MODELING VECTOR TIME SERIES DATA

by

Yi Liu

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

2013

Approved by:

Dr. Jiancheng Jiang

Dr. Yanqing Sun

Dr. Weihua Zhou

Dr. Aixi Zhou

ABSTRACT

YI LIU. Modeling vector time series data.
(Under the direction of DR. JIANCHENG JIANG)

In this dissertation, firstly, I study spatial quantile regression estimation of multivariate threshold time series models. Asymptotic normality of the proposed spatial quantile regression estimator is established. Simulations and a real example are used to evaluate the performance of the proposed estimator. Secondly, I study the multivariate time-varying coefficient models for time series data. An explicit solution of the coefficient estimators is given in the paper. Furthermore, I propose generalized likelihood ratio test for the multivariate time-varying coefficient models, my aim is to construct some test statistics to test whether the coefficients are constants or of some specific parametric functional for the time-varying coefficient model. The asymptotic null distribution of the proposed test statistics is presented and shown to be independent of the nuisance parameters. Simulation results for the power of the test and a real example are reported at the end of this dissertation.

ACKNOWLEDGMENTS

First of all, I would like to gratefully and sincerely thank my advisor Dr. Jiancheng Jiang for his excellent guidance, caring and patience during my graduate studies at UNC Charlotte. Besides providing me with an effective guidance on my research, his enthusiasm to work also inspired me to push myself for better. This dissertation would not be possible without his guidance and encouragement. He and his family have also offered a lot of help during my whole phd life.

Appreciation also extends to Dr. Yanqing Sun, Dr. Weihua Zhou, and Dr. Aixi Zhou for serving as very valuable members in my advisory committee who gave very precious suggestions and great support on my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Vector time series data are frequently observed in practice. A real example was given in Tsay(1998) for analysis of two daily river flow series of Iceland, where a bivariate thresholding $AR(15)$ model was successfully used. In financial markets, multiple time series are usually correlated. For instance, the yields of three-month, six-month and twelve-month treasury bills (see Example 1.3 in Fan and Yao, 2003), they are highly correlated. When analyzing several interdependent time series, it is natural to consider them as a single vector time series, for example, using the linear or thresholding vector autoregressive model where the current value of the vector time series depends on its past.

As nonlinear features widely exist in time series(Tong and Lim,1980;Tong,1983, 1990;Chen and Tsay,1993;Chen,Liu and Tsay, 1995;Yao and Tong, 1994,1995; Tsay 1989,1998;Fan and Yao,2003), it is important to model the nonlinear features using nonparametric vector time series models, which requires little prior information on the model structure and give some insights into further parametric fitting. However, a full nonparametric method suffers from "curse of dimensionality" in multivariate cases when the dimension is high. These motivate us to propose a multivariate time-varying coefficient regression model for modeling vector time series data. The newly proposed model releases some restrictions on the model structure while avoiding the "curse of dimensionality". Many works have been contributed to modeling nonlinearity in univariate time series using parametric methods(Tong,1990,1995). Successful examples include, but are not limited to, the threshold AR models(Tong,1983;Tsay,1989), the ARCH/GARCH models (Engle,1982; Bollerslev, 1986),and their variants such as the double threshold ARCH model(Li and Li,1996; Hui and Jiang,2005). For parametric modeling of vector time series data, there are many extensions to the above univariate models as well, which gives the well-known vector AR, ARCH and GARCH models(for overview, see Bauwens et al.,2006). Nevertheless, there are infinitely many nonlinear parametric forms needing to explore. Nonparametric techniques give an alternative and some useful insights to nonlinear parametric modeling(Fan and Yao,2003). Therefore, its better for us to model nonlinearity through the nonparametric method for vector time series data.

There is relatively much less work available in the literature for vector time series data since existing nonparametric regression techniques mainly focus on univariate response models. This is partially due to the "curse of dimensionality" and the complexity of smoothing vector time series data. Generally, for modeling vector time series data,we prefer multivariate models than univariate models because univariate models for each time series can not capture the correlation structure of different time series, so they may be inefficient. Moreover, multivariate models provide a convenient tool for modeling interdependencies among multiple time series and hence for simultaneously analyzing feedback effects. Recently Li and Genton (2009) proposed a single-index additive VAR model with constant conditional variance of the error. Motivated by analyzing the aforementioned interest rates, we fit the vector time series with exogenous variables by using multivariate time-varying coefficient models.

1.1 Multivariate Threshold Time Series Model

We begin with the well-known multivariate threshold model(Tsay,1998) to introduce our multivariate nonparametric time series model. Consider at first a k-dimensional time series $\mathbf{y}_i = (y_{1i}, y_{2i}, ..., y_{ki})'$ and v-dimensional exogenous time series $\mathbf{x}_i = (x_{1i}, x_{2i}, ..., x_{vi})'$, where $i = 1, 2, ..., n$. Let $-\infty = r_0 < r_1 < r_2 < ... < r_s = \infty$. The multivariate threshold model with threshold variable z_i and delay d formulates $\mathbf{y}_i$ by

$$\mathbf{y}_i = \mathbf{c}_j + \sum_{t=1}^{p} \alpha_t^{(j)} \mathbf{y}_{i-t} + \sum_{t=1}^{q} \beta_t^{(j)} \mathbf{x}_{i-t} + \mathbf{e}_i^{(j)}, \text{if } r_{j-1} < z_{i-d} < r_j, \tag{1.1}$$

where $j = 1, 2...s$, p and q are nonnegative integers (see Tsay 1998). The innovations satisfy $\mathbf{e}_i^{(j)} = \gamma_j^{1/2} \mathbf{a}_i$, where $\gamma_j^{1/2}$ are symmetric positive definite matrices and $\mathbf{a}_i$ is a sequence of serially uncorrelated random vectors with mean $\mathbf{0}$ and identity covariance matrix $\mathbf{I}_k$. The threshold variable z_i and exogenous time series $\mathbf{x}_i$ are assumed to be strictly stationary with continuous distributions. Model (1.1) is piecewise linear in the threshold space z_{i-d}, but it is nonlinear when $s > 1$(Tsay,1998). These motivate us to work on the spatial QR estimation for the multivariate threshold time series model proposed by Tsay(1998). This model has nice features: It characterizes different regression relationships at different regions of the lagged variable, a nonlinear regression relationship. As a natural extension to Tong's threshold model(see Tong and Lim,1980), the double-threshold structure allow us to capture nonlinear phenomena such as asymmetric cycles, jump resonance and amplitude frequency dependence. No restriction on the form of the error distribution enables robust inference for the model. To the best of my knowledge, there is little solid work about the spatial QR for vector time series data in the literature. The proposed spatial QR methodology is vary useful in detecting the nonlinear dependence on the covariates in the lower and upper tails, as well as in the central, of the vector time series.

1.2 Multivariate Time-varying Coefficient Time Series Model

In practice, there are many successful examples applying multivariate parametric models such as (1.1). See Tsay(1998) and the references therein. However, the assumption for the threshold model (1.1) that the coefficients are usually evolving or changing slowly through time and the coefficient functions may vary smoothly. This reveals that the coefficients might be functions of time. As a matter of fact, model (1.1) is a special case when the coefficients are piecewise constant functions of z_{t-d}. This motivates us to propose the multivariate time-varying coefficient model:

$$\mathbf{y}_t = \mathbf{c}(t/T) + \sum_{i=1}^{p} \boldsymbol{\alpha}_i(t/T)\mathbf{y}_{t-i} + \sum_{j=1}^{q} \boldsymbol{\beta}_j(t/T)\mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t, t = 1, ..., T. \tag{1.2}$$

where $\mathbf{y}_t$ is $k \times 1$ vector, $\mathbf{x}_t$ is $v \times 1$ vector. $\mathbf{c}(.)$ is a $k \times 1$ vector, $\boldsymbol{\alpha}_i$ is $k \times k$ smooth matrix and $\boldsymbol{\beta}_j$ is $k \times v$ smooth matrix. The innovations satisfy $\boldsymbol{\varepsilon}_t = \gamma_t^* \mathbf{a}_t$, where γ_t^* are symmetric positive definite matrices and $\mathbf{a}_t$ is a sequence of uncorrelated random vectors with mean zero and convariance matrix $\mathbf{I}_k$. For model (1.2), we are interested in estimating the regression part. In addition, I develop a new test procedure and propose new test statistics to perform simultaneous inference about the parameters.

1.3 Generalized Likelihood Ratios

It is well known that likelihood ratio theory is one of the most important statistic results and it develops a useful principle that generally applicable to most parametric hypothesis problems. A key fundamental property that contributes very significantly to the success of the maximum likelihood ratio tests is that their asymptotic null distribution are independent of nuisance parameters. Many computationally intensive nonparametric techniques and theories have been rapidly developed to exploit hidden structures and to reduce modelling biases of traditional parametric methods. Methods such as local linear fitting, local polynomial fitting, orthogonal series expansions and spline approximations, also dimensionality reduction techniques have been studied in great details in many statistical contexts. Yet, there are no generally applicable methods available for the inferences in multivariate nonparametric models. Owen(1988) extending the scope of the likelihood inferences through the empirical likelihood which is applicable to a class of nonparametric functionals. Usually, these functionals are smooth that they can be estimated at root-n rate. See also Owen (1990), Hall and Owen (1993),Chen and Qin (1993), Li, Hollander, Mckeague and Yang(1996) for applications of the empirical likelihood. A further extension of the empirical likelihood, called "random-sieve likelihood", can be found in Shen, Shi and Wong(1999). The random-sieve likelihood allows one to handle the situations where observable variables and stochastic errors are not necessarily one-on-one. In addition, various efforts have been put on nonparametric hypothesis testing. For instance, see, Bickel and Ritov (1992),Fan(1996), Fan and Li(1996), Kallenberg and Ledwina (1997). However, most of the studies focus only on the one-dimensional nonparametric regression problem. It is difficult to extend them to multivariate semiparametric and nonparametric models. In order to derive a generally applicable testing procedure for multivariate semiparametric and nonparametric models. Fan, Zhang and Zhang (2001) proposed generalized likelihood ratio tests. The work is motivated by the fact that the nonparametric maximum likelihood ratio test may not exist in many nonparametric problems. Generalized likelihood ratio statistics, obtained by replacing unknown functions by reasonable nonparametric estimators have several nice properties. For instance additive models (Fan and Jiang 2005):

$$\mathbf{Y} = m_1(\mathbf{X}_1) + \ldots + m_p(\mathbf{X}_p) + \boldsymbol{\varepsilon} \tag{1.3}$$

or time-varying coefficient models(Dr Hoover 1998):

$$\mathbf{Y} = a_1(t/T)\mathbf{X}_1 + \ldots + a_p(t/T)\mathbf{X}_p + \boldsymbol{\varepsilon} \tag{1.4}$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_p$ are covariates. One would ask if certain parametric forms such as linear models fit the data adequately, after fitting the model. This means testing if each additive component is linear in the additive model (1.3) or if the coefficient functions in (1.4) are really time-varying or not.

Let us begin with a simple nonparametric regression model to motivate the generalized likelihood ratio statistics. Suppose we have n sample data $\{X_i, Y_i\}$ from the nonparametric regression model, for $i = 1, ...n$,

$$Y_i = m(X_i) + \varepsilon_i \tag{1.5}$$

where $\{\varepsilon_i\}$ are a sequence of i.i.d. random variables from $N(0, \sigma^2)$ and X_i has a density f with support $[0, 1]$. Denote the parameter space is

$$\mathcal{F}_k = \{m \in L^2[0,1] : \int m^{(k)}(x)^2 dx \leqslant C\} \tag{1.6}$$

for a given C. Consider the testing problem:

$$H_0 : m(x) = \alpha_0 + \alpha_1 x \longleftrightarrow H_1 : m(x) \neq \alpha_0 + \alpha_1 x \tag{1.7}$$

Hence, the conditional log-likelihood function is: $l_n(m) = -nlog(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - m(X_i))^2$. Let $(\hat{\alpha}_0, \hat{\alpha}_1)$ be the maximum likelihood estimator (MLE) under H_0, and $\hat{m}_{MLE}(\centerdot)$ be the MLE under the full model: $min \sum_{i=1}^n (Y_i - m(X_i))^2$, subject to $\int m^{(k)}(x)^2 dx \leqslant C$. The resulting estimator $\hat{m}_{MLE}$ is a smoothing spline. Define the residual sum of squares RSS_0 and RSS_1 as follows:

$$RSS_0 = \sum_{i=1}^n (Y_i - \hat{\alpha}_0 - \hat{\alpha}_1 X_i)^2, RSS_1 = \sum_{i=1}^n (Y_i - \hat{m}_{MLE}(X_i))^2. \tag{1.8}$$

So it is easy to see that the logarithm of the conditional maximum likelihood ratio statistic for the problem (1.7) is given by:

$$\lambda_n = l_n(\hat{m}_{MLE}) - l_n(H_0) = \frac{n}{2} log \frac{RSS_0}{RSS_1} \approx \frac{n}{2} \frac{RSS_0 - RSS_1}{RSS_1}$$

Technically, the maximum likelihood ratio test is not convenient to manipulate and is either not optimal due to restriction of choosing smoothing parameters. In general, MLEs under nonparametric regression models are hard to obtain. Therefore, we replace the maximum likelihood estimator under the alternative nonparametric model by any reasonable nonparametric estimator, giving the generalized likelihood ratio

$$\lambda_n = l_n(H_1) - l_n(H_0) \tag{1.9}$$

Here $l_n(H_1)$ is the likelihood with unknown regression function replaced by a reasonable nonparametric regression estimator. We can find similar ideas in Severini and Wong (1992) for construction of semi-parametric efficient estimators. We notice that the nonparametric estimator does not have to belong to $\mathcal{F}_k$. Thus the assumption that the constant C in (1.6) is given can be removed. The above generalized likelihood method can be readily be applied to other statistical methods such as additive models, varying-coefficient models, and any nonparametric regression model with a parametric regression model with a parametric error distribution. Using suitable nonparametric estimators, we need to compute the likelihood function under null and alternative models. The generalized likelihood ratio tests are expected to be powerful with appropriate choice of smoothing parameters.

