mathbf{X}_t\mathbf{W}_{t,h}^T\} \\ =& vec\{\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})[(1, \frac{z_{t-d} - z_0}{h}) \otimes \mathbf{X}_t\mathbf{X}_t^T]\}.\end{aligned}$$

Hence, by Condition (A1) and (A2),

$$\begin{aligned}&E\left\{[\mathbf{W}_{t,h} \otimes \{\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\frac{h^2}{2}\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})\mathbf{X}_t\}] K^{(2)}(\frac{z_{t-d} - z_0}{h})\right\} \\ =& E\left\{vec\{\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\boldsymbol{\Phi}''(\xi_{t-d}; \mathbf{u})[(1, \frac{z_{t-d} - z_0}{h}) \otimes \mathbf{X}_t\mathbf{X}_t^T]\} K^{(2)}(\frac{z_{t-d} - z_0}{h})\right\} \\ =& hf(z_0)vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0; \mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\}(1 + o(1))\end{aligned}$$

and

$$E(\boldsymbol{\xi}_{n21}) = \frac{h^2}{2}\sqrt{nh}f(z_0)\{vec(\boldsymbol{\gamma})\}^T vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\}(1+o(1)).$$

Note that under Condition (A5), $var(\boldsymbol{\xi}_{n21}) = O(h^4)$. Then $\boldsymbol{\xi}_{n21} = O_p(\sqrt{nh^5})$. Similarly, we can show that $\boldsymbol{\xi}_{n22} = o_p(\sqrt{nh^5})$. It follows that

$$\boldsymbol{\xi}_{n2} = \boldsymbol{\xi}_{n21}(1+o(1)) = \frac{h^2}{2}\sqrt{nh}f(z_0)\{vec(\boldsymbol{\gamma})\}^T vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\} + o_p(\sqrt{nh^5}).$$

As $\boldsymbol{\xi}_{n1} = \{vec(\boldsymbol{\gamma})\}^T \mathbf{Z}_n(\mathbf{u})$, by (B.1), we complete the proof. □

Lemma B.3. *Suppose Condition A holds. Then* $\mathbf{Z}_n(\mathbf{u}) = \frac{1}{\sqrt{nh}}\sum_{t=s'+1}^{n}[\mathbf{W}_{t,h} \otimes \psi_u(\epsilon_t - q_u(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h})$ *has*

(i) $E[\mathbf{Z}_n(\mathbf{u})] = 0$ *and* $Var[\mathbf{Z}_n(\mathbf{u})] = f(z_0)(\mathbf{V} \otimes \mathbf{M}(z_0)) \otimes \mathbf{N}_u(z_0)(1+o(1))$*;*

(ii) $\mathbf{Z}_n(\mathbf{u})$ *is asymptotically normal with mean 0 and variance of* $f(z_0)(\mathbf{V}\otimes\mathbf{M}(z_0))\otimes\mathbf{N}_u(z_0)$.

Proof Lemma B.3. Since $E[\boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))|z_{t-d}] = 0$, by taking iterative expectations, it is easy to obtain $E[\mathbf{Z}_n(\mathbf{u})] = 0$.

By $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$,

$$\mathbf{W}_{t,h}^{\otimes 2} = \begin{bmatrix} 1 & h^{-1}(z_{t-d}-z_0) \\ h^{-1}(z_{t-d}-z_0) & h^{-2}(z_{t-d}-z_0)^2 \end{bmatrix} \otimes (\mathbf{X}_t\mathbf{X}_t^T),$$

where $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^T$. Similarly,

$$\begin{aligned} E[\mathbf{Z}_n(\mathbf{u})] &= \frac{1}{nh}\sum_{t=s'+1}^{n} E[\mathbf{W}_{t,h}^{\otimes 2} \otimes \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))^{\otimes 2} K(\frac{z_{t-d}-z_0}{h})] \\ &= f(z_0)(\mathbf{V}\otimes\mathbf{M}(z_0))\otimes\mathbf{N}_u(z_0)(1+o(1)) \end{aligned}$$

Then by the martingale central limit theorem, $\mathbf{Z}_n(\mathbf{u})$ is asymptotically normal with mean 0 and variance of $f(z_0)(\mathbf{V}\otimes\mathbf{M}(z_0))\otimes\mathbf{N}_u(z_0)$. □

Proof of Theorem 2.1. Since

$$\begin{aligned}\mathbf{y}_t-[\hat{\mathbf{A}}+\hat{\mathbf{B}}(z_{t-d}-z_0)]\mathbf{X}_t &= \mathbf{y}_t-[\mathbf{A}+\mathbf{B}(z_{t-d}-z_0)]\mathbf{X}_t\\ &\quad -[\hat{\mathbf{A}}-\mathbf{A}+(\hat{\mathbf{B}}-\mathbf{B})(z_{t-d}-z_0)]\mathbf{X}_t\\ &= \boldsymbol{\eta}_t^u-\frac{1}{\sqrt{nh}}\hat{\boldsymbol{\gamma}}\mathbf{W}_{t,h},\end{aligned}$$

the objective function (2.6) can be written as

$$\sum_{t=s'+1}^{n} Q_u(\boldsymbol{\eta}_t^u-\frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})K(\frac{z_{t-d}-z_0}{h}).$$

Therefore, minimizing (2.6) over A, B is equivalent to minimizing

$$\sum_{t=s'+1}^{n}[Q_u(\boldsymbol{\eta}_t^u-\frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})-Q_u(\boldsymbol{\eta}_t^u)]K(\frac{z_{t-d}-z_0}{h}). \qquad \text{(B.2)}$$

over $\boldsymbol{\gamma}$. In the following part we approximate (B.2) by a quadratic function. To this end, we define

$$V_{n,t}=\left\{Q_u(\boldsymbol{\eta}_t^u-\frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})-Q_u(\boldsymbol{\eta}_t^u)+\frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T[\mathbf{W}_{t,h}\otimes\boldsymbol{\psi}_u(\boldsymbol{\eta}_t^u)]\right\}K(\frac{z_{t-d}-z_0}{h}).$$

Since $Q_u(\boldsymbol{\eta}_t^u-\lambda\frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})-Q_u(\boldsymbol{\eta}_t^u)$ is a convex function of λ, the gradient of the function in λ, $-\frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T[\mathbf{W}_{t,h}\otimes\boldsymbol{\psi}_u(\boldsymbol{\eta}_t^u-\lambda\frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})]$ is non-decreasing in λ. Then simple geometry leads to

$$\begin{aligned}0\le V_{n,t}\le &-\frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T[\mathbf{W}_{t,h}\otimes\boldsymbol{\psi}_u(\boldsymbol{\eta}_t^u-\frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})]K(\frac{z_{t-d}-z_0}{h})\\ &+\frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T[\mathbf{W}_{t,h}\otimes\boldsymbol{\psi}_u(\boldsymbol{\eta}_t^u)]K(\frac{z_{t-d}-z_0}{h})\equiv V_{n,t}^*.\end{aligned}$$

For $||vec(\boldsymbol{\gamma})|| \leq M$,

$$V^*_{n,t} = -\frac{1}{\sqrt{nh}}\{vec(\boldsymbol{\gamma})\}^T \left\{\mathbf{W}_{t,h} \otimes [\boldsymbol{\psi}_u(\boldsymbol{\eta}^u_t - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h}) - \boldsymbol{\psi}_u(\boldsymbol{\eta}^u_t)]\right\} K(\frac{z_{t-d} - z_0}{h})$$

converges to zero almost surely as $n \to \infty$. Let $\eta_n \equiv \sum_{t=s'+1}^n V^2_{n,t}$. Since $\boldsymbol{\psi}_u$ is bounded,

$$E(\eta_n) \leq \sum_{t=s'+1}^{n} E[V^{*2}_{n,t}] = O(1)E\left[||\mathbf{W}_{t,h}||^2 \frac{1}{h} K^2(\frac{z_{t-d} - z_0}{h})\right] < \infty$$