1.4 Outline of the Dissertation

The rest of this dissertation is organized as follows.

In Chapter 2, consistency and asymptotic normality of our estimator are established, and an algorithm is suggested to compute the proposed estimates. The performance of our estimator is evaluated via finite sample simulations. A real example is presented to illustrate the use of the proposed methodology.

In Chapter 3, I discuss the estimation of coefficients in a time-varying coefficient multivariate regression model by using local linear technique and then derive the explicit expression of the proposed estimator. Then, I derive the asymptotic theory for the nonparametric estimator. At last, I propose the new GLR test for the multivariate time-varying coefficient model to test if varying coefficients for the time-varying nonparametric regression model are some known constants or of some specific time-varying functional forms. The test statistics are constructed based on the comparison of the likelihood under null and alternative hypotheses respectively. I derive the asymptotic distributions of the test statistic under null and alternative hypotheses. In addition, Monte Carlo simulation is done to show the finite sample performance of the proposed methods, power curves are presented for different error distributions and different sample sizes. Finally an real example of monthly US interest rate is given for the application of the methodology.

Chapter 4 concludes the dissertation. The detailed proofs of the main results in each chapter are relegated to the last section of the corresponding chapter.

CHAPTER 2: MULTIVARIATE THRESHOLD TIME SERIES MODEL

There is a rich literature on quantile regression (QR) in the analysis of time series, examples include but not limit to Koul and Saleh (1995), Davis and Dunsmuir (1997), Jiang, Zhao and Hui (2001), Peng and Yao (2003), etc. However, all of them are restricted to univariate cases. For vector time series, to the best of our knowledge, there is little solid mathematical theory on QR in the literature, although much work has been contributed using the maximum likelihood or least squares estimation. See for example Bollerslev (1990), Engle and Kroner (1995), Chen and Tsay (1993), Pan and Yao (2008), and the references therein. A main difficulty with the multivariate QR is about how to define a multivariate quantile, and an additional difficulty is to deal with the aurocorrelation in vector time series in the QR setting.

Based on the L1-norm, Chaudhuri (1996) and Koltchinskii (1997) proposed a compelling form of multivariate quantiles, the spatial quantiles, as a certain form of generalization of the univariate case.This kind of multivariate 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 Sering (2004), it also has some appealing features: the equivariance and outlyingness with respect to shift, orthogonal, and homogeneous scale transformations. These motivate us to work on the spatial QR estimation for the multivariate threshold time series model proposed by Tsay (1998).

2.1 Spatial QR for Multivariate Threshold Time Series Model

Recall for model (1.1), consider at first a k-dimensional time series $\mathbf{y}_i = (y_{1i}, y_{2i}, ..., y_{ki})'$ and v-dimensional exogenous time series $\mathbf{x}_i = (x_{1i}, x_{2i}, ..., x_{vi})'$, where $i =$

$1, 2, ..., n$. Let $-\infty = r_0 < r_1 < r_2 < ... < r_s = \infty$. The multivariate threshold model with threshold variable z_i and delay d formulates $\mathbf{y}_i$ by

$$\mathbf{y}_i = \mathbf{c}_j + \sum_{t=1}^{p} \alpha_t^{(j)} \mathbf{y}_{i-t} + \sum_{t=1}^{q} \beta_t^{(j)} \mathbf{x}_{i-t} + \mathbf{e}_i^{(j)} \tag{2.1}$$

if $r_{j-1} < z_{i-d} < r_j$, where $j = 1, 2...s$, p and q are nonnegative integers (see Tsay 1998). The innovations satisfy $\mathbf{e}_i^{(j)} = \gamma_j^{1/2} \mathbf{a}_i$, where $\gamma_j^{1/2}$ are symmetric positive definite matrices and $\mathbf{a}_i$ is a sequence of serially uncorrelated random vectors with mean $\mathbf{0}$ and identity covariance matrix $\mathbf{I}_k$. The threshold variable z_i and exogenous time series $\mathbf{x}_i$ are assumed to be strictly stationary with continuous distributions.

Let $I_{i,j} = I(r_{j-1} < z_{i-d} \leq r_j)$, $\mathbf{X}_i = vec(1, \mathbf{y}_{i-1}, ..., \mathbf{y}_{i-p}, \mathbf{x}_{i-1}, ..., \mathbf{x}_{i-q})$, $\mathbf{Z}_i = (I_{i1}, I_{i2}, ..., I_{is})' \otimes \mathbf{X}_i$, $\mathbf{e}_i = \sum_{j=1}^{s} I_{i,j} \mathbf{e}_i^{(j)}$, $\mathbf{\Phi}_j = (\mathbf{c}_j, \alpha_1^{(j)}, ..., \alpha_p^{(j)}, \beta_1^{(j)}, ..., \beta_q^{(j)})$, and $\mathbf{\Phi} = (\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_s)$. Then $\mathbf{\Phi}$ is a $k \times s(1+kp+vq)$ matrix, and model (2.1) becomes

$$\mathbf{y}_i = \mathbf{\Phi} \mathbf{Z}_i + \mathbf{e}_i \tag{2.2}$$

Following Chaudhuri (1996), for any $\mathbf{u}$ in the unit ball $\mathbb{B}^k$ centered at zero in $\mathbb{R}^k$, we define the $\mathbf{u}$-th spatial QR estimators of the parameters by

$$(\hat{\mathbf{\Phi}}_n(\mathbf{u}), \hat{\mathbf{e}}_\mathbf{u}) = \arg \min_{\mathbf{\Phi}, \mathbf{e}_\mathbf{u}} \sum_{i=s^*+1}^{n} \boldsymbol{\rho}_\mathbf{u}(\mathbf{y}_i - \mathbf{\Phi} \mathbf{Z}_i - \mathbf{e}_\mathbf{u}), \tag{2.3}$$

where $s^* = \max(p, q)$, $\boldsymbol{\rho}_\mathbf{u}(\mathbf{t}) = \|\mathbf{t}\| + \mathbf{u}^T \mathbf{t}$, and $\mathbf{e}_\mathbf{u}$ is the $\mathbf{u}$-th quantile of $\mathbf{e}$. Then $\hat{\mathbf{\Phi}}_n(\mathbf{u}) \mathbf{Z}_i + \hat{\mathbf{e}}_\mathbf{u}$ is the spatial QR estimate of the $\mathbf{u}$-th quantile of $\mathbf{y}_i$ conditional on $\mathbf{Z}_i$, and $\hat{\mathbf{\Phi}}_n(\mathbf{u})$ is the spatial QR estimate of $\mathbf{\Phi}$.

2.2 Asymptotic Normality

Let us begin with introducing some notations and definitions. For any $\mathbf{t} \in \mathbb{R}^k$, define $\boldsymbol{\varphi}_\mathbf{u}(\mathbf{t}) = \mathbf{t}/\|\mathbf{t}\| + \mathbf{u}$ for $\mathbf{t} \neq \mathbf{0}$ and $\boldsymbol{\varphi}_\mathbf{u}(\mathbf{0}) = \mathbf{u}$. Let $\mathbf{\Psi}(\mathbf{t})$ denote the $k \times k$

Hessian matrix, so for $\mathbf{t} \neq \mathbf{0}$,

$$\mathbf{\Psi}(\mathbf{t}) = \|\mathbf{t}\|^{-1}(\mathbf{I}_k - \mathbf{t}\mathbf{t}^T\|\mathbf{t}\|^{-2}),$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix. We will adopt the convention that $\mathbf{\Psi}(\mathbf{0}) = \mathbf{0}$, the zero matrix.

Let the marginal density of the error vectors $\mathbf{e}_i$'s be $h(\cdot)$. Define $\mathbf{Q}(\mathbf{u}) = \arg\min_{\mathbf{Q}\in\mathbb{R}^k} E[\boldsymbol{\rho}_{\mathbf{u}}(\mathbf{e}_1 - \mathbf{Q}) - \boldsymbol{\rho}_{\mathbf{u}}(\mathbf{e}_1)]$, the $\mathbf{u}$-th quantile of the error vector $\mathbf{e}_1$. Denote by $D_1(\mathbf{u}) = E[\mathbf{\Psi}(\mathbf{e}_1 - \mathbf{Q}(\mathbf{u}))]$, and $D_2(\mathbf{u}) = E[\boldsymbol{\varphi}_{\mathbf{u}}(\mathbf{e}_1 - \mathbf{Q}(\mathbf{u}))\{\boldsymbol{\varphi}_{\mathbf{u}}(\mathbf{e}_1 - \mathbf{Q}(\mathbf{u}))\}^T]$. To derive asymptotic distributions of the spatial QR estimators, we need the following assumptions:

Assumption A:

(A1) $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i^T = \mathbf{S}$, where $\mathbf{S}$ is a positive definite matrix.

(A2) The processes $\{\mathbf{x}_t, \mathbf{y}_t\}$ are strictly stationary with α-mixing coefficients $\alpha(k)$ such that $\sum_k k^c[\alpha(k)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$.

(A3) There exists a positive $\gamma > 0$ such that $E\|\mathbf{e}_t\|^{2+\gamma} < \infty$.

Theorem 2.1 Assume that the Assumptions (A1)-(A3) hold. Then

$$\sqrt{n}(\hat{\mathbf{\Phi}}_n(\mathbf{u}) - \tilde{\mathbf{\Phi}}(\mathbf{u})) \xrightarrow{\mathcal{D}} N(0, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = \{[D_1(\mathbf{u})]^{-1}[D_2(\mathbf{u})][D_1(\mathbf{u})]^{-1}\}\otimes\mathbf{S}^{-1}$ and $\tilde{\mathbf{\Phi}}(\mathbf{u}) = \mathbf{\Phi} + [\mathbf{Q}(\mathbf{u}), 0, ..., 0]_{k\times s(1+kp+vq)}$

Proof: See section 2.5.

By the definition of $\tilde{\mathbf{\Phi}}(\mathbf{u})$, we can partition it as $\tilde{\mathbf{\Phi}}(\mathbf{u}) = (\tilde{\mathbf{\Phi}}_1(\mathbf{u}), \tilde{\mathbf{\Phi}}_2(\mathbf{u}))$, where $\tilde{\mathbf{\Phi}}_1(\mathbf{u})$ is the first column, and $\tilde{\mathbf{\Phi}}_2(\mathbf{u})$ are the remaining columns which do not depend on $\mathbf{u}$. We also partition $\hat{\mathbf{\Phi}}_n(\mathbf{u})$ as $\hat{\mathbf{\Phi}}_n(\mathbf{u}) = (\hat{\mathbf{\Phi}}_{1n}(\mathbf{u}), \hat{\mathbf{\Phi}}_{2n}(\mathbf{u})$ in the same way as that for $\tilde{\mathbf{\Phi}}(\mathbf{u})$. Then, by Theorem 2.1, $\hat{\mathbf{\Phi}}_{2n}(\mathbf{u})$ are always consistent estimators of $\tilde{\mathbf{\Phi}}_2$ for different $\mathbf{u}$. Since averaging can reduce the variance of estimator, one may use the following weighted estimator

$$\hat{\mathbf{\Phi}}_{2\boldsymbol{\omega}} = \sum_{k=1}^{K} \omega_k \hat{\mathbf{\Phi}}_{2n}(\mathbf{u}_k),$$

where $\sum_{k=1}^{K} \omega_k = 1$. In practice, given a specific value of K, one may use equally-spaced $\{\mathbf{u}_k\}$ in the unit ball centered at zero in R^k. For simplicity, one may also employ $\omega_k = 1/K$. Let $D_3(\mathbf{u}_k, \mathbf{u}_{k'}) = E[\boldsymbol{\varphi}_{\mathbf{u}_k}(\mathbf{e}_1 - \mathbf{Q}(\mathbf{u}_k))\{\boldsymbol{\varphi}_{\mathbf{u}_{k'}}(\mathbf{e}_1 - \mathbf{Q}(\mathbf{u}_{k'}))\}^T]$. Then:

Theorem 2.2 Assume that the Assumptions (A1)-(A3) hold. Then

$$\sqrt{n}(\hat{\mathbf{\Phi}}_{2\boldsymbol{\omega}} - \tilde{\mathbf{\Phi}}_2) \xrightarrow{\mathcal{D}} N(0, \mathbf{\Omega}(\omega)),$$

where $\mathbf{\Omega}(\omega) = \sum_{k,k'=1}^{K} \omega_k \omega_{k'} \{[D_1(\mathbf{u}_k)]^{-1} D_3(\mathbf{u}_k, \mathbf{u}_{k'})[D_1(\mathbf{u}_{k'})]^{-1}\} \otimes \mathbf{S}^{-1}$. Proof: See section 2.5.

Let D be a $k \times k$ matrix with the (k, k')th entry being $D_{kk'} = \text{tr}[D_1(\mathbf{u}_k)]^{-1} D_3(\mathbf{u}_k, \mathbf{u}_{k'})[D_1(\mathbf{u}_{k'})]^{-1}$. Then, under the constraint $\sum_{k=1}^{K} \omega_k = 1$, minimizing $\text{tr}[\mathbf{\Omega}(\omega)]$ over ω is equivalent to minimizing $\sum_{k,k'=1}^{K} \omega_k \omega_{k'} D_{kk'}$ over ω. The minimizer is $\omega_{opt} = (\mathbf{1}^T D^{-2} \mathbf{1})^{-1/2} D^{-1} \mathbf{1}$, and the corresponding minimum is $(\mathbf{1}^T D^{-1} \mathbf{1})^{-1}$, where $\mathbf{1} = (1, \ldots, 1)^T$ is a $K \times 1$ vector of all entries being ones.

2.3 Simulations

We conduct a 500 Monte Carlo time simulation study to demonstrate the performance of our QR estimator: $\hat{\mathbf{\Phi}}_{2n}(\mathbf{u})$ (QR). We set the dimension of $\mathbf{y}$, k=2, the dimension of $\mathbf{x}$, v=1. we set two different sample sizes n=1800 and n=3600. We chose $\mathbf{u} = [0.0, 0.0]$ for the $\hat{\mathbf{\Phi}}_{2n}(\mathbf{u})$. In model (2.1): $\mathbf{y}_i = \mathbf{c}_j + \sum_{t=1}^{p} \alpha_t^{(j)} \mathbf{y}_{i-t} + \sum_{t=1}^{q} \beta_t^{(j)} \mathbf{x}_{i-t} + \mathbf{e}_i^{(j)}$, we have taken $s = 2, p = 2, q = 1$. We set $\mathbf{c} = (c_1, c_2) = (0, 0)$, the first two lagged $\mathbf{y}_1 = (0.15, 0.2)$, $\mathbf{y}_2 = (-0.25, -0.04)$. We let $\mathbf{x}_i$ be a stationary AR(1) time series such that $\mathbf{x}_i = 0.5\mathbf{x}_{i-1} + \boldsymbol{\varepsilon}_i$ with $E(\boldsymbol{\varepsilon}_i) = 0$ and the threshold variable $\mathbf{z}_{i-1} = \mathbf{x}_{i-1}$ We set the threshold point $r_0 = 0.5$, we have generated two different errors as $\mathbf{e} = \mathbf{I}_{(\mathbf{z}<r_0)}\mathbf{e}_1 + \mathbf{I}_{(\mathbf{z}\geq r_0)}\mathbf{e}_2$ with $\mathbf{e}_1 \sim \mathbf{N}(\mathbf{0}, \mathbf{V}_1)$, $\mathbf{e}_2 \sim \mathbf{N}(\mathbf{0}, \mathbf{V}_2)$. Here $\mathbf{V}_1 = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$, $\mathbf{V}_2 = \left(\begin{smallmatrix} 1 & 0.25 \\ 0.25 & 1 \end{smallmatrix}\right)$ for the two-dimensional normal error; $\mathbf{V}_1 = \left(\begin{smallmatrix} 3 & 0.25 \\ 0.25 & 3 \end{smallmatrix}\right)$, $\mathbf{V}_2 = \left(\begin{smallmatrix} 3 & 0.75 \\ 0.75 & 3 \end{smallmatrix}\right)$ for the two-dimensional t-distribution error with degrees of freedom 3. Also, we included additive outliers with an artificial magnitude of 15 added to the values of $\mathbf{y}$ at three time points in the middle using 2-dimensional standard normal error. Finally, we let $\alpha_1^{(1)} = 0.075$, $\alpha_1^{(2)} = 0.15$, $\alpha_2^{(1)} = -0.1$, $\alpha_2^{(2)} = -0.075$ and $\beta_1^{(1)} = 0.025$, $\beta_1^{(2)} = 0.1$. Then we have $ks(kp + vq) = 20$ components of the parameter estimates,using these $\mathbf{e}_i$, $\mathbf{x}_1$, $\mathbf{y}_1$ and $\mathbf{y}_2$ we have generated the observations $(\mathbf{Z}_i, \mathbf{y}_i)$ for $i = 1, 2, ..., n$. Our estimate is initiated by the ordinary least square estimate (LSE).

Figure 2.1 to Figure 2.3 are Box-plots showing the comparison between $\hat{\mathbf{\Phi}}_{2n}^j(\mathbf{u}) - \tilde{\mathbf{\Phi}}^j$ and $LSE - \tilde{\mathbf{\Phi}}^j$, $j = 1, 2..., 20$. Figure 2.4 represents the 95% coverage probability of the estimates with the corresponding different error distributions calculated from sample mean of $\{\hat{\mathbf{\Phi}}_{2n}^j\} - 1.96 \cdot std(\hat{\mathbf{\Phi}}_{2n}^j)/\sqrt{n}$ to sample mean of $\{\hat{\mathbf{\Phi}}_{2n}^j\} + 1.96 \cdot std(\hat{\mathbf{\Phi}}_{2n}^j)/\sqrt{n}$. Here, $std(\hat{\mathbf{\Phi}}_{2n}^j)$ are the square roots of the diagonal elements of the estimate matrix of $\mathbf{\Sigma}$, i.e. the estimator of $\{[D_1(\mathbf{u})]^{-1} D_2(\mathbf{u})[D_1(\mathbf{u})]^{-1}\} \otimes \mathbf{S}^{-1}$. $i = 1, 2, ..., n$, the estimator of $D_1(\mathbf{u})$ is calculated as the sample mean of $\mathbf{\Psi}(\mathbf{e}_i - \mathbf{Q}(\mathbf{u}))$, $D_2(\mathbf{u})$ is estimated by the sample mean of $\boldsymbol{\varphi}_{\mathbf{u}}(\mathbf{e}_i - \mathbf{Q}(\mathbf{u}))\{\boldsymbol{\varphi}_{\mathbf{u}}(\mathbf{e}_i - \mathbf{Q}(\mathbf{u}))\}^T$.

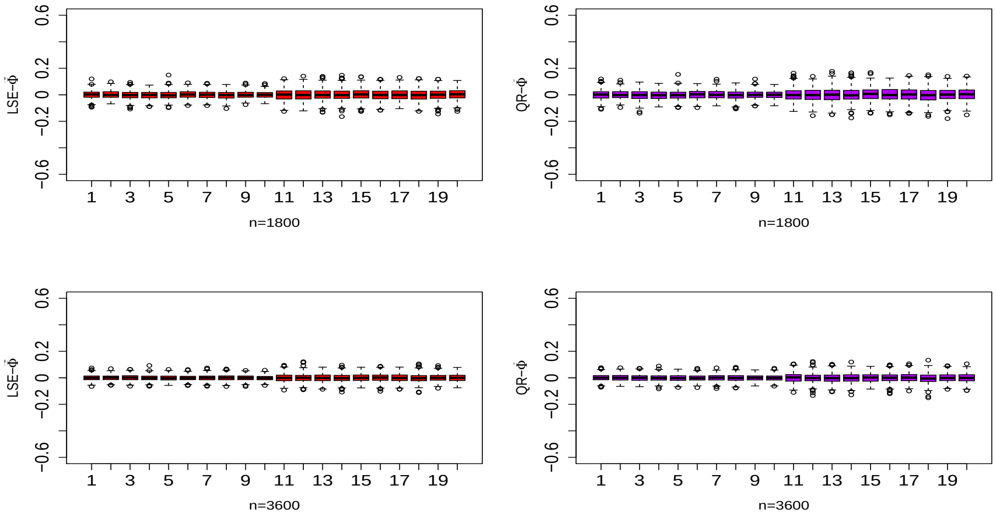

Figure 2.1 : Comparison between $LSE - \tilde{\mathbf{\Phi}}$(left) and $QR - \tilde{\mathbf{\Phi}}$(right) for Normal error when sample size n=1800 and n=3600 .

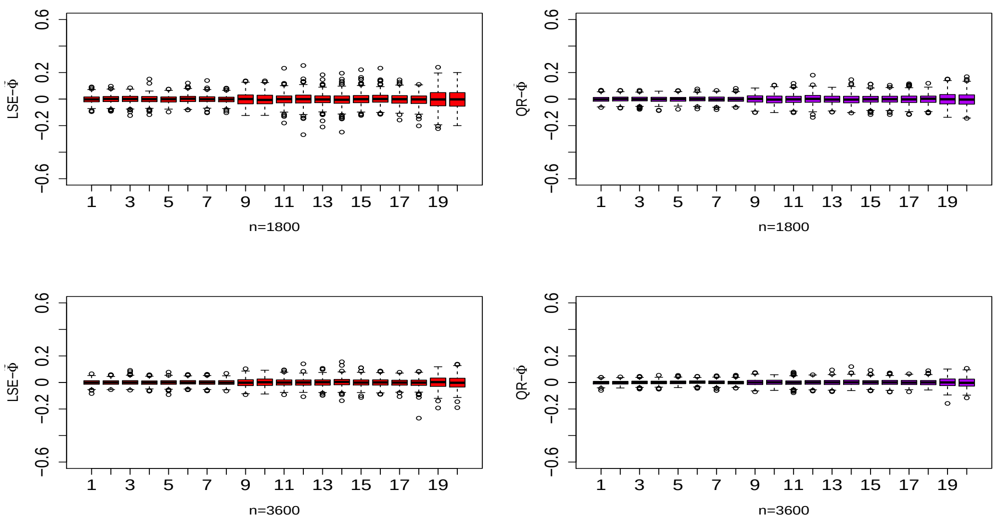

Figure 2.2 : Comparison between $LSE - \tilde{\mathbf{\Phi}}$(left) and $QR - \tilde{\mathbf{\Phi}}$(right) for t(3) error when sample size n=1800 and n=3600.

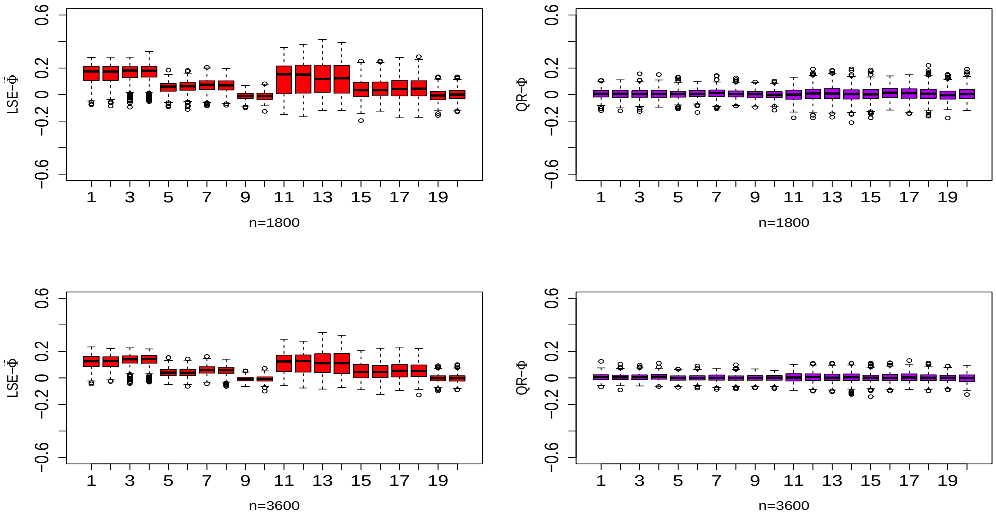

Figure 2.3 : Comparison between $LSE - \tilde{\mathbf{\Phi}}$(left) and $QR - \tilde{\mathbf{\Phi}}$(right) for additive error when sample size n=1800 and n=3600.

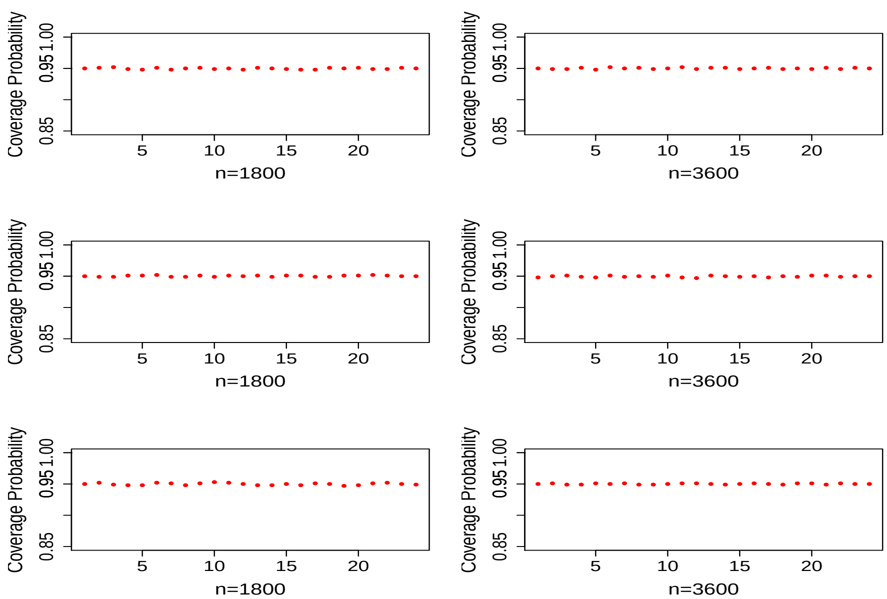

Figure 2.4 : 95%-Coverage probability of QR for Normal, t(3) errors and additive outliers with different sample sizes.