. By the Lebesgue dominated convergence theorem, $E(\eta_n) = \sum_{t=s'+1}^n E[V^2_{n,t}] \to 0$. Then, by the Chebyshev's inequality,

$$\begin{aligned}\sum_{t=s'+1}^{n} V_{n,t} &= E(\sum_{t=s'+1}^{n} V_{n,t}) + O_p(\{var(\sum_{t=s'+1}^{n} V_{n,t})\}^{1/2}) \\ &= (n - s')E(V_{n,t}) + o_p(1).\end{aligned} \tag{B.3}$$

By the definition of $V_{n,t}$ and Taylor's expansion at $\boldsymbol{\gamma} = 0$, we have

$$E(V_{n,t}) = \frac{1}{2nh}\{vec(\boldsymbol{\gamma})\}^T E\left\{(\mathbf{W}_{t,h} \otimes \mathbf{I}_k)\boldsymbol{\Psi}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))(\mathbf{W}_{t,h} \otimes \mathbf{I}_k)^T K(\frac{z_{t-d} - z_0}{h})\right\} vec(\boldsymbol{\gamma})(1+o(1)).$$

Using the identity for conforming matrices $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, we have

$$(\mathbf{W}_{t,h} \otimes \mathbf{I}_k)\boldsymbol{\Psi}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))(\mathbf{W}_{t,h} \otimes \mathbf{I}_k)^T = (\mathbf{W}_{t,h}\mathbf{W}^T_{t,h}) \otimes \boldsymbol{\Psi}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})).$$

Then

$$E(V_{n,t}) = \frac{1}{2nh}\{vec(\boldsymbol{\gamma})\}^T E\left\{(\mathbf{W}_{t,h}\mathbf{W}^T_{t,h}) \otimes \boldsymbol{\Psi}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))K(\frac{z_{t-d} - z_0}{h})\right\} vec(\boldsymbol{\gamma})(1 + o(1)).$$

Simple algebra leads to $\mathbf{W}_{t,h}\mathbf{W}_{t,h}^T = \mathbf{S}_n \otimes \mathbf{X}_t\mathbf{X}_t^T$, where

$$\mathbf{S}_n = \begin{bmatrix} 1 & h^{-1}(z_{t-d} - z_0) \\ h^{-1}(z_{t-d} - z_0) & h^{-2}(z_{t-d} - z_0)^2 \end{bmatrix}.$$

Taking iterative expectation, we obtain that

$$E(V_{n,t}) = \frac{1}{2n} f(z_0)\{vec(\boldsymbol{\gamma})\}^T((\mathbf{S} \otimes \mathbf{M}(z_0)) \otimes \mathbf{D}_u(z_0))vec(\boldsymbol{\gamma})(1 + o(1)).$$

This, combined with (B.3), yields that

$$\sum_{t=s'+1}^{n} V_{n,t} = \frac{1}{2} f(z_0)\{vec(\boldsymbol{\gamma})\}^T((\mathbf{S} \otimes \mathbf{M}(z_0)) \otimes \mathbf{D}_u(z_0))vec(\boldsymbol{\gamma}) + o_p(1). \quad \text{(B.4)}$$

By Lemma B.2 and the definition of $V_{n,t}$, we have

$$\begin{aligned} \sum_{t=s'+1}^{n} V_{n,t} = & \sum_{t=s'+1}^{n} [Q_u(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h}) - Q_u(\boldsymbol{\eta}_t^u)]K(\frac{z_{t-d} - z_0}{h}) \\ & + \frac{h^2}{2}\sqrt{nh} f(z_0)\{vec(\boldsymbol{\gamma})\}^T vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0; \mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\} \\ & + \{vec(\boldsymbol{\gamma})\}^T \boldsymbol{Z}_n(\mathbf{u}) + o_p(\sqrt{nh^5}) \end{aligned} \quad \text{(B.5)}$$

Combine (B.4) and (B.5) leads to the following quadratic approximation

$$\sum_{t=s'+1}^{n} [Q_u(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h}) - Q_u(\boldsymbol{\eta}_t^u)]K(\frac{z_{t-d} - z_0}{h}) = G_n(\boldsymbol{\gamma}) + o_p(1), \quad \text{(B.6)}$$

where $nh^5 = O(1)$ and

$$\begin{aligned} G_n(\boldsymbol{\gamma}) = & \frac{1}{2} f(z_0)\{vec(\boldsymbol{\gamma})\}^T((\mathbf{S} \otimes \mathbf{M}(z_0)) \otimes \mathbf{D}_u(z_0))vec(\boldsymbol{\gamma}) \\ & - \frac{h^2}{2}\sqrt{nh} f(z_0)\{vec(\boldsymbol{\gamma})\}^T vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0; \mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\} - \{vec(\boldsymbol{\gamma})\}^T \boldsymbol{Z}_n(\mathbf{u}) \end{aligned}$$

is a quadratic function of $vec(\boldsymbol{\gamma})$. Since the left-hand side of (B.6) is convex, by Lemma 3 of

Niemiro (1992), (B.6) holds uniformly for $||vec(\boldsymbol{\gamma})|| \leq M$. That is, for any $\epsilon > 0$ and $M > 0$, when n is large, with probability at least $1 - \epsilon$, we have

$$\sup_{||vec(\boldsymbol{\gamma})|| \leq M} |\sum_{t=s'+1}^{n} [Q_u(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h}) - Q_u(\boldsymbol{\eta}_t^u)]K(\frac{z_{t-d} - z_0}{h}) - G_n(\boldsymbol{\gamma})| < \epsilon. \tag{B.7}$$

For function $G_n(\boldsymbol{\gamma})$, applying Lemma B.1 with

$$A = \frac{1}{2}f(z_0)((\mathbf{S} \otimes \mathbf{M}(z_0)) \otimes \mathbf{D}_u(z_0)),$$

$$\mathbf{b} = -\frac{h^2}{2}\sqrt{nh}f(z_0)vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0; \mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\} - \boldsymbol{Z}_n(\mathbf{u}),$$

and $c = 0$, we get the minimizer, $vec(\boldsymbol{\gamma}_0) = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$, of $G_n(\boldsymbol{\gamma})$. Let $K = \sqrt{2\lambda_{min}^{-1}(\mathbf{A})}$, where $\lambda_{min}^{-1}(\mathbf{A})$ is the smallest eigenvalue of $\mathbf{A}$. Since $\mathbf{A}$ is positive definite, $K < \infty$. Consider the ball centered at $vec(\boldsymbol{\gamma}_0)$,

$$\mathbf{O}_\gamma = \{\boldsymbol{\gamma} : ||vec(\boldsymbol{\gamma}) - vec(\boldsymbol{\gamma}_0)|| \leq K\sqrt{\epsilon}\}.$$

For any $\boldsymbol{\gamma}$ on the surface the ball $\mathbf{O}_\gamma$, by Lemma B.1 (ii), we have

$$G_n(\boldsymbol{\gamma}) \geq (K\sqrt{\epsilon})^2\lambda_{min}(A) + G_n(\boldsymbol{\gamma}_0) = 2\epsilon + G_n(\boldsymbol{\gamma}_0). \tag{B.8}$$

By (B.7), it is seen that

$$\sum_{t=s'+1}^{n} [Q_u(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h}) - Q_u(\boldsymbol{\eta}_t^u)]K(\frac{z_{t-d} - z_0}{h}) < G_n(\boldsymbol{\gamma}) + \epsilon.$$

By (B.7) and (B.8), for any $\boldsymbol{\gamma}$ on the surface of $\mathbf{O}_\gamma$,

$$\sum_{t=s'+1}^{n} [Q_u(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h}) - Q_u(\boldsymbol{\eta}_t^u)]K(\frac{z_{t-d} - z_0}{h}) > G_n(\boldsymbol{\gamma}) - \epsilon \geq G_n(\boldsymbol{\gamma}_0) + \epsilon.$$

Therefore, the minimum of the convex function (B.2) must be achieved at the interior of the ball $\mathbf{O}_\gamma$. This amounts to

$$||vec(\hat{\boldsymbol{\gamma}}) - vec(\boldsymbol{\gamma}_0)|| \leq K\sqrt{\epsilon}$$

or equivalently

$$vec(\hat{\boldsymbol{\gamma}}) - f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1} \\ \times \left\{ \frac{h^2}{2}\sqrt{nh} f(z_0) vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\} + \mathbf{Z}_n(\mathbf{u}) \right\} = o_p(1) \quad \text{(B.9)}$$

By the identity $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})vec\mathbf{X}$, we have

$$\begin{aligned} &(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1} vec\{\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]\} \\ &= vec\{\boldsymbol{\Phi}''(z_0;\mathbf{u})(\mathbf{s}^T\mathbf{S}^{-1}) \otimes \mathbf{I}_m\} \\ &= vec\{(\mathbf{S}^{-1}\mathbf{s}^T) \otimes \boldsymbol{\Phi}''(z_0;\mathbf{u})\} \\ &= (\mathbf{S}^{-1}\mathbf{s}^T) \otimes vec(\boldsymbol{\Phi}''(z_0;\mathbf{u})). \end{aligned}$$

Then (B.9) is equivalent to

$$vec(\hat{\boldsymbol{\gamma}}) - \sqrt{nh}\mathbf{B}_n(z_0;\mathbf{u}) = f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1}\mathbf{Z}_n + o_p(1),$$

which completes the proof of the theorem. □

Proof of Theorem 2.2. By Lemma B.3, the variance matrix of $f^{-1}(z_0)(\mathbf{S}\otimes\mathbf{M}(z_0)\otimes\mathbf{D}_u(z_0))^{-1}\mathbf{Z}_n$ is

$$\Omega(z_0) = f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1}(\mathbf{V} \otimes \mathbf{M}(z_0) \otimes \mathbf{N}_u(z_0))(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1}.$$

Applying $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, we establish that

$$\boldsymbol{\Omega}(z_0) = f^{-1}(z_0)(\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}) \otimes \mathbf{M}^{-1}(z_0) \otimes (\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0)).$$

This, together with Theorem 2.1 and Lemma B.3, completes the proof of the theorem. □

Proof of Theorem 2.3. Using similar arguments from Theorem 2.1, we can complete the proof. Here we just sketch the proof for saving space.

With new notations $\boldsymbol{\zeta}$, minimizing (2.9) is equivalent to minimizing

$$\sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} [Q_{u_j}(\boldsymbol{\eta}_t^{u_j} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_{1j}\mathbf{W}_{1t,h} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_2\mathbf{W}_{2t,h}) - Q_{u_j}(\boldsymbol{\eta}_t^{u_j})]K(\frac{z_{t-d} - z_0}{h}) \quad \text{(B.10)}$$

Define

$$\begin{aligned} V_{n,t2} = & \sum_{j=1}^{J} \omega_j \Big\{ Q_{u_j}(\boldsymbol{\eta}_t^{u_j} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_{1j}\mathbf{W}_{1t,h} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_2\mathbf{W}_{2t,h}) - Q_{u_j}(\boldsymbol{\eta}_t^{u_j}) \\ & + \frac{1}{\sqrt{nh}} \begin{pmatrix} vec(\boldsymbol{\zeta}_{1j}) \\ vec(\boldsymbol{\zeta}_2) \end{pmatrix}^T \Big[\begin{pmatrix} \mathbf{W}_{1t,h} \\ \mathbf{W}_{2t,h} \end{pmatrix} \otimes \boldsymbol{\psi}_{u_j}(\boldsymbol{\eta}_t^{u_j}) \Big] \Big\} K(\frac{z_{t-d} - z_0}{h}). \end{aligned}$$

With the same argument as between (B.2) and (B.4), we obtain that

$$\begin{aligned} V_{n,t2} = & \frac{1}{2} f(z_0) \sum_{j=1}^{J} \omega_j \Big[\{vec(\boldsymbol{\zeta}_{1j})\}^T \mathbf{S} \otimes \mathbf{D}_{u_j}(z_0) vec(\boldsymbol{\zeta}_{1j}) \\ & + \{vec(\boldsymbol{\zeta}_2)\}^T \mathbf{S} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{D}_{u_j}(z_0) vec(\boldsymbol{\zeta}_2) \\ & + \{vec(\boldsymbol{\zeta}_2)\}^T \mathbf{S} \otimes \boldsymbol{\mu}^*(z_0) \otimes \mathbf{D}_{u_j}(z_0) vec(\boldsymbol{\zeta}_2) \Big] + o_p(1). \end{aligned} \quad \text{(B.11)}$$

Similar to (B.6), if $nh^5 = O(1)$, (B.10) can be approximated by the quadratic function

$$G_{n2}(\boldsymbol{\zeta}) = G_{n21} + G_{n22} - \sum_{j=1}^{J} \omega_j [\{vec(\boldsymbol{\zeta}_{1j})\}^T \mathbf{Z}_{n1j} + \{vec(\boldsymbol{\zeta}_2)\}^T \mathbf{Z}_{n2j}],$$

where

$$G_{n21} = \frac{1}{2} f(z_0) \sum_{j=1}^{J} \omega_j \Big[\{vec(\boldsymbol{\zeta}_{1j})\}^T \mathbf{S} \otimes \mathbf{D}_{u_j}(z_0) vec(\boldsymbol{\zeta}_{1j})$$
$$+ \{vec(\boldsymbol{\zeta}_2)\}^T \mathbf{S} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{D}_{u_j}(z_0) vec(\boldsymbol{\zeta}_2)$$
$$+ \{vec(\boldsymbol{\zeta}_2)\}^T \mathbf{S} \otimes \boldsymbol{\mu}^*(z_0) \otimes \mathbf{D}_{u_j}(z_0) vec(\boldsymbol{\zeta}_2) \Big],$$

and

$$G_{n22} = \frac{h^2}{2} \sqrt{nh} f(z_0) \sum_{j=1}^{J} \omega_j \Big\{ \{vec(\boldsymbol{\zeta}_{1j})\}^T \big[vec\{\mathbf{D}_{u_j}(z_0) \boldsymbol{\Phi}_1''(z_0) \mathbf{s}^T\}$$
$$+ vec\{\mathbf{D}_{u_j}(z_0) \boldsymbol{\Phi}_2''(z_0) [\mathbf{s}^T \otimes \boldsymbol{\mu}^*(z_0)]\} \big]$$
$$+ \{vec(\boldsymbol{\zeta}_2)\}^T \big[vec\{\mathbf{D}_{u_j}(z_0) \boldsymbol{\Phi}_1''(z_0) [\mathbf{s}^T \otimes \boldsymbol{\mu}^*(z_0)]\}$$
$$+ vec\{\mathbf{D}_{u_j}(z_0) \boldsymbol{\Phi}_2''(z_0) [\mathbf{s}^T \otimes \mathbf{M}^*(z_0)]\} \big] \Big\}.$$

By finding the minimizer of $G_{n2}(\boldsymbol{\zeta})$, it can be shown that

$$vec(\hat{\boldsymbol{\zeta}}_2) - \sqrt{nh} \mathbf{B}_{n2}(z_0) = f^{-1}(z_0) (\mathbf{S} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{D}(z_0; \boldsymbol{\omega}))^{-1} Z_n(\boldsymbol{\omega}) + o_p(1), \tag{B.12}$$

where $\mathbf{B}_{n2}(z_0) = \frac{1}{2} h^2 (\mathbf{S}^{-1}\mathbf{s}) \otimes vec(\boldsymbol{\Phi}_2''(z_0; \mathbf{u}))$. □