2.4 A real example

To illustrate the use of spatial QR, we use Iceland river flow dataset (Tong, 1990). The data are daily observations from January 1, 1972 to December 31, 1974 on 4 variables. The dependent variables are $y_t = (y_{1t}, y_{2t})'$, where y_{1t} is the daily flow of Vatnsdalsá river and y_{2t} is the daily flow of Jökulsá Eystri river. The exogeneous variables are daily precipitation (x_t) and temperature (z_t) observed in Hveravellir. The threshold variable is z_t. We aim to investigate how the daily mean flow of both rivers change over time using a bivariate two regime TAR(15) model selected from Tsay (1998). Hence we focus on the effect from the first 15 lags of response variables and the first 3 lags of the exogenous variables. So in this case, n=1095, k=v=2, the number of the lags of $\mathbf{y}_t$; p=15, the number of the lags of $(\mathbf{x}_t, \mathbf{z}_t)$; q=3, and the number of regions s=2. We set threshold temperature $r_0 = 0$ and $\mathbf{c} = (0, 0)$. Similarly as model (2.1), we have: $\mathbf{y}_t = \sum_{i=1}^{15} \alpha_i^{(j)} \mathbf{y}_{t-i} + \sum_{i=1}^{3} \beta_i^{(j)} \mathbf{x}_{t-i} + \mathbf{e}_t^{(j)}$, for $j = 1, 2$.

Our estimators are computed in order to minimize $\sum_{t=16}^{n} \boldsymbol{\rho}_{\mathbf{u}}(\mathbf{y}_t - \boldsymbol{\Phi}\mathbf{Z}_t)$ at a given $\mathbf{u}$ with $\mathbf{Z}_t = (I_{t1}, I_{t2})' \otimes [vec(\mathbf{x}_t, \mathbf{z}_t)]$, indicator vector $(I_{t1}, I_{t2}) = (I(\mathbf{z}_t < r_0), I(\mathbf{z}_t \geq r_0))$ using the proposed algorithm and compare with the least square estimates. Figure 2.5 shows that the mean flow of the two rivers are highly correlated so we have to employ a bivariate threshold model to perform the estimations of the coefficients instead of one-dimensional estimation. Figure 2.6 shows the conditional quantiles of the daily flows of both rivers at different times. It occurs to both rivers that the daily flow changed over time and they change in the same pattern even at different quantiles. Table 2.1 to Table 2.3 show our QR estimators and their estimated standard deviations using a bivariate two regime TAR(15) model selected from Tsay (1998) at three given different quantiles.

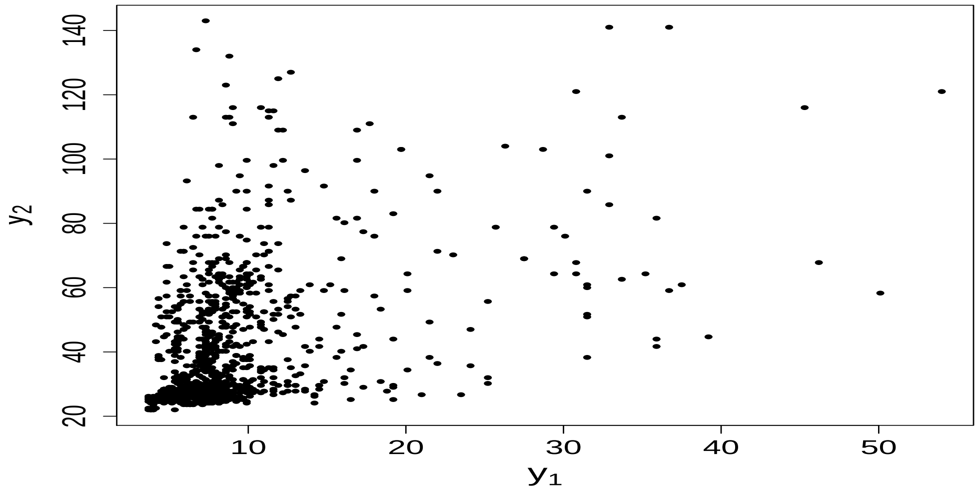

Figure 2.5 : The scatter plots showing the mean flow of Jokulsa river versus the mean flow of Vatnsdalsa river.

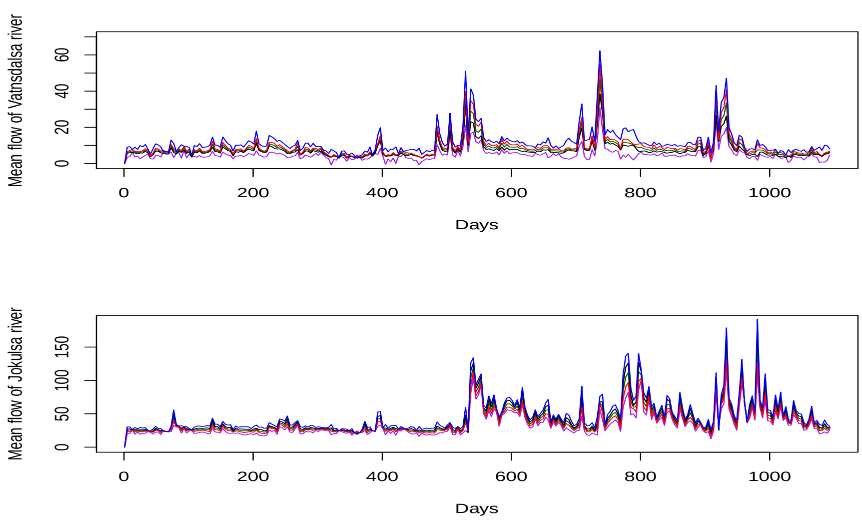

Figure 2.6 : The time series of mean daily flows of Vatnsdalsa and Jokulsa river at different quantiles.($\mathbf{u} = [0.0, 0.0]$ (green), $\mathbf{u} = [-0.9, -0.9]$ (purple),$\mathbf{u} = [-0.5, 0.5]$ (black),$\mathbf{u} = [0.5, -0.5]$ (red) and $\mathbf{u} = [0.9, 0.9]$ (blue).

Table 2.1 : QR estimates and their estimated standard deviations for a selected bivariate two-regime TAR(15) model for the Iceland river flow data at quantile $\mathbf{u} = (0, 0)$. The threshold value is $0^o C$, and the numbers of observations in each regime are 479 and 601, respectively.

	Regime 1				Regime 2			
	y_{1t}		y_{2t}		y_{1t}		y_{2t}	
	Coef.	*Std*	*Coef.*	*Std*	*Coef.*	*Std*	*Coef.*	*Std*
$y_{1,t-1}$	0.94	0.04	0.98	0.02	0.98	0.11	1.17	0.06
$y_{1,t-2}$	0.47	0.05	0.46	0.01	0.35	0.13	0.59	0.08
$y_{1,t-3}$	-0.48	0.05	-0.56	0.02	-0.22	0.16	-0.40	0.07
$y_{1,t-4}$	-0.33	0.04	-0.32	0.01	-0.09	0.06	0.54	0.08
$y_{1,t-5}$	0.13	0.05	0.22	0.02	0.28	0.14	0.25	0.02
$y_{1,t-6}$	0.06	0.02	-0.41	0.03	0.08	0.06	-0.45	0.04
$y_{1,t-7}$	-0.02	0.04	-0.55	0.03	-0.09	0.13	-0.38	0.03
$y_{1,t-8}$	-0.02	0.02	0.05	0.01	0.22	0.06	-0.13	0.01
$y_{1,t-9}$	0.21	0.05	0.80	0.02	0.06	0.02		
$y_{1,t-10}$	-0.03	0.01	0.62	0.03				
$y_{1,t-11}$	0.05	0.01	-0.12	0.01				
$y_{1,t-12}$	-0.02	0.02	0.15	0.01				
$y_{1,t-13}$	-0.07	0.04						
$y_{1,t-14}$	0.02	0.01						
$y_{1,t-15}$	-0.03	0.01						
$y_{2,t-1}$	0.16	0.04	0.78	0.03	1.87	0.08	0.66	0.07
$y_{2,t-2}$	0.10	0.04	-0.52	0.03	-1.12	0.09	-0.49	0.08
$y_{2,t-3}$	0.08	0.03	0.08	0.01	-0.11	0.13	0.25	0.04
$y_{2,t-4}$	0.15	0.04	0.11	0.01	-0.21	0.12	-0.37	0.08
$y_{2,t-5}$	-0.03	0.01	-0.40	0.02	0.04	0.04	-0.42	0.04
$y_{2,t-6}$	-0.05	0.04	0.13	0.01	0.17	0.13	-0.51	0.08
$y_{2,t-7}$	0.08	0.02	0.18	0.01	0.43	0.04	-0.42	0.08
$y_{2,t-8}$	0.07	0.04	-0.12	0.03	-0.20	0.02	0.04	0.01
$y_{2,t-9}$	0.02	0.02	0.14	0.01	0.13	0.02	0.10	0.03
$y_{2,t-10}$	0.05	0.04	-0.02	0.03				
$y_{2,t-11}$	-0.02	0.01	-0.07	0.01				
$y_{2,t-12}$	0.04	0.02	-0.12	0.03				
$y_{2,t-13}$			-0.04	0.01				
$y_{2,t-14}$			0.21	0.02				
$y_{2,t-15}$			0.11	0.01				
x_{t-1}	0.05	0.02	0.30	0.04	0.10	0.06	0.19	0.03
x_{t-2}	-0.02	0.01	0.06	0.03	0.01	0.04	0.17	0.11
x_{t-3}	-0.03	0.01	0.03	0.01	-0.03	0.03	0.05	0.03
z_{t-1}	0.07	0.02	0.30	0.04	-0.10	0.06	-0.18	0.06
z_{t-2}	-0.03	0.01	-0.25	0.01	0.03	0.03	-0.05	0.03
z_{t-3}			0.19	0.03	-0.06	0.01		

Table 2.2 : QR estimates and their estimated standard deviations for a selected bivariate two-regime TAR(15) model for the Iceland river flow data at quantile $\mathbf{u} = (-0.9, -0.9)$. The threshold value is 0^oC, and the numbers of observations in each regime are 479 and 601, respectively.

	Regime 1				Regime 2			
	y_{1t}		y_{2t}		y_{1t}		y_{2t}	
	$Coef.$	Std	$Coef.$	Std	$Coef.$	Std	$Coef.$	Std
$y_{1,t-1}$	0.59	0.22	-1.56	0.27	0.57	0.36	-1.28	0.16
$y_{1,t-2}$	-0.84	0.38	-1.19	0.30	0.37	0.19	-0.79	0.18
$y_{1,t-3}$	-0.83	0.39	1.00	0.28	-0.56	0.21	-0.60	0.16
$y_{1,t-4}$	-0.02	0.23	0.42	0.09	-0.11	0.03	0.64	0.27
$y_{1,t-5}$	0.23	012	0.47	0.28	0.41	0.08	0.48	0.28
$y_{1,t-6}$	0.14	0.22	0.36	0.13	0.15	0.46	0.76	0.29
$y_{1,t-7}$	0.16	0.09	-0.48	0.29	-0.11	0.02	-0.28	0.11
$y_{1,t-8}$	-0.08	0.01	-0.08	0.14	0.08	0.03		
$y_{1,t-9}$	0.32	0.16	0.22	0.13				
$y_{1,t-10}$	-0.11	0.22	0.35	0.13				
$y_{1,t-11}$	0.30	0.53	0.38	0.19				
$y_{1,t-12}$	0.07	0.22						
$y_{1,t-13}$	0.29	0.49						
$y_{1,t-14}$	0.19	0.06						
$y_{1,t-15}$								
$y_{2,t-1}$	0.46	0.18	-1.32	0.20	0.90	0.16	3.41	0.61
$y_{2,t-2}$	0.35	0.19	0.95	0.26	0.49	0.11	-2.54	0.66
$y_{2,t-3}$	0.36	0.16	0.91	0.15	-0.46	0.26	0.27	0.31
$y_{2,t-4}$	0.32	0.23	0.65	0.31	-0.43	0.13	-1.64	0.33
$y_{2,t-5}$			-0.49	0.16	0.05	0.34	0.59	0.24
$y_{2,t-6}$			0.38	0.10			0.45	0.26
$y_{2,t-7}$			0.21	0.04			0.37	0.30
$y_{2,t-8}$			0.38	0.11				
$y_{2,t-9}$			0.44	0.06				
$y_{2,t-10}$			0.25	0.02				
$y_{2,t-11}$			-0.39	0.12				
$y_{2,t-12}$			-0.51	0.26				
$y_{2,t-13}$			0.20	0.03				
$y_{2,t-14}$			0.15	0.09				
$y_{2,t-15}$			0.10	0.02				
x_{t-1}	0.08	0.15	0.18	0.23	0.15	0.20	0.45	0.25
x_{t-2}	-0.07	0.22	-0.06	0.13	0.05	0.10	0.31	0.18
x_{t-3}	-0.05	0.12	0.08	0.09				
z_{t-1}	0.32	0.22	0.43	0.18	-0.31	0.28	-0.88	0.43
z_{t-1}	-0.04	0.12	-0.15	0.11	0.09	0.25	0.57	0.36
z_{t-3}	-0.12	0.18					-0.11	0.13

Table 2.3 : QR estimates and their estimated standard deviations for a selected bivariate two-regime TAR(15) model for the Iceland river flow data at quantile $\mathbf{u} = (0.9, 0.9)$. The threshold value is $0^o C$, and the numbers of observations in each regime are 479 and 601, respectively.