Proof of Theorem 2.4. Note that

$$cov(\mathbf{Z}_{nj}, \mathbf{Z}_{nl}) = f(z_0) \mathbf{V} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{N}_{u_j, u_l}(z_0) + o(1),$$

where $\mathbf{N}_{u_j,u_l}(z_0) = E[\boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))\boldsymbol{\psi}_{u_l}^T(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_l}(z_{t-d}))|z_{t-d} = z_0]$. It follow that

$$\begin{aligned} var(\mathbf{Z}_n) &= \sum_{j,l=1}^{J} \omega_j \omega_l \; cov(\mathbf{Z}_{nj}, \mathbf{Z}_{nl}) \\ &= f(z_0)\mathbf{V} \otimes \mathbf{M}^*(z_0) \otimes \sum_{j,l=1}^{J} \omega_j \omega_l \mathbf{N}_{u_j,u_l}(z_0) + o(1). \end{aligned} \tag{B.13}$$

Using the Slutsky theorem and (B.12)-(B.13), we obtain the asymptotic variance of $vec(\hat{\boldsymbol{\zeta}}_2)$

$$\boldsymbol{\Omega}_2(z_0; \boldsymbol{\omega}) = f^{-1}(z_0)(\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}) \otimes \mathbf{M}^{*-1}(z_0) \otimes (\mathbf{D}^{-1}(z_0; \boldsymbol{\omega})\mathbf{N}(z_0; \boldsymbol{\omega})\mathbf{D}^{-1}(z_0; \boldsymbol{\omega})).$$

□

APPENDIX C: PROOFS of THEOREMS IN CHAPTER 3

Notation

Except the notations from Appendix B, the following notations are needed throughout the proofs. Let $\tilde{\boldsymbol{\gamma}} = \sqrt{nh}[\tilde{\mathbf{A}} - \boldsymbol{\Phi}(z_0;\mathbf{u}), h(\tilde{\mathbf{B}} - \boldsymbol{\Phi}'(z_0;\mathbf{u})]$ and $\tilde{\boldsymbol{\zeta}} = (\tilde{\boldsymbol{\zeta}}_{11}, \ldots, \tilde{\boldsymbol{\zeta}}_{1J}, \tilde{\boldsymbol{\zeta}}_2)$, where $\tilde{\boldsymbol{\zeta}}_{1j} = \sqrt{nh}[\tilde{\mathbf{c}}_{u_j} - \boldsymbol{\Phi}_1(z_0;\mathbf{u}_j), h(\tilde{\mathbf{d}}_{u_j} - \boldsymbol{\Phi}_1'(z_0;\mathbf{u}_j)]$ and $\tilde{\boldsymbol{\zeta}}_2 = \sqrt{nh}[\tilde{\mathbf{A}}_2 - \boldsymbol{\Phi}_2(z_0), h(\tilde{\mathbf{B}}_2 - \boldsymbol{\Phi}_2'(z_0)]$. By the definition (3.1), $\tilde{\boldsymbol{\gamma}} = arg\ \min_{\boldsymbol{\gamma}}\ L_n(\boldsymbol{\gamma})$, where

$$L_n(\boldsymbol{\gamma}) = \sum_{t=s'+1}^{n} Q_{u,\delta}(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})K(\frac{z_{t-d}-z_0}{h}).$$

And $\tilde{\boldsymbol{\zeta}} = arg\ \min_{\boldsymbol{\gamma}}\ L_n(\boldsymbol{\zeta};\boldsymbol{\omega})$, as defined in (3.2), where

$$L_n(\boldsymbol{\zeta};\boldsymbol{\omega}) \equiv \sum_{j=1}^{J}\omega_j \sum_{t=s'+1}^{n} Q_{u_j,\delta}(\boldsymbol{\eta}_t^{u_j} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_{1j}\mathbf{W}_{1t,h} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_2\mathbf{W}_{2t,h})K(\frac{z_{t-d}-z_0}{h}).$$

Lemma C.1. $L_n'(0) = -\sum_{t=s'+1}^{n}\frac{1}{\sqrt{nh}}\mathbf{W}_{t,h}\otimes\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\eta}_t^u)K(\frac{z_{t-d}-z_0}{h})$. *Suppose condition A holds and* $\delta^k nh \to \infty$, *then*

$$\begin{aligned} L_n'(0) = &-\mathbf{Z}_n(\mathbf{u}) - \sqrt{nh}f(z_0)\mathbf{s}_0\otimes\boldsymbol{\mu}(z_0)\otimes\mathbf{n}_{u,\delta}(z_0) \\ &-\frac{h^2}{2}\sqrt{nh}f(z_0)vec(\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T\otimes\mathbf{M}(z_0)]) \qquad + o_p(\delta^k\sqrt{nh}) + o_p(\sqrt{nh^5}). \end{aligned}$$

Proof of Lemma C.1. $L_n'(0)$ can be written as

$$\begin{aligned} L_n'(0) = &-\sum_{t=s'+1}^{n}\frac{1}{\sqrt{nh}}\mathbf{W}_{t,h}\otimes\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\eta}_t^u)K(\frac{z_{t-d}-z_0}{h}) \\ = &-\{\sum_{t=s'+1}^{n}\frac{1}{\sqrt{nh}}\mathbf{W}_{t,h}\otimes\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))K(\frac{z_{t-d}-z_0}{h}) \\ &+\sum_{t=s'+1}^{n}\frac{1}{\sqrt{nh}}\mathbf{W}_{t,h}\otimes[\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\eta}_t^u) - \boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h})\} \\ = &-\{\vartheta_1 + \vartheta_2\}. \end{aligned} \tag{C.1}$$

For the first term,

$$\vartheta_1 = \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} \mathbf{W}_{t,h} \otimes \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))K(\frac{z_{t-d}-z_0}{h})$$
$$+ \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} \mathbf{W}_{t,h} \otimes [\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) - \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h}) \quad \text{(C.2)}$$
$$= \mathbf{Z}_n(\mathbf{u}) + \vartheta_{11}.$$

Denote $\mathbf{n}_{u,\delta}(z_0) = E[\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) - \boldsymbol{\psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))|z_{t-d} = z_0]$, $\boldsymbol{\mu}(z_0) = E[\mathbf{X}_t|z_{t-d} = z_0]$, $\mu_i = \int u^i K(u)\, du$, $\mathbf{s}_0 = (1, \mu_1)^T$. By taking iterative expectations, we have

$$E(\vartheta_{11}) = \sqrt{nh} f(z_0)\mathbf{s}_0 \otimes \boldsymbol{\mu}(z_0) \otimes \mathbf{n}_{u,\delta}(z_0).$$

Note that, by the definition of $\boldsymbol{\psi}_{u,\delta}(\cdot)$,

$$\mathbf{n}_{u,\delta}(z_0) = \int_{\mathbf{x}\in R^k} [\boldsymbol{\psi}_{u,\delta}(\mathbf{x}) - \boldsymbol{\psi}_u(\mathbf{x})]g_{z_0}(\mathbf{x})\, d\mathbf{x} = \int_{\mathbf{x}\in \mathcal{B}_\delta} [\boldsymbol{\psi}_{u,\delta}(\mathbf{x}) - \boldsymbol{\psi}_u(\mathbf{x})]g_{z_0}(\mathbf{x})\, d\mathbf{x} = O(\delta^k), \quad \text{(C.3)}$$

where $\mathcal{B}_\delta$ is a k-dimensional ball centered at 0 with radius δ, $g_{z_0}(\cdot)$ is the conditional density function of $\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})$ under $z_{t-d} = z_0$. Under conditions (A2) and (A4),

$$E(\vartheta_{11}) = O(\delta^k \sqrt{nh}).$$

Similarly we can obtain that $Var(\vartheta_{11}) = O(\delta^k)$. By Chebyshev's theorem,

$$\vartheta_{11} = E(\vartheta_{11}) + O_p(\{var(\vartheta_{11})\}^{1/2}) = E(\vartheta_{11}) + O_p(\delta^{k/2}) = E(\vartheta_{11}) + o_p(\delta^k \sqrt{nh}).$$

Hence

$$\vartheta_{11} = \sqrt{nh} f(z_0)\mathbf{s}_0 \otimes \boldsymbol{\mu}(z_0) \otimes \mathbf{n}_{u,\delta}(z_0) + o_p(\delta^k \sqrt{nh}). \tag{C.4}$$

For the second term, we have $\boldsymbol{\eta}_t^u = \boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}) + \mathbf{R}(z_{t-d}; \mathbf{u})\mathbf{X}_t$ and $\mathbf{R}(z_{t-d}; \mathbf{u}) = \frac{1}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})(z_{t-d} - z_0)^2$, where z^* is between z_{t-d} and z_0. Let $K^{(j)}(x) = x^j K(x)$ and $\boldsymbol{\chi}_{t,u,\delta} = \boldsymbol{\psi}_{u,\delta}(\boldsymbol{\eta}_t^u) - \boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) - \boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\mathbf{R}(z_{t-d}; \mathbf{u})\mathbf{X}_t$. Then

$$\begin{aligned}
\vartheta_2 &= \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} \mathbf{W}_{t,h} \otimes [\boldsymbol{\psi}_{u,\delta}(\boldsymbol{\eta}_t^u) - \boldsymbol{\psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))] K(\frac{z_{t-d} - z_0}{h}) \\
&= \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} \mathbf{W}_{t,h} \otimes [\boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) \frac{1}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})(z_{t-d} - z_0)^2 \mathbf{X}_t] K(\frac{z_{t-d} - z_0}{h}) \\
&+ \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} \mathbf{W}_{t,h} \otimes [\boldsymbol{\chi}_{t,u,\delta} \frac{1}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})(z_{t-d} - z_0)^2 \mathbf{X}_t] K(\frac{z_{t-d} - z_0}{h}) \\
&= \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} \mathbf{W}_{t,h} \otimes [\boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) \frac{h^2}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})\mathbf{X}_t] K^{(2)}(\frac{z_{t-d} - z_0}{h}) \\
&+ \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} \mathbf{W}_{t,h} \otimes [\boldsymbol{\chi}_{t,u,\delta} \frac{h^2}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})\mathbf{X}_t] K^{(2)}(\frac{z_{t-d} - z_0}{h}) \\
&= \vartheta_{21} + \vartheta_{22}.
\end{aligned}$$

By $vec(\mathbf{A}\mathbf{B}^T) = \mathbf{B} \otimes \mathbf{A}$,

$$\begin{aligned}
\vartheta_{21} &= \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} vec(\boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) \frac{h^2}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})\mathbf{X}_t \mathbf{W}_{t,h}^T) K^{(2)}(\frac{z_{t-d} - z_0}{h}) \\
&= \sum_{t=s'+1}^{n} \frac{1}{\sqrt{nh}} vec(\boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) \frac{h^2}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})[(1, \frac{z_{t-d} - z_0}{h}) \otimes \mathbf{X}_t \mathbf{X}_t^T]) K^{(2)}(\frac{z_{t-d} - z_0}{h}).
\end{aligned}$$

By similar arguments for (C.3), it can be obtained that

$$E[\boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) | z_{t-d} = z_0] = E[\boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) | z_{t-d} = z_0](1 + o(1)) = \mathbf{D}_u(z_0)(1 + o(1)).$$

Denoting $\mathbf{s} = (\mu_2, \mu_3)^T$ and $\mathbf{M}(z_0) = E[\mathbf{X}_t \mathbf{X}_t^T | z_{t-d} = z_0]$, along with conditions (A1)-(A2), we have

$$E(vec(\mathbf{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\frac{h^2}{2}\mathbf{\Phi}''(z^*;\mathbf{u})[(1, \frac{z_{t-d}-z_0}{h}) \otimes \mathbf{X}_t\mathbf{X}_t^T])K^{(2)}(\frac{z_{t-d}-z_0}{h}))$$
$$= hf(z_0)vec(\mathbf{D}_u(z_0)\mathbf{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)])(1+o(1)).$$

Hence,

$$E(\vartheta_{21}) = \frac{h^2}{2}\sqrt{nh}f(z_0)vec(\mathbf{D}_u(z_0)\mathbf{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)])(1+o(1)).$$

And under condition (A5), $var(\vartheta_{21}) = O(h^4)$. Hence $\vartheta_{21} = O_p(\sqrt{nh^5})$. Since $E[\boldsymbol{\chi}_{t,u,\delta}|z_{t-d} = z_0] = o(h^2)$, similarly we have $E(\vartheta_{22}) = o(\sqrt{nh^5})$ and $var(\vartheta_{22}) = o(h4)$. Then $\vartheta_{22} = o_p(\sqrt{nh^5})$ and

$$\vartheta_2 = \vartheta_{21}(1+o_p(1)) = \frac{h^2}{2}\sqrt{nh}f(z_0)vec(\mathbf{D}_u(z_0)\mathbf{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)]) + o_p(\sqrt{nh^5}).$$

By above we complete the proof the lemma. □

Lemma C.2. $L_n''(0) = \sum_{t=s'+1}^n \boldsymbol{\xi}_t'^{\,T}\mathbf{\Psi}_{u,\delta}(\boldsymbol{\eta}_t^u)\boldsymbol{\xi}_t' K(\frac{z_{t-d}-z_0}{h})$. *Suppose condition A hold, then*

$$L_n''(0) = f(z_0)\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0)(1+o_p(1)).$$

Proof of Lemma C.2. Plug in $\xi_t' = -\frac{1}{\sqrt{nh}}\mathbf{W}_{t,h}^T \otimes \mathbf{I}_k$,

$$L_n''(0) = \frac{1}{nh}\sum_{t=s'+1}^n \mathbf{W}_{t,h} \otimes \mathbf{I}_k \mathbf{\Psi}_{u,\delta}(\boldsymbol{\eta}_t^u)\mathbf{W}_{t,h}^T \otimes \mathbf{I}_k K(\frac{z_{t-d}-z_0}{h}).$$

By $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$,

$$L_n''(0) = \frac{1}{nh}\sum_{t=s'+1}^n \mathbf{W}_{t,h}\mathbf{W}_{t,h}^T \otimes \mathbf{\Psi}_{u,\delta}(\boldsymbol{\eta}_t^u)K(\frac{z_{t-d}-z_0}{h})$$
$$= \frac{1}{nh}\sum_{t=s'+1}^n \mathbf{S}_n \otimes (\mathbf{X}_t\mathbf{X}_t^T) \otimes \mathbf{\Psi}_{u,\delta}(\boldsymbol{\eta}_t^u)K(\frac{z_{t-d}-z_0}{h}),$$

where

$$\mathbf{S}_n = \begin{bmatrix} 1 & h^{-1}(z_{t-d} - z_0) \\ h^{-1}(z_{t-d} - z_0) & h^{-2}(z_{t-d} - z_0)^2 \end{bmatrix}.$$

With similar techniques in the proof of Lemma C.1, $L_n''(0)$ can be rewritten as

$$\begin{aligned} L_n''(0) &= \frac{1}{nh}\sum_{t=s'+1}^{n} \mathbf{S}_n \otimes (\mathbf{X}_t\mathbf{X}_t^T) \otimes \boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))K(\frac{z_{t-d} - z_0}{h}) \\ &+ \frac{1}{nh}\sum_{t=s'+1}^{n} \mathbf{S}_n \otimes (\mathbf{X}_t\mathbf{X}_t^T) \otimes [\boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) - \boldsymbol{\Psi}_u(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))]K(\frac{z_{t-d} - z_0}{h}) \\ &+ \frac{1}{nh}\sum_{t=s'+1}^{n} \mathbf{S}_n \otimes (\mathbf{X}_t\mathbf{X}_t^T) \otimes [\boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\eta}_t^u) - \boldsymbol{\Psi}_{u,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))]K(\frac{z_{t-d} - z_0}{h}) \\ &\equiv \tau_1 + \tau_2 + \tau_3. \end{aligned}$$