	Regime 1				Regime 2			
	y_{1t}		y_{2t}		y_{1t}		y_{2t}	
	Coef.	*Std*	*Coef.*	*Std*	*Coef.*	*Std*	*Coef.*	*Std*
$y_{1,t-1}$	1.30	0.41	1.54	0.43	1.91	0.74	1.69	0.38
$y_{1,t-2}$	0.86	0.22	1.47	0.13	1.80	0.72	1.61	0.30
$y_{1,t-3}$	0.81	0.19	0.85	0.22	1.79	0.66	0.97	0.29
$y_{1,t-4}$	0.64	0.27	-0.62	0.24	1.68	0.56	0.96	0.33
$y_{1,t-5}$	0.69	0.18	0.75	0.14	1.63	0.69	0.98	0.39
$y_{1,t-6}$	-0.43	0.17	0.61	0.20	1.43	0.33		
$y_{1,t-7}$	-0.19	0.12	-0.51	0.19	0.99	0.46		
$y_{1,t-8}$	0.18	0.09	0.39	0.06	0.72	0.24		
$y_{1,t-9}$	-0.33	0.14						
$y_{1,t-10}$	0.32	0.10						
$y_{1,t-11}$	0.28	0.04						
$y_{1,t-12}$	0.22	0.11						
$y_{1,t-13}$	-0.18	0.07						
$y_{1,t-14}$	0.31	0.04						
$y_{1,t-15}$	0.10	0.03						
$y_{2,t-1}$	-0.71	0.37	-1.19	0.43	1.59	0.50	1.66	0.73
$y_{2,t-2}$	-0.67	0.27	0.67	0.44	1.58	0.49	-1.43	0.66
$y_{2,t-3}$	-0.64	0.09	-0.55	0.23	1.51	0.44	1.14	0.49
$y_{2,t-4}$	0.69	0.15	0.53	0.21	1.08	0.16	1.21	0.48
$y_{2,t-5}$			-0.44	0.39	0.65	0.12	1.05	0.44
$y_{2,t-6}$			0.49	0.14	0.57	0.08	-1.02	0.56
$y_{2,t-7}$			-0.50	0.28	0.45	0.08	0.33	0.06
$y_{2,t-8}$			0.27	0.12	0.27	0.02		
$y_{2,t-9}$			0.45	0.18	0.45	0.18		
$y_{2,t-10}$			0.21	0.13	0.27	0.02		
$y_{2,t-11}$			0.22	0.12				
$y_{2,t-12}$			0.21	0.01				
$y_{2,t-13}$								
$y_{2,t-14}$								
$y_{2,t-15}$								
x_{t-1}	0.32	0.05	0.57	0.19	0.74	0.08	1.35	0.16
x_{t-2}	0.21	0.03	0.43	0.18	0.50	0.04	0.42	0.04
x_{t-3}			-0.05	0.01	-0.04	0.40	-0.34	0.04
z_{t-1}	0.41	0.11	-0.36	0.16	-0.41	0.13	0.54	0.02
z_{t-2}	0.31	0.18	-0.15	0.07	0.19	0.03	0.39	0.02
z_{t-3}	-0.18	0.02	-0.15	0.05			-0.46	0.04

2.5 Proofs

In this section, we give the proofs of our theorems in detail. For convenience we first introduce a lemma, which is from Chakraborty (2003), then we prove the theorems.

Lemma 2.1 Let $\boldsymbol{\phi}_n(\boldsymbol{\beta})$, n=1,2,... be a sequence of random functions on $\mathbb{R}^k$ and convex in $\boldsymbol{\beta}$. Let $\boldsymbol{\phi}(\boldsymbol{\beta})$ be a random function such that, for each fixed $\boldsymbol{\beta}$, $\boldsymbol{\phi}_n(\boldsymbol{\beta}) \to \boldsymbol{\phi}(\boldsymbol{\beta})$ in probability. Then for each $M > 0$,

$$\sup_{\|\boldsymbol{\beta}\| \leq M} \mid \boldsymbol{\phi}_n(\boldsymbol{\beta}) - \boldsymbol{\phi}(\boldsymbol{\beta}) \mid \to 0$$

in probability.

Proof of Theorem 2.1: Following the same arguments as for Theorem 4.1 of Chakraborty (2003), we obtain that

$$\sqrt{n}(\hat{\mathbf{\Phi}}_n(\mathbf{u}) - \tilde{\mathbf{\Phi}}(\mathbf{u})) = n^{-1/2}[D_1(\mathbf{u})]^{-1}[\sum_{i=1}^{n} \boldsymbol{\varphi}_{\mathbf{u}}(\mathbf{e}_i - \mathbf{Q}(\mathbf{u}))\mathbf{Z}_i^T]\mathbf{S}^{-1} + o_p(1). \quad (2.4)$$

Define $\mathbf{V}_i = \boldsymbol{\varphi}_{\mathbf{u}}(\mathbf{e}_i - \mathbf{Q}(\mathbf{u}))\mathbf{Z}_i^T$. Since function $\boldsymbol{\varphi}$ is bounded and smooth, $\mathbf{Z}_i = \{y_{i-1}, y_{i-2}, ...y_{i-p}, x_{i-1}, ...x_{i-q}\}$ has mean zero. Hence, we know $\{\mathbf{V}_i\}$; $i = 1, 2...n$, is also a strictly stationary and α-mixing process. Since $\mathbf{e}_i$ and $\mathbf{Z}_i$ are uncorrelated, we have $E(\mathbf{V}_i) = E(E[\mathbf{V}_i|\mathbf{Z}_i]) = \mathbf{0}$.

In the following we show the asymptotic normality of $\mathbf{T}_n = \sum_{i=1}^n \mathbf{V}_i$. Since $\{\mathbf{V}_i, i = 1, ...n\}$ are dependent, we employ the standard small-block and large-block arguments to complete this task. To this end, we partition the set $\{1, 2..., n\}$ into $2k_n + 1$ subsets with large blocks of size l_n and small blocks of size s_n. A large block is followed by a small block, and the last remaining set has size $n - k_n(l_n + s_n)$, where l_n and s_n are selected such that $s_n \longrightarrow \infty$, $s_n/l_n \longrightarrow 0$, $l_n/n \longrightarrow 0$, and the number of the blocks $k_n = [n/(l_n + s_n)] = O(s_n)$. Let $l_n = O(n^{(r-1)/r})$ and $s_n = O(n^{1/r})$ for any $r > 2$, then $k_n = O(n^{1/r}) = O(s_n)$. For $j = 1, 2 \ldots k_n$, define

$$\boldsymbol{\xi}_j = \sum_{i=(j-1)(l_n+s_n)+1}^{jl_n+(j-1)s_n} \mathbf{V}_i, \quad \boldsymbol{\eta}_j = \sum_{i=jl_n+(j-1)s_n+1}^{j(l_n+s_n)} \mathbf{V}_i, \quad \boldsymbol{\zeta} = \sum_{i=k_n(l_n+s_n)+1}^{n} \mathbf{V}_i.$$

Note that $\alpha(n) = o(n^{-1})$ and $k_n s_n/n \longrightarrow 0$. It follows from Proposition 2.7 of Fan and Yao (2003) that

$$\frac{1}{n} E(\sum_{j=1}^{k_n} \boldsymbol{\eta}_j)^2 \longrightarrow 0 \text{ and } \frac{1}{n} E(\boldsymbol{\zeta}^2) \longrightarrow 0.$$

This means that the summations over the small blocks and the residual block are asymptotically negligible. Thus,

$$\frac{1}{\sqrt{n}} \mathbf{T}_n = \frac{1}{\sqrt{n}} (\sum_{j=1}^{k_n} \boldsymbol{\xi}_j + \sum_{j=1}^{k_n} \boldsymbol{\eta}_j + \boldsymbol{\zeta}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} \boldsymbol{\xi}_j + o_p(1).$$

It follows from Proposition 2.6 of Fan and Yao (2003) that, as $n \longrightarrow \infty$,

$$|E\{exp(\frac{it}{\sqrt{n}} \sum_{j=1}^{k_n} \boldsymbol{\xi}_j)\} - \prod_{j=1}^{k_n} E\{exp(it\boldsymbol{\xi}_j/\sqrt{n})\}| \leq 16(k_n - 1)\alpha(s_n) \longrightarrow 0,$$

which implies the summations over the large blocks $\{\boldsymbol{\xi}_j\}$ are asymptotically independent. Now, by stationary, we have

$$\frac{1}{n} Var(\mathbf{T}_n) = \frac{1}{n} \sum_{j=1}^{n} Var(\mathbf{V}_j) + \frac{2}{n} \sum_{1 \leq i < j \leq n} Cov(\mathbf{V}_i, \mathbf{V}_j) = \gamma(0) + 2 \sum_{l=1}^{n-1} (1 - \frac{l}{n}) \gamma(l),$$

where $\gamma(k) = Cov(\mathbf{V}_{i+k}, \mathbf{V}_i)$ is the autocovariance function of $\mathbf{V}_i$. Define $\boldsymbol{\Sigma}_1 = D_2(\mathbf{u})\mathbf{S}$. It is straightforward to show that $\boldsymbol{\Sigma}_1 = D_2(\mathbf{u})\mathbf{S} = \gamma(0) + 2\sum_{j=1}^{\infty} \gamma(j)$. Applying Theorem 2.20 of Fan and Yao (2003), we have $l_n^{-1} E(\boldsymbol{\xi}_1^2) \longrightarrow \boldsymbol{\Sigma}_1$, which implies the Feller condition

$$\frac{1}{n}\sum_{j=1}^{k_n} E(\boldsymbol{\xi}_j^2) = \frac{k_n l_n}{n}\frac{1}{l_n}E(\boldsymbol{\xi}_1^2) \longrightarrow \boldsymbol{\Sigma}_1.$$

Note that for any $\epsilon > 0$,

$$\begin{aligned} E[\boldsymbol{\xi}_1^2 I(|\boldsymbol{\xi}_1| > \sqrt{n}\epsilon|\boldsymbol{\Sigma}_1|^{1/2})] &\leq E(\boldsymbol{\xi}_1^4)^{1/2} P[|\boldsymbol{\xi}_1| > \sqrt{n}\epsilon|\boldsymbol{\Sigma}_1|^{1/2}] \\ &\leq C l_n \frac{1}{n\epsilon^2}|\boldsymbol{\Sigma}_1|^{-1}E(\boldsymbol{\xi}_1^2) = O(l_n^2/n). \end{aligned}$$

It follows that

$$\frac{1}{n}\sum_{j=1}^{k_n} E[\boldsymbol{\xi}_j^2 I(|\boldsymbol{\xi}_j| \geq \sqrt{n}\epsilon|\boldsymbol{\Sigma}_1|^{1/2})] = O(k_n l_n^2/n^2) = O(l_n/n) \longrightarrow 0,$$

which is the Lindberg condition. Under the Lindberg and Feller conditions, by the central limit theorem, we have

$$\prod_{j=1}^{k_n} E[exp(i\mathbf{t}\boldsymbol{\xi}_j/\sqrt{n})] \longrightarrow exp\{-\mathbf{t}\boldsymbol{\Sigma}_1\mathbf{t}^T/2\},$$

for any $\mathbf{t}$. That is, $\frac{1}{\sqrt{n}}\mathbf{T}_n \longrightarrow \mathrm{N}(0,\boldsymbol{\Sigma}_1)$ in distribution as $n \longrightarrow \infty$. This combined with equation (2.4) leads to

$$\sqrt{n}(\hat{\boldsymbol{\Phi}}_n(\mathbf{u}) - \tilde{\boldsymbol{\Phi}}(\mathbf{u})) = n^{-1/2}[D_1(\mathbf{u})]^{-1}\mathbf{T}_n\mathbf{S}^{-1} + o(1) \xrightarrow{\mathcal{D}} N(0,\boldsymbol{\Sigma}). \tag{2.5}$$

Proof of Theorem 2.2: The result follows from equation (2.5).

CHAPTER 3: MULTIVARIATE TIME-VARYING COEFFICIENT MODEL

3.1 Local Linear Smoother for Multivatiate Time-varying Coefficient Model

This section mainly discusses the local linear estimation of the time-varying coefficient and asymptotic distribution of the nonparametric estimator.

3.1.1 Estimation

Recall for model (1.2), I propose the multivariate time-varying coefficient model:

$$\mathbf{y}_t = \mathbf{c}(t/T) + \sum_{i=1}^{p} \boldsymbol{\alpha}_i(t/T)\mathbf{y}_{t-i} + \sum_{j=1}^{q} \boldsymbol{\beta}_j(t/T)\mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t, t = 1, ..., T. \tag{3.1}$$

where $\mathbf{y}_t$ is $k \times 1$ vector, $\mathbf{x}_t$ is $v \times 1$ vector. $\mathbf{c}(\centerdot)$ is a $k \times 1$ vector, $\boldsymbol{\alpha}_i$ is $k \times k$ smooth matrix and $\boldsymbol{\beta}_j$ is $k \times v$ smooth matrix. The innovations satisfy $\boldsymbol{\varepsilon}_t = \gamma_t^* \mathbf{a}_t$, where γ_t^* are symmetric positive definite matrices and $\mathbf{a}_t$ is a sequence of uncorrelated random vectors with mean zero and convariance matrix $\mathbf{I}_k$. Let $\mathbf{X}_t = vec(1, \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-p}, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-q})$ be a $d \times 1$ vector with $d = 1+kp+vq$, and $\boldsymbol{\Phi}(t/T) = (\mathbf{c}(t/T), \boldsymbol{\alpha}_1(t/T), ...\boldsymbol{\alpha}_p(t/T), \boldsymbol{\beta}_1(t/T), ...\boldsymbol{\beta}_q(t/T)$, Then model (3.1) becomes:

$$\mathbf{y}_t = \boldsymbol{\Phi}(t/T)\mathbf{X}_t + \boldsymbol{\varepsilon}_t, t = 1, ...T. \tag{3.2}$$

where $\boldsymbol{\Phi}(\centerdot)$ is $k \times d$ matrix and $\mathbf{X}_t$ is $d \times 1$ vector.

For any t in the neighborhood of $t_0 \in (0, T)$, i.e $\mid \dfrac{t - t_0}{T} \mid \leq h$, using the Taylor expansion, we obtain:

$$\boldsymbol{\Phi}(t/T) \approx \boldsymbol{\Phi}(t_0/T) + \boldsymbol{\Phi}'(t_0/T)(\frac{t - t_0}{T})$$

$$\equiv \mathbf{P} + \mathbf{Q}(\frac{t-t_0}{T})$$

Running the local linear smoother for model (3.2), we minimize:

$$\sum_{t=s+1}^{T} \parallel \mathbf{y}_t - \mathbf{P}\mathbf{X}_t - \mathbf{Q}\mathbf{X}_t(\frac{t-t_0}{T}) \parallel^2 K_h(t-t_0) \tag{3.3}$$

over $\mathbf{P}$ and $\mathbf{Q}$, where $\parallel \centerdot \parallel$ denotes the Euclidean norm, $s = max(p,q)$ and $K_h(x) = \frac{1}{h}K(\frac{x}{hT})$ for a kernel function $K(\centerdot)$ and a bandwidth h controlling the amount of smoothing. Let the resulting minimizers for $(\mathbf{P},\mathbf{Q})$ be $(\hat{\mathbf{P}},\hat{\mathbf{Q}})$.

Let $\mu_i = \int u^i K(u)du$, $\nu_i = \int u^i K^2(u)du$, $\omega = (\mu_2,\mu_3)^T$,

$$U = \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}, V = \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}$$

Define $\mathbf{M} = E[(\boldsymbol{X}_t\boldsymbol{X}_t^T) \otimes I_k]$ and $\mathbf{N} = E[(\boldsymbol{X}_t\boldsymbol{X}_t^T) \otimes (\gamma_t^*)^2]$. Next, I derive the explicit representation of the estimator by using local linear fitting .

Theorem 3.1 The solution $(\hat{\mathbf{P}},\hat{\mathbf{Q}})$ for (3.3) admits the following closed form:

$$\begin{bmatrix} vec(\hat{\mathbf{P}}) \\ vec(h\hat{\mathbf{Q}}) \end{bmatrix} = \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h(t-t_0) \\ \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h^{(1)}(t-t_0) \end{bmatrix} \tag{3.4}$$

where $\otimes$ denotes the kronecker product. I_k is the $k \times k$ indentity matrix, $\mathbf{S}_{Ti} = \sum_{t=s+1}^{T}(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k K_h^{(i)}(t-t_0)$ and $K_h^{(i)}(t-t_0) = (Th)^{(-i)}(t-t_0)^i K_h(t-t_0)$, for $i = 0,1,2$.

Proof: See Section 3.5. □.

3.1.2 Asymptotic Distribution

To derive the asymptotic distribution of the above estimators, we need the following assumptions.

Assumption B:

(B1) For any $u = t_0/T \in (0,1)$, the second derivative of $\mathbf{\Phi}(\centerdot)$ exists and is continuous at u.

(B2) The kernel function $K(v)$ is symmetrical with a bounded support s.t $\mu_0(K) = 1$ and $\mu_1(K) = 0$ i.e. $\int K(v)dv = 1$ and $\int vK(v)dv = 0$. Further, the functions $v^3K(v)$ and $v^3K'(v)$ are bounded with $v^4K(v) < \infty$.

(B3) There exists a positive $\rho > 0$ such that $E \parallel \mathbf{a}_t \parallel^{1+\rho} < \infty$.

(B4) Assume that γ_t^* is measurable with respect to the σ-field generated by historical information $\mathcal{F}_{t-1} = \{\mathbf{w}_s; s \leq t-1\}$, where $\mathbf{w}_{t-1} = (\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...\mathbf{y}_{t-p}, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-q})$.

(B5) $\mathbf{M}$ and $\mathbf{N}$ are both invertible positive definite matrices.

(B6) The processes $\{\mathbf{X}_t, \mathbf{a}_t\}$ are strictly stationary with α-mixing coefficients $\alpha(s)$ such that $\sum_s s^c[\alpha(s)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$.

(B7) $h \longrightarrow 0$ in such a way that $hT \longrightarrow 0$. There exists a sequence of positive integers $\{r_T\}$ s.t. $r_T \longrightarrow \infty$, $r_T = o(\sqrt{hT})$ and $\sqrt{T/h}\alpha(r_T) \longrightarrow 0$ as $T \longrightarrow \infty$.

Theorem 3.2 Suppose the assumptions (B1)-(B7) hold. Then, for any $u = t_0/T \in (0,1)$, we have:

$$\sqrt{Th}\{\begin{pmatrix} vec(\hat{\mathbf{P}} - \mathbf{\Phi}(u)) \\ vec(h(\hat{\mathbf{Q}} - \mathbf{\Phi}'(u))) \end{pmatrix} - \mathbf{B}_T(u)\} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$$

where $\mathbf{B}_T(u) = \frac{h^2}{2}(U^{-1}\omega) \otimes vec(\mathbf{\Phi}"(u))(1 + o_p(1))$ and $\mathbf{\Sigma} = (U^{-1}VU^{-1}) \otimes (\mathbf{M}^{-1}\mathbf{N}\mathbf{M}^{-1})$.

Proof: See Section 3.5. $\square$.

3.2 Generalized Likelihood Ratio for Multivariate Time-varying Coefficient Model

This section briefly discussed the testing hypotheses about whether the coefficients of the time-varying regression models are of some specific functional forms or constants.

3.2.1 Generalized Likelihood Ratios

The likelihood ratio type test was proposed by Fan, Zhang and Zhang (2001) and studied extensively by Fan and Jiang (2005). For the varying coefficient regression model: $Y = A(U)^T \mathbf{X} + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2)$ with $\mathbf{X} = (x_1, ..., xp)^T$ and $A(U) = (a_1(U), ...a_p)^T$. After fitting the regression models via local linear technique. Fan, Zhang and Zhang(2001) raises one interesting problem to check whether the varying coefficients are of some specific functional forms. This is equivalent to the following hypotheses:

$$H_0 : A(U) = A_0(U) \longleftrightarrow H_a : A(U) \neq A_0(U) \tag{3.5}$$

where $A_0(U)$ is a vector of known functionals. One special case of (3.5) is when $A_0(U)$ is a vector of constants. Then the test hypothesis becomes to checking whether the varying coefficients are indeed varying. That is equivalent to:

$$H_0 : A(U) = A_0 \longleftrightarrow H_a : A(U) \neq A_0 \tag{3.6}$$

where A_0 is a vector of known or unknown constants.
The test statistic is defined as:

$$\lambda_n = \mathcal{L}_n(H_a) - \mathcal{L}_n(H_0) = \frac{n}{2} log(\frac{RSS_0}{RSS_a}) \approx \frac{n}{2}(\frac{RSS_0 - RRS_a}{RSS_a})$$

where $\mathcal{L}_n(H_a)$ and $\mathcal{L}_n(H_0)$ are the log-likelihood under H_a and H_0, respectively. $RSS_a = \sum_{k=1}^{n}(Y_k - \hat{A}^T(U_k)\mathbf{X}_k)^2$ and $RSS_0 = \sum_{k=1}^{n}(Y_k - \hat{A}_0^T(U_k)\mathbf{X}_k)^2$. Here $\hat{A}(U)$ is the corresponding nonparametric estimator of $A(U)$ and $\hat{A}_0(U)$ is the true or estimated value of coefficients under H_0.

3.2.2 Test Statistics

Motivated by Fan, Zhang and Zhang (2001) and Fan and Jiang(2005). Suppose $\{\mathbf{y}_t, \mathbf{x}_t\}_{t=1}^T$ are a random sample from the multivariate time-varying coefficient model (3.2). Namely,

$$\mathbf{y}_t = \mathbf{\Phi}(t/T)\mathbf{X}_t + \boldsymbol{\varepsilon}_t \tag{3.7}$$

Now, we assume $\mathbf{\Sigma}_0^{-1/2}\boldsymbol{\varepsilon}_t$ has mean zero and covariance $\mathbf{I}_k$. where $\mathbf{\Sigma}_0$ is a symmetric positive definite constant matrix. $\mathbf{X}_t = vec(1, \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-p}, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-q})$ is a $d \times 1$ vector with $d = 1 + kp + vq$ and $\mathbf{\Phi}(t/T) = (\mathbf{c}(t/T), \boldsymbol{\alpha}_1(t/T), ...\boldsymbol{\alpha}_p(t/T), \boldsymbol{\beta}_1(t/T), ...$ $\boldsymbol{\beta}_q(t/T))$ is a $k \times d$ matrix.

I consider the simple null hypothesis testing problem:

$$H_0 : \mathbf{\Phi}(t/T) \in \Theta_0(t/T) \longleftrightarrow H_a : \mathbf{\Phi}(t/T) \notin \Theta_0(t/T) \tag{3.8}$$

where $\Theta_0(t/T)$ is a set of functionals of matrix . Denote $\hat{\mathbf{\Phi}}(t/T)$ as the corresponding nonparametric estimator of $\mathbf{\Phi}$. $\hat{\mathbf{\Phi}}_0(t/T)$ is the true or estimated value of coefficients under H_0. I propose the similar test statistic for the testing problem in (3.8) as:

$$\lambda_T = \mathcal{L}(H_a) - \mathcal{L}(H_0) = \frac{T}{2} log(\frac{R\tilde{S}S_0}{R\tilde{S}S_a}) \approx \frac{T}{2}(\frac{R\tilde{S}S_0 - R\tilde{S}S_a}{R\tilde{S}S_a})$$

where $\mathcal{L}(H_a)$ and $\mathcal{L}(H_0)$ are the log-likelihood under H_a and H_0, respectively. $R\tilde{S}S_a = \sum_{t=1}^T(\mathbf{y}_t - \hat{\mathbf{\Phi}}(t/T)\mathbf{X}_t)^T\mathbf{\Sigma}^{-1}(\mathbf{y}_t - \hat{\mathbf{\Phi}}(t/T)\mathbf{X}_t)$ and $R\tilde{S}S_0 = \sum_{t=1}^T(\mathbf{y}_t - \hat{\mathbf{\Phi}}_0(t/T)\mathbf{X}_t)^T$ $\mathbf{\Sigma}^{-1}(\mathbf{y}_t - \hat{\mathbf{\Phi}}_0(t/T)\mathbf{X}_t)$. Where Σ is a positive definite matrix as a working covariance of $\boldsymbol{\varepsilon}_t$. So $\mathbf{\Sigma}^{-1}$ can be written in terms of spectral decomposition as:

$$\mathbf{\Sigma}^{-1} = \mathbf{Q}^T \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix} \mathbf{Q}$$

where $\lambda_1 \geqslant \lambda_2 \cdots \geqslant \lambda_k > 0$ are the eigenvalues of $\mathbf{\Sigma}^{-1}$ and $\mathbf{Q}$ is the orthogonal matrix having rows $\mathbf{q}_1, ...\mathbf{q}_k$ which are normalized eigen-vectors corresponding to $\lambda_1, ..., \lambda_k$. With this spectral decomposition of $\mathbf{\Sigma}^{-1}$, model (3.7) becomes equivalently to:

$$\mathbf{Q}\mathbf{y}_t = \mathbf{Q}\mathbf{\Phi}(t/T)\mathbf{X}_t + \mathbf{Q}\boldsymbol{\varepsilon}_t$$

Denote $\mathbf{y}_t^* = \mathbf{Q}\mathbf{y}_t$, $\mathbf{\Phi}^*(t/T) = \mathbf{Q}\mathbf{\Phi}(t/T)$, $\boldsymbol{\varepsilon}_t^* = \mathbf{Q}\boldsymbol{\varepsilon}_t$. Hence, from now on, we focus on the model:

$$\mathbf{y}_t^* = \mathbf{\Phi}^*(t/T) + \boldsymbol{\varepsilon}_t^* \tag{3.9}$$

where $\boldsymbol{\varepsilon}_t^*$ has mean zero and covariance matrix $\mathbf{Q}\mathbf{\Sigma}_0\mathbf{Q}^T$.

Accordingly, the testing hypothesis problem (3.8) becomes:

$$H_0 : \mathbf{\Phi}^*(t/T) \in \Theta_0^*(t/T) \longleftrightarrow H_a : \mathbf{\Phi}^*(t/T) \notin \Theta_0^*(t/T) \tag{3.10}$$

3.2.3 Asymptotic Null Distribution

To derive the asymptotic distribution of $\lambda_T(\mathbf{\Phi}_0^*)$ under H_0, we need the following assumptions.