Taking iterative expectations, we have

$$E[\tau_1] = f(z_0)\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0).$$

And since $Var[\tau_1] = O((nh)^{-1})$, by Chebyshev's theorem,

$$\tau_1 = f(z_0)\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0)(1 + o_p(1)).$$

For the second term, using similar arguments for (C.3), we have $\tau_2 = O_p(\delta^k) = o_p(\tau_1)$. As $\boldsymbol{\eta}_t^u - (\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d})) = \frac{1}{2}\boldsymbol{\Phi}''(z^*; \mathbf{u})(z_{t-d} - z_0)^2\mathbf{X}_t = O(h^2)$, thus $\tau_3 = O_p(h^2) = o_p(\tau_1)$. Combining above,

$$L_n''(0) = f(z_0)\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0)(1 + o_p(1)).$$

□

Proof of Theorem 3.2. Let $\boldsymbol{\xi}_t = \boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h}$. By $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{X})$, we have

$$vec(\boldsymbol{\xi}_t) = vec(\boldsymbol{\eta}_t^u) - \frac{1}{\sqrt{nh}}(\mathbf{W}_{t,h}^T \otimes \mathbf{I}_k)vec(\boldsymbol{\gamma})$$

and

$$\boldsymbol{\xi}_t' = \frac{\partial \boldsymbol{\xi}_t}{\partial vec(\boldsymbol{\gamma})} = -\frac{1}{\sqrt{nh}}\mathbf{W}_{t,h}^T \otimes \mathbf{I}_k.$$

Then,

$$L_n'(\boldsymbol{\gamma}) = \frac{\partial L_n(\boldsymbol{\gamma})}{\partial vec(\boldsymbol{\gamma})} = \sum_{t=s'+1}^{n} \boldsymbol{\xi}_t'^T \boldsymbol{\psi}_{u,\delta}(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})K(\frac{z_{t-d} - z_0}{h}),$$

and

$$L_n''(\boldsymbol{\gamma}) = \frac{\partial^2 L_n(\boldsymbol{\gamma})}{\partial vec(\boldsymbol{\gamma})\partial\{vec(\boldsymbol{\gamma})\}^T} = \sum_{t=s'+1}^{n} \boldsymbol{\xi}_t'^{\,T} \Psi_{u,\delta}(\boldsymbol{\eta}_t^u - \frac{1}{\sqrt{nh}}\boldsymbol{\gamma}\mathbf{W}_{t,h})\boldsymbol{\xi}_t' K(\frac{z_{t-d} - z_0}{h}).$$

By Taylor expansion at 0,

$$L_n'(\hat{\boldsymbol{\gamma}}) = L_n'(0) + L_n''(0)(vec(\hat{\boldsymbol{\gamma}}) - 0) + o(\hat{\boldsymbol{\gamma}}).$$

Noting that by definition $L_n'(\hat{\boldsymbol{\gamma}}) = 0$, with Lemma C.1 and Lemma C.2, we obtain

$$\begin{aligned} vec(\hat{\boldsymbol{\gamma}}) &= -\frac{L_n'(0)}{L_n''(0)}(1 + o(1)) \\ &= f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1}\mathbf{Z}_n(u) + \sqrt{nh}\mathbf{B}_{n,\delta} + o_p(1), \end{aligned}$$

where

$$\begin{aligned} \mathbf{B}_{n,\delta} &= f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}(z_0) \otimes \mathbf{D}_u(z_0))^{-1}\{f(z_0)\mathbf{s}_0 \otimes \boldsymbol{\mu}(z_0) \otimes \mathbf{n}_{u,\delta}(z_0) \\ &+ \frac{h^2}{2} f(z_0) vec(\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0; \mathbf{u})[\mathbf{s}^T \otimes \mathbf{M}(z_0)])\} \\ &= \kappa_1 + \kappa_2. \end{aligned}$$

For the first term,

$$\begin{aligned}\kappa_1 &= (\mathbf{S}\otimes\mathbf{M}(z_0)\otimes\mathbf{D}_u(z_0))^{-1}\mathbf{s}_0\otimes\boldsymbol{\mu}(z_0)\otimes\mathbf{n}_{u,\delta}(z_0)\\ &= \mathbf{S}^{-1}\otimes\mathbf{M}(z_0)^{-1}\otimes\mathbf{D}_u(z_0)^{-1}\mathbf{s}_0\otimes\boldsymbol{\mu}(z_0)\otimes\mathbf{n}_{u,\delta}(z_0)\\ &= \mathbf{S}^{-1}\mathbf{s}_0\otimes[(\mathbf{M}(z_0)^{-1}\boldsymbol{\mu}(z_0))\otimes(\mathbf{D}_u(z_0)^{-1}\mathbf{n}_{u,\delta}(z_0))].\end{aligned}$$

For the second term, by $vec(\mathbf{AXB}) = \mathbf{B}^T\otimes\mathbf{A}\ vec(\mathbf{X})$,

$$\begin{aligned}\kappa_2 &= f^{-1}(z_0)(\mathbf{S}\otimes\mathbf{M}(z_0)\otimes\mathbf{D}_u(z_0))^{-1}\frac{h^2}{2}f(z_0)\\ &\quad vec(\mathbf{D}_u(z_0)\boldsymbol{\Phi}''(z_0;\mathbf{u})[\mathbf{s}^T\otimes\mathbf{M}(z_0)])\\ &= \frac{h^2}{2}\mathbf{S}^{-1}\otimes\mathbf{M}^{-1}(z_0)\otimes\mathbf{D}_u^{-1}(z_0)\mathbf{s}\otimes\mathbf{M}(z_0)\otimes\mathbf{D}_u^{-1}(z_0)vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))\\ &= \frac{h^2}{2}\mathbf{S}^{-1}\mathbf{s}\otimes[\mathbf{M}^{-1}(z_0)\otimes\mathbf{D}_u^{-1}(z_0)\mathbf{M}(z_0)\otimes\mathbf{D}_u^{-1}(z_0)]vec(\boldsymbol{\Phi}''(z_0;\mathbf{u}))\\ &= \frac{h^2}{2}\mathbf{S}^{-1}\mathbf{s}\otimes vec(\boldsymbol{\Phi}''(z_0;\mathbf{u})).\end{aligned}$$

This complete the proof of Theorem 3.2. □

Proof of Theorem 3.3. Using Theorem 3.2 and Lemma B.3, it follows from matrices algebra that

$$\begin{aligned}\Omega(z_0) &= f^{-1}(z_0)(\mathbf{S}\otimes\mathbf{M}(z_0)\otimes\mathbf{D}_u(z_0))^{-1}(\mathbf{V}\otimes\mathbf{M}(z_0)\otimes\mathbf{N}_u(z_0))(\mathbf{S}\otimes\mathbf{M}(z_0)\otimes\mathbf{D}_u(z_0))^{-1}\\ &= f^{-1}(z_0)(\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1})\otimes\mathbf{M}^{-1}(z_0)\otimes\mathbf{D}_u^{-1}(z_0)\mathbf{N}_u(z_0)\mathbf{D}_u^{-1}(z_0).\end{aligned}$$

This, combined with Theorem 3.2 and Lemma B.3, completes the proof the theorem. □