Assumption C

(C1) $\mathbf{\Phi}^*(u)$ has the continuous second derivative at any $u = t_0/T \in (0, 1)$.

(C2) The kernel function $K(v)$ is symmetrical with a bounded support s.t $\mu_0(K) = 1$ and $\mu_1(K) = 0$ i.e. $\int K(v)dv = 1$ and $\int vK(v)dv = 0$. Further, the functions $v^3K(v)$

and $v^3K'(v)$ are bounded with $v^4K(v) < \infty$.

(C3) $E|\boldsymbol{\varepsilon}_t^*|^4 < \infty$.

(C4) $\mathbf{X}_t$ is bounded. The $d \times d$ matrix $E[\mathbf{X}_t\mathbf{X}_t^T]$ is invertible. $(E[\mathbf{X}_t\mathbf{X}_t^T])^{-1}$ and $E[(\mathbf{X}_t\mathbf{X}_t^T) \otimes \mathbf{\Sigma}_0]$ are both Lipschitz continuous.

We define: $\mathbf{\Gamma} = E[\mathbf{X}_t\mathbf{X}_t^T]$ and $\omega_0 = \iint t^2(s+t)^2K(t/T)K((s+t)/T)dtds$.

Denote $D = k \times d$, $\mathbf{\Omega} \equiv \mathbf{Q}\mathbf{\Sigma}_0\mathbf{Q}^T = (\sigma_{ij}^2)_{i,j=1}^k$ i.e, it has (i,j)-th element as σ_{ij}^2, $i,j = 1,...k$. For $j = 1,2...k$, Let $\varepsilon_{tj}^* = y_{tj}^* - \mathbf{\Phi}_{0j}^*(t/T)\mathbf{X}_t$,

$R_{T10}^j = \dfrac{1}{\sqrt{T}}[\sum_{t=1}^T \varepsilon_{tj}^* \mathbf{\Phi}_{0j}^{"*}\mathbf{X}_t \int t^2K(t/T)dt](1+O(h)+O(T^{-1/2}))$,

$R_{T20}^j = \dfrac{1}{2\sqrt{T}}\sum_{t=1}^T \varepsilon_{tj}^*\mathbf{X}_t^T\mathbf{\Gamma}^{-1}(\mathbf{\Phi}_{0j}^{*"}(t/T)\mathbf{X}_t)E(\mathbf{X}_t)\omega_0$,

$R_{T30}^j = \dfrac{1}{8}E[\mathbf{\Phi}_{0j}^{*"}(t/T)\mathbf{X}_t\mathbf{X}_t^T\mathbf{\Phi}_{0j}^{*"}(t/T)^T]\omega_0(1+O(T^{-1/2}))$,

$d_{1Tj} = \sigma_{jj}^{-2}[Th^4R_{T30}^j - T^{1/2}h^2(R_{T10}^j - R_{T20}^j)] = O_p(Th^4 + \sqrt{T}h^2)$,

$\mu_T = \dfrac{D}{h}(K(0) - \dfrac{1}{2}\int K^2(x)dx)$, $\sigma_T^2 = \dfrac{D}{2h}\int (2K(x) - K*K(x))^2dx$,

$d_{1T}^* = \dfrac{1}{k}\sum_{j=1}^k \lambda_j\sigma_{jj}^2 d_{1Tj}$, $\mu_T^* = \dfrac{\mu_T}{k}\sum_{j=1}^k \lambda_j\sigma_{jj}^2$,

$\sigma^{*2} = \dfrac{\sigma_T^2}{k^2}(\sum_{j=1}^k \lambda_j^2\sigma_{jj}^4 + 2\sum_{i<j}^k \lambda_i\lambda_j\sigma_{ij}^4)$, where $K*K$ denotes the convolution product of K, note that both R_{T10}^j and R_{T20}^j are asymptotically normal and hence are stochastically bounded.

Then, we have the following theorem.

Theorem 3.3 Suppose Assumptions (C1)-(C4) hold. Then under H_0, as $h \longrightarrow 0$ and $Th^{3/2} \longrightarrow \infty$,

$$\sigma^{*-1}(\lambda_T(\mathbf{\Phi}_0) - \mu_T^* - d_{1T}^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Proof: See Section 3.5. □

One special case of the hypothesis in (3.10) is to check whether the coefficient functions are actually varying. This means when $\Theta_0^*(t/T)$ is some known constant matrix $\mathbf{\Phi}_0^*$. In this case, we have the following asymptotic result.

Theorem 3.4 Suppose Assumption (C1)-(C4) hold. Then under H_0, as $h \longrightarrow 0$ and

$Th^{3/2} \longrightarrow \infty$,

$$\sigma^{*-1}(\lambda_T(\mathbf{\Phi}_0) - \mu_T^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Proof: See Section 3.5. □

3.2.4 Power of Test

In this section, we consider the power of the quasi-likelihood ratio test based on local linear fitting. For simplicity, we fix the null hypothesis in (3.10) with a known matrix.

For any $u = t/T \in (0,1)$, if we rewrite matrix $\mathbf{\Phi}^*(u)$ as a vector:

$$\mathbf{\Delta}(u) \equiv vec(\Phi_1^*(u), \Phi_2^*(u), ...\Phi_D^*(u))$$

Denote: $\mathbf{\Delta}_0(u) \equiv vec(\Phi_{01}^*(u), \Phi_{02}^*(u), ...\Phi_{0D}^*(u))$, then the power of the test is considered under the local alternatives as follows:

$$H_a : \mathbf{\Delta}(u) = \mathbf{\Delta}_0(u) + \mathbf{G}_T(u)$$

where $\mathbf{G}_T(u) = \dfrac{1}{\sqrt{Th}}(g_1(u), g_2(u), ..., g_D(u))^T$ is a $D \times 1$ vector of functions. So, the power of the test under H_a can be approximated by using the following theorem.

Theorem 3.5 Suppose that Assumptions (C1)-(C4) hold and $\mathbf{\Delta}(u)$ is linear in u or $Th^5 \longrightarrow 0$. If $ThE\{\mathbf{G}_T^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}_T(u)\} \longrightarrow C(\mathbf{G})$ and $E\{(\mathbf{G}_T^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}_T(u))\boldsymbol{\varepsilon}_t^{T*}\boldsymbol{\varepsilon}_t^*\}^2 = O((Th)^{-3/2})$ for some constant $C(\mathbf{G})$, then under H_a:

$$\sigma_1^{*-1}(\lambda_T(\mathbf{\Phi}_0) - \mu_T^* - d_{2T}^* + \nu_T^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where $\sigma_1^{2*} = \sigma^{2*} + \dfrac{T}{k^2}E\{\mathbf{G}_T^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}_T(u)\}(\sum_{j=1}^k \lambda_j^2\sigma_{jj}^2 + 2\sum_{i<j}^k \lambda_i\lambda_j\sigma_{ij}^2)$,

$d^*_{2T} = \dfrac{T}{2k}E\{\mathbf{G}_T^T(u)[(\mathbf{X}_t\mathbf{X}_t^T)\otimes I_k]\mathbf{G}_T(u)\}(\sum_{j=1}^k \lambda_j\sigma_{jj}^2)$, $\nu^*_T = \dfrac{Th^4}{8k}E\{\mathbf{\Delta}^{"T}(u)[(\mathbf{X}_t\mathbf{X}_t^T)\otimes I_k]\mathbf{\Delta}^{"}(u)\}\omega_0(\sum_{j=1}^k \lambda_j)$ and μ^*_T is given in Theorem 3.1.

Proof: See Section 3.5. □.

3.3 Simulations

In this section, I conduct Monte Carlo simulation to demonstrate the power of the proposed GLR. The effect of the error distribution on the performance of the proposed test is also investigated. Throughout this section, the Gaussian Kernel $K(u) = \dfrac{1}{\sqrt{2\pi}}e^{-u^2/2}$ is used. Simulation procedures and results are given below.

3.3.1 Simulation Procedures: Conditional Bootstrap

To implement the GLR tests, we need to obtain the null distribution of the test statistic. In Section 3.2.2, we give the theoretical asymptotic distribution of the statistic. For a finite sample, the null distribution can be approximated by simulation via fixing nuisance parameters/functions at their reasonable estimates. This simulation method is referred to as the conditional bootstrap, which is detailed as follows:

1) Fix the optimal bandwidth at its estimated value $\hat{h}_{opt}$ and then obtain the estimators of the coefficient $\hat{\mathbf{\Phi}}(t/T)$ under both null and alternative models.

2) Compute the GLR test statistic $\lambda_T(H_0)$ by definition and the residuals $\mathbf{e}_t$ (for $t = 1, 2, ...T$) from the unrestricted model under H_a.

3) For each X_t, draw a bootstrap residual $\mathbf{e}^*_t$ from the centered empirical distribution of $\mathbf{e}_t$ and compute $\mathbf{y}^*_t = \hat{\mathbf{\Phi}}(t/T)\mathbf{X}_t + \mathbf{e}^*_t$. where $\hat{\mathbf{\Phi}}(t/T)$ is the estimated regression coefficients under H_a in step 1). This forms a conditional bootstrap sample $\{\mathbf{X}_t, \mathbf{y}^*_t\}_{t=1}^T$.

4) Using the bootstrap sample in step 3) with the bandwidth $\hat{h}_{opt}$, obtain the GLR $\lambda^*_T(H_0)$ in the same manner as $\lambda_T(H_0)$.

5) Repeat steps 3) and 4) many times, say 1000 times to get a sample of statistic

$\lambda_T^*(H_0)$. The critical value at significant level α is given by the $(1-\alpha)$th quantile.

3.3.2 Simulation Results

In this section, I consider the following data generating model:

$$\mathbf{y}_t = \mathbf{\Phi}(t/T)\mathbf{X}_t + \mathbf{e}_t, t = 1, ...T. \tag{3.11}$$

where $k = 2$, $v = p = q = 1$, $D = k \times d = 6$, $\mathbf{\Delta} = vec(\Phi_1, .., \Phi_6) = (0.5, 0.0075, 0.08, 0.65, 0.25, 0.75)^T$, set the initial value $\mathbf{y}_1 = (0.15, 0.2)$ and $x_1 = 0$. In this case, $\mathbf{X}_t = vec(y_{1,t-1}, y_{2,t-1}, x_{t-1})$, for $t = 2, ...T$. $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$. where $\Sigma = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$. For the power assessment, we evaluate the power for a sequence of alternative models indexed by θ:

$$H_\theta : \mathbf{\Delta}_\theta = (0.5, 0.0075, 0.08, 0.65, 0.25, 0.75)^T + \frac{\theta}{\sqrt{Th}}\mathbf{G}(t/T) \tag{3.12}$$

where $\mathbf{G}(t/T) = (sin(\sqrt{2}\pi t/T), -0.09cos(\pi t/T), 0.16sin(\sqrt{3}\pi t/T), 0.8sin(\sqrt{2\pi}t/T), 0.3sin(\pi t/T), cos(\sqrt{1.5}\pi t/T))^T$. The simulation is repeated 600 times for each sample size $n = 200$, $n = 400$ and $n = 800$ and for each $\theta = 0$, $\theta = 0.2$, $\theta = 0.4$, $\theta = 0.6$ $\theta = 0.8$ and $\theta = 1.0$. For each given value of θ, I use 1000 Monte Carlo replications for the calculation of the critical values via the conditional bootstrap method (see section 3.3.1). Given the significance of level 5% and 10%, the power function $\rho(\theta)$ is estimated based on the relative frequency of $\lambda_T(\Phi)$ over 600 simulations. In addition to the bivariate normal distribution, the bivariate $t(5)$ and bivariate lognormal$(\mathbf{0}, \Sigma)$ distribution.where Σ is the same variance matrix as of the bivariate normal distribution. I plot the power curves in Figure 3.1 and Figure 3.2 at significance levels 10% and 5% for all settings.