Proof of Theorem 3.4. Let $\boldsymbol{\xi}_t(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2) = \boldsymbol{\eta}_t^{u_j} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_{1j}\mathbf{W}_{1t,h} - \frac{1}{\sqrt{nh}}\boldsymbol{\zeta}_2\mathbf{W}_{2t,h}$. Define the first order derivative with respective to $vec(\boldsymbol{\zeta})$ of $L_n(\boldsymbol{\zeta};\boldsymbol{\omega})$ as $L_n'(\boldsymbol{\zeta};\boldsymbol{\omega})$ and the second order derivative as $L_n''(\boldsymbol{\zeta};\boldsymbol{\omega})$. Then,

$$L_n'(\boldsymbol{\zeta};\boldsymbol{\omega}) = \sum_{j=1}^{J}\omega_j \sum_{t=s'+1}^{n}\{\boldsymbol{\xi}_t'(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2)\}^T \boldsymbol{\psi}_{u_j,\delta}(\boldsymbol{\xi}_t(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2))K(\frac{z_{t-d}-z_0}{h}),$$

and

$$L_n''(\boldsymbol{\zeta};\boldsymbol{\omega}) = \sum_{j=1}^{J}\omega_j \sum_{t=s'+1}^{n}\{\boldsymbol{\xi}_t'(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2)\}^T \boldsymbol{\Psi}_{u_j,\delta}(\boldsymbol{\xi}_t(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2))\boldsymbol{\xi}_t'(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2)K(\frac{z_{t-d}-z_0}{h}),$$

where

$$\boldsymbol{\xi}_t'(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2) = \frac{\partial \boldsymbol{\xi}_t(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2)}{\partial vec(\boldsymbol{\zeta})} = -\frac{1}{\sqrt{nh}}(\mathbf{e}_j^T \otimes \mathbf{W}_{1t,h}^T, \mathbf{W}_{2t,h}^T)\otimes \mathbf{I}_k,$$

$\mathbf{e}_j$ is a $J\times 1$ vector with the j-th component being 1 and the remaining components being 0, $\mathbf{I}_k$ is a $k\times k$ identity matrix. As $\boldsymbol{\xi}_t'(\boldsymbol{\zeta}_{1j},\boldsymbol{\zeta}_2)$ does not depend on $\boldsymbol{\zeta}$, it can be written as $\boldsymbol{\xi}_{tj}'$.

For $L_n''(0;\boldsymbol{\omega})$, we have

$$\begin{aligned} L_n''(0;\boldsymbol{\omega}) &= \sum_{j=1}^{J}\omega_j \sum_{t=s'+1}^{n}\{\boldsymbol{\xi}_{tj}'\}^T \boldsymbol{\Psi}_{u_j,\delta}(\boldsymbol{\eta}_t^{u_j})\boldsymbol{\xi}_{tj}'K(\frac{z_{t-d}-z_0}{h}) \\ &= \sum_{j=1}^{J}\omega_j \sum_{t=s'+1}^{n}\{\boldsymbol{\xi}_{tj}'\}^T \boldsymbol{\Psi}_{u_j,\delta}(\boldsymbol{\epsilon}_t-\mathbf{q}_{u_j}(z_{t-d}))\boldsymbol{\xi}_{tj}'K(\frac{z_{t-d}-z_0}{h}) \\ &\quad + \sum_{j=1}^{J}\omega_j \sum_{t=s'+1}^{n}\{\boldsymbol{\xi}_{tj}'\}^T \Big[\boldsymbol{\Psi}_{u_j,\delta}(\boldsymbol{\eta}_t^{u_j}) - \boldsymbol{\Psi}_{u_j,\delta}(\boldsymbol{\epsilon}_t-\mathbf{q}_{u_j}(z_{t-d}))\Big]\boldsymbol{\xi}_{tj}'K(\frac{z_{t-d}-z_0}{h}) \\ &\equiv L_{n1}'' + L_{n2}''. \end{aligned}$$

With similar arguments in the proof of Lemma C.2, it can be shown that

$$L_n''(0;\boldsymbol{\omega}) = f(z_0)\sum_{j=1}^{J}\omega_j \mathbf{T}_j \otimes \mathbf{D}_{u_j}(z_0)(1+o_p(1)),$$

where

$$\mathbf{T}_j = \begin{bmatrix} \mathbf{e}_j^{\otimes 2}\otimes \mathbf{S} & \mathbf{e}_j \otimes \mathbf{S}\otimes \{\boldsymbol{\mu}^*\}^T \\ \mathbf{e}_j^T \otimes \mathbf{S}\otimes \boldsymbol{\mu}^* & \mathbf{S}\otimes E[\mathbf{X}_t^*\mathbf{X}_t^{*T}|z_{t-d}=z_0] \end{bmatrix}.$$

Denote $\mathbf{W} = \sum_{j=1}^{J} \omega_j \mathbf{T}_j \otimes \mathbf{D}_{u_j}(z_0)$, we have

$$(L_n''(0;\boldsymbol{\omega}))^{-1} = f^{-1}(z_0)\mathbf{W}^{-1}(1+o_p(1)),$$

where $\mathbf{W}^{-1} = \begin{bmatrix} \mathbf{W}^{11} & \mathbf{W}^{12} \\ \mathbf{W}^{21} & \mathbf{W}^{22} \end{bmatrix}$ is the inverse of $\mathbf{W}$, with

$$\mathbf{W}^{21} = -\sum_{j=1}^{J} \mathbf{e}_j^T \otimes \mathbf{S}^{-1} \otimes \{(\mathbf{M}^*(z_0))^{-1}\boldsymbol{\mu}^*\} \otimes \{\mathbf{D}(z_0;\boldsymbol{\omega})\}^{-1}$$

and $\mathbf{W}^{22} = \{\mathbf{S} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{D}(z_0;\boldsymbol{\omega})\}^{-1}$. Tedious calculations are skipped here to save space.

Similar to (C.1) and (C.2), $L_n'(0;\boldsymbol{\omega})$ can be decomposed to

$$L_n'(0;\boldsymbol{\omega}) = L_{n1}' + L_{n2}' + L_{n3}',$$

where

$$L_{n1}' = \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} \{\boldsymbol{\xi}_{tj}'\}^T \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))K(\frac{z_{t-d}-z_0}{h}),$$

$$L_{n2}' = \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} \{\boldsymbol{\xi}_{tj}'\}^T [\boldsymbol{\psi}_{u_j,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d})) - \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h}),$$

$$L_{n3}' = \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} \{\boldsymbol{\xi}_{tj}'\}^T [\boldsymbol{\psi}_{u_j,\delta}(\boldsymbol{\eta}_{\boldsymbol{t}}^{\boldsymbol{u_j}}) - \boldsymbol{\psi}_{u_j,\delta}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h}).$$

Using an arugument similar to that for Lemma C.1, we obtain that $L_{n1}' = [\mathbf{L}^{11}, \mathbf{L}^{12}]^T(1+o_p(1))$, where

$$\mathbf{L}^{11} = -\frac{1}{\sqrt{nh}} \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} \mathbf{e}_j \otimes \mathbf{W}_{1t,h} \otimes \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))K(\frac{z_{t-d}-z_0}{h}),$$

$$\mathbf{L}^{12} = -\frac{1}{\sqrt{nh}} \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} \mathbf{W}_{2t,h} \otimes \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))K(\frac{z_{t-d}-z_0}{h}),$$

and $L'_{n2} = [\mathbf{L}^{21}, \mathbf{L}^{22}]^T(1 + o_p(1))$, where

$$\mathbf{L}^{21} = -\sqrt{nh}f(z_0)\sum_{j=1}^{J}\omega_j\mathbf{e}_j \otimes \mathbf{s}_0 \otimes \mathbf{n}_{u_j,\delta}(z_0),$$

$$\mathbf{L}^{22} = -\sqrt{nh}f(z_0)\sum_{j=1}^{J}\omega_j\mathbf{s}_0 \otimes \boldsymbol{\mu}^* \otimes \mathbf{n}_{u_j,\delta}(z_0),$$

and $L'_{n3} = [\mathbf{L}^{31}, \mathbf{L}^{32}]^T(1 + o_p(1))$, where

$$\mathbf{L}^{31} = -\frac{h^2}{2}\sqrt{nh}f(z_0)\sum_{j=1}^{J}\omega_j\mathbf{e}_j \otimes \big[vec(\mathbf{D}_{u_j}(z_0)(z_0)\boldsymbol{\Phi}_1''(z_0)\mathbf{s}^T) + vec\{\mathbf{D}_{u_j}(z_0)(z_0)\boldsymbol{\Phi}_2''(z_0)(\mathbf{s}^T \otimes \boldsymbol{\mu}^*)\}\big],$$

$$\begin{aligned}\mathbf{L}^{32} = & -\frac{h^2}{2}\sqrt{nh}f(z_0)\sum_{j=1}^{J}\omega_j\big[vec(\mathbf{D}_{u_j}(z_0)(z_0)\boldsymbol{\Phi}_1''(z_0)(\mathbf{s}^T \otimes \boldsymbol{\mu}^{*T})) \\ & + vec\{\mathbf{D}_{u_j}(z_0)(z_0)\boldsymbol{\Phi}_2''(z_0)(\mathbf{s}^T \otimes E[\mathbf{X}_t^*\mathbf{X}_t^{*T}|z_{t-d} = z_0])\}\big].\end{aligned}$$

Then the $(J+1)$th block component of $(L_n''(0;\boldsymbol{\omega}))^{-1}L_n'(0;\boldsymbol{\omega})$ is

$$\begin{aligned}\sum_{l=1}^{3}(\mathbf{W}^{21}\mathbf{L}^{l1} + \mathbf{W}^{22}\mathbf{L}^{l2}) = \{& -f^{-1}(z_0)(\mathbf{S} \otimes \mathbf{M}^*(z_0) \otimes \mathbf{D}(z_0;\boldsymbol{\omega}))^{-1}\mathbf{Z}_n(\boldsymbol{\omega}) \\ & - \frac{1}{2}h^2\sqrt{nh}(\mathbf{S}^{-1}\mathbf{s}) \otimes vec(\boldsymbol{\Phi}_2''(z_0)\}(1 + o_p(1)),\end{aligned}$$

where $\mathbf{Z}_n(\boldsymbol{\omega}) = \sum_{j=1}^{J} w_j[\mathbf{Z}_{n2j} - (\mathbf{I}_2 \otimes \boldsymbol{\mu}^*(z_0) \otimes \mathbf{I}_k)\mathbf{Z}_{n1j}]$ and $\mathbf{Z}_{nij} = \frac{1}{\sqrt{nh}}\sum_{t=s'+1}^{n}[\mathbf{W}_{it,h} \otimes \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d}))]K(\frac{z_{t-d}-z_0}{h})$. This completes the proof of the theorem. $\square$

Proof of Theorem 3.5. Using the same notations from the proof of Theorem 3.4 and by taking iterative iterative expectations, we have $E(L'_{n1} = 0$ and

$$
\begin{aligned}
var(L'_{n1}) = E\{ \sum_{j=1}^{J} \omega_j \sum_{t=s'+1}^{n} \{\boldsymbol{\xi}'_{tj}\}^T \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d})) K(\frac{z_{t-d} - z_0}{h}) \\
\cdot \sum_{l=1}^{J} \omega_l \sum_{s=s'+1}^{n} [\boldsymbol{\psi}_{u_l}(\boldsymbol{\epsilon}_s - \mathbf{q}_{u_l}(z_{s-d}))]^T \boldsymbol{\xi}'_{sl} K(\frac{z_{s-d} - z_0}{h}) \} \\
= E\{ \sum_{j,l=1}^{J} \omega_j \omega_l \sum_{t=s'+1}^{n} \{\boldsymbol{\xi}'_{tj}\}^T \boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_j}(z_{t-d})) [\boldsymbol{\psi}_{u_l}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_l}(z_{t-d}))]^T \boldsymbol{\xi}'_{tl} K^2(\frac{z_{t-d} - z_0}{h}) \}.
\end{aligned}
$$

With the similar arguments in the proof of Lemma C.2, we obtain that

$$
var(L'_{n1}) = f(z_0)\mathbf{W}^*(1 + o_p(1)),
$$

where

$$
\mathbf{W}^* = \sum_{j,l=1}^{J} \omega_j \omega_l \begin{bmatrix} \mathbf{e}_j \mathbf{e}_l^T \otimes \mathbf{V} & \mathbf{e}_j \otimes \mathbf{V} \otimes \boldsymbol{\mu}^{*T} \\ \mathbf{e}_l^T \otimes \mathbf{V} \otimes \boldsymbol{\mu}^* & \mathbf{V} \otimes E[\mathbf{X}^* \mathbf{X}^{*T} | z_{t-d} = z_0] \end{bmatrix} \otimes \mathbf{N}_{u_j, u_l},
$$

and $\mathbf{N}_{u_j,u_l} = E[\boldsymbol{\psi}_{u_j}(\boldsymbol{\epsilon}_t - \mathbf{q}_u(z_{t-d}))\{\boldsymbol{\psi}_{u_l}(\boldsymbol{\epsilon}_t - \mathbf{q}_{u_l}(z_{t-d}))\}^T | z_{t-d} = z_0]$. Then by the martingale central limit theorem, L'_{n1} is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $f(z_0)\mathbf{W}^*$. Hence, by the Slutsky theorem, $(L''_n(0; \boldsymbol{\omega}))^{-1} L'_{n1}$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $f^{-1}(z_0)\mathbf{W}^{-1}\mathbf{W}^*(\mathbf{W}^{-1})^T$. This, combined with Theorem 3.4 completes the proof. □

Proof of Theorem 3.1. As $a(\mathbf{t}'\mathbf{t})^2 + b(\mathbf{t}'\mathbf{t}) + c + \mathbf{u}'\mathbf{t}$ and $Q_u(\mathbf{t})$ are C^2-continuous respectively in $\mathcal{B}_\delta(\mathbf{0})$ and $R^k/\mathcal{B}_\delta(\mathbf{0})$, denoting $a(\mathbf{t}'\mathbf{t})^2 + b(\mathbf{t}'\mathbf{t}) + c + \mathbf{u}'\mathbf{t}$ as $z_u(\mathbf{t})$ it suffices to show that $z_u(\mathbf{t}) = Q_u(\mathbf{t})$, $\partial z_u(\mathbf{t})/\partial \mathbf{t} = \partial Q_u(\mathbf{t})/\partial \mathbf{t}$ and $\partial^2 z_u(\mathbf{t})/\partial \mathbf{t}\mathbf{t}^T = \partial^2 Q_u(\mathbf{t})/\partial \mathbf{t}\mathbf{t}^T$ for all $\mathbf{t}$ with $||\mathbf{t}|| = \delta$. With simple algebra, the above three equations can be written as

$$
\begin{cases}
a(\mathbf{t}'\mathbf{t})^2 + b(\mathbf{t}'\mathbf{t}) + c = \sqrt{\mathbf{t}'\mathbf{t}}, \\
4a(\mathbf{t}'\mathbf{t})\mathbf{t} + 2b\mathbf{t} = \frac{\mathbf{t}}{\sqrt{\mathbf{t}'\mathbf{t}}}, \\
[4a(\mathbf{t}'\mathbf{t}) + 2b]\mathbf{I}_k + 8a(\mathbf{t}\mathbf{t}') = \frac{1}{\sqrt{\mathbf{t}'\mathbf{t}}}\mathbf{I}_k - \frac{1}{(\mathbf{t}'\mathbf{t})^{3/2}}(\mathbf{t}\mathbf{t}').
\end{cases}
$$

Note $\mathbf{t}'\mathbf{t} = \delta^2$, to match the coefficients of the left and right sides, we have

$$\begin{cases} a\delta^4 + b\delta^2 + c = \delta \\ 4a\delta^2 + 2b = \frac{1}{\delta} \\ 8a = -\frac{1}{\delta^3}. \end{cases}$$

Solving the above linear system completes the proof of the theorem. $\square$