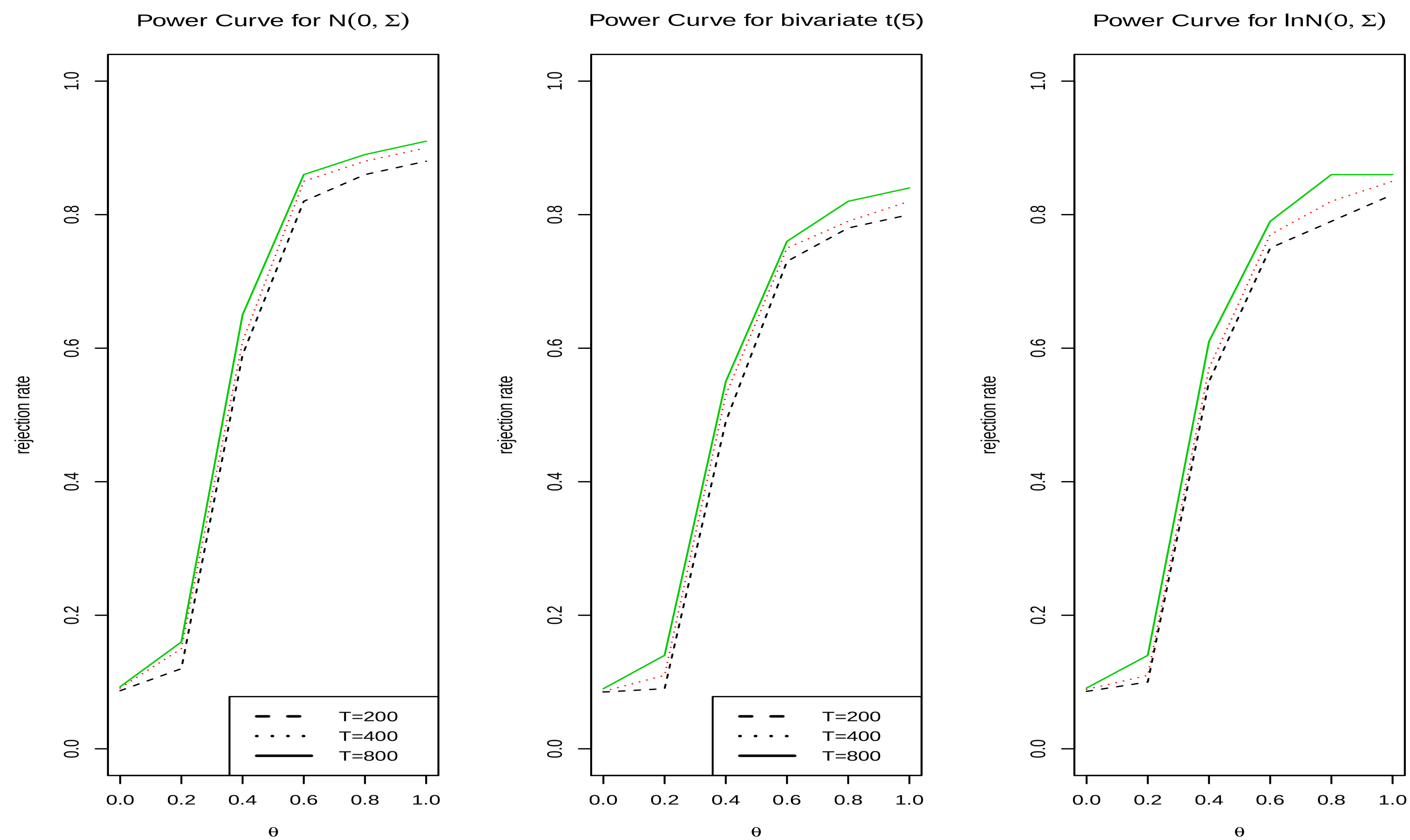

Figure 3.1 : The power curves for the testing hypothesis in (3.10) with the nominal size 10%. The dashed line is for $T = 200$, the dotted line is for $T = 400$ and the solid line is for $T = 800$ in Section 3.3.2.

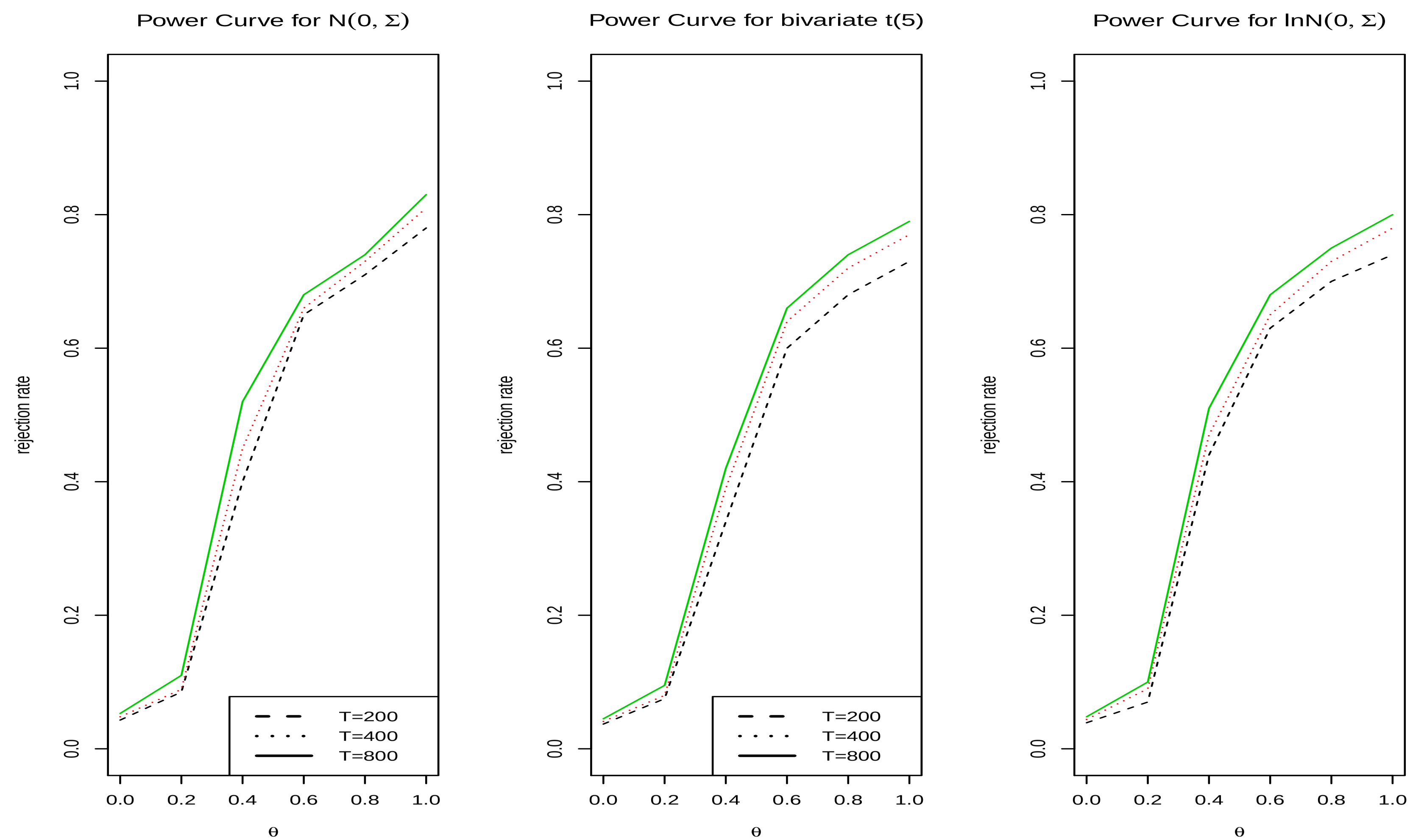

Figure 3.2 : The power curves for the testing hypothesis in (3.10) with the nominal size 5%. The dashed line is for $T = 200$, the dotted line is for $T = 400$ and the solid line is for $T = 800$ in Section 3.3.2.

3.4 Real Example

In previous section, I conducted Monte Carlo simulation to illustrate the effectiveness and the validity of the proposed test statistics. In this section, I consider the application of these methodologies to a real example. Here I analyze a subset of the interest rates of the Federal Reserve Bank of St.Louis (http://research.stlouisfed.org /fred2/). They are monthly 1-year and 10-year Treasury constant maturity rates, which represent short-term and long-term series, respectively. The data consist of 571 monthly observations from January 1984 to October 2000.

Let Y_{1t} and Y_{2t} be the interest rate series of the 1-year and 10-year Treasury, respectively. Denote $s_{1t} = ln(Y_{1t})$ and $s_{2t} = ln(Y_{2t})$. I use the logarithm return $\mathbf{y}_t = (y_{1t}, y_{2t})^T$, where $y_{it} = s_{it} - s_{i,t-1}, i = 1, 2$. I fit the data using the following bivariate AR(2) model:

$$\mathbf{y}_t = \mathbf{a}\mathbf{y}_{t-1} + \mathbf{b}\mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t \tag{3.13}$$

where $\mathbf{a} = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$ and $\mathbf{b} = \left(\begin{smallmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{smallmatrix}\right)$. We are interested in using the proposed GLR test statistic to test:

$$H_0 : \mathbf{y}_t = \mathbf{a}\mathbf{y}_{t-1} + \mathbf{b}\mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t \longleftrightarrow H_a : \mathbf{y}_t = \mathbf{a}(t/T)\mathbf{y}_{t-1} + \mathbf{b}(t/T)\mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t \tag{3.14}$$

If we rewrite the coefficient matrices into vectors, denote:

$\boldsymbol{\Delta}_0 = (\hat{a}_{11}, \hat{a}_{21}, \hat{a}_{12}, \hat{a}_{22}, \hat{b}_{11}, \hat{b}_{21}, \hat{b}_{12}, \hat{b}_{22})^T$, $\boldsymbol{\Delta}_a = (a_{11}(t/T), a_{21}(t/T), a_{12}(t/T), a_{22}(t/T), b_{11}(t/T), b_{21}(t/T), b_{12}(t/T), b_{22}(t/T))^T = (\hat{a}_{11}, \hat{a}_{21}, \hat{a}_{12}, \hat{a}_{22}, \hat{b}_{11}, \hat{b}_{21}, \hat{b}_{12}, \hat{b}_{22})^T + \frac{1}{\sqrt{Th}}$ $\mathbf{G}(t/T)$, where $(\hat{a}_{11}, \hat{a}_{21}, \hat{a}_{12}, \hat{a}_{22}, \hat{b}_{11}, \hat{b}_{21}, \hat{b}_{12}, \hat{b}_{22})^T = (0.230, -0.032, 0.334, 0.398, 0.024, 0.008, -0.184, -0.152)^T$ is the estimated coefficients using software R for model (3.13). $\mathbf{G}(t/T) = (0.08sin(\sqrt{2}\pi t/T), 0.3sin(\pi t/T), 0.16sin(\sqrt{3}\pi t/T), cos(\sqrt{1.5}\pi t/T), -0.09cos(\pi t/T), 0.3sin(\pi t/T), 0.8sin(\sqrt{2\pi}t/T), cos(\sqrt{1.5}\pi t/T))^T$. Then, it is equivalent to test:

$$H_0 : \boldsymbol{\Delta} = \boldsymbol{\Delta}_0 \longleftrightarrow H_a : \boldsymbol{\Delta} = \boldsymbol{\Delta}_a \tag{3.15}$$

To compute the p-value of the test statistic, I need to find the null distribution of the GLR statistic $\lambda_T(H_0)$. This can be estimated by the conditional bootstrap method mentioned in section 3.3.1. The p-values are computed from 500 bootstrap replicates for using different bandwidths. The corresponding p values are reported in Table 3.1. Therefore, one can see all the p-values are greater than significant level 0.05, which implies that the varying coefficients are indeed time-varying.

Table 3.1 : The p-values for testing constancy in hypothesis (3.15)

h	$h=\frac{1}{2}\hat{h}_{opt}$	$h=\hat{h}_{opt}$	$h=\frac{3}{2}\hat{h}_{opt}$
$p-value$	0.032	0.006	0.019

3.5 Proofs

In this section, we give the derivation of the main results presented in previous sections of this chapter.

Proof of Theorem 3.1: We can prove it by following the similar steps in the proof of Lemma 1 in Jiang (2013): The proof involves taking the derivative of a generic matrix-valued function $\mathbf{F(X)}$ with respect to a matrix $\mathbf{X}$. Taking derivative over $\mathbf{P}$ and $\mathbf{Q}$ for (3.3), we obtain the score equations:

$$\left\{\begin{array}{l}\sum_{t=s+1}^{T}(\mathbf{X}_t\otimes I_k)[\mathbf{y}_t-\hat{\mathbf{P}}\mathbf{X}_t-\hat{\mathbf{Q}}(\frac{t-t_0}{T})]K_h(t-t_0)=0\\ \sum_{t=s+1}^{T}(\mathbf{X}_t\otimes I_k)[\mathbf{y}_t-\hat{\mathbf{P}}\mathbf{X}_t-\hat{\mathbf{Q}}(\frac{t-t_0}{T})]K_h^{(1)}(t-t_0)=0\end{array}\right. \tag{3.16}$$

For conforming matrices, we have the identity:

$$vec(\mathbf{AXB})=(\mathbf{B}^T\otimes\mathbf{A})vec(\mathbf{X}) \tag{3.17}$$

This combined with the identity:

$$(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C}=\mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C}) \tag{3.18}$$

yields that

$$(\mathbf{X}_t \otimes I_k)\hat{\mathbf{P}}\mathbf{X}_t = vec((\mathbf{X}_t \otimes I_k)\hat{\mathbf{P}}\boldsymbol{X}_t)$$
$$= ((\mathbf{X}_t^T \otimes \mathbf{X}_t) \otimes I_k)vec(\hat{\mathbf{P}}) \qquad (3.19)$$
$$= ((\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k)vec(\hat{\mathbf{P}})$$

it follows from (3.16) that:

$$\begin{cases} \mathbf{S}_{T0}vec(\hat{\mathbf{P}}) + \mathbf{S}_{T1}vec(h\hat{\mathbf{Q}}) = \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h(t-t_0) \\ \mathbf{S}_{T1}vec(\hat{\mathbf{P}}) + \mathbf{S}_{T2}vec(h\hat{\mathbf{Q}}) = \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h^{(1)}(t-t_0) \end{cases} \qquad (3.20)$$

□.

Proof of Theorem 3.2: By taking iterative expectation, we get that:

$$E(T^{-1}\mathbf{S}_{Ti}) = \frac{1}{T}E(\sum_{t=s+1}^{T}(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k K_h^{(i)}(t-t_0))$$
$$= \frac{1}{T}\sum_{t=s+1}^{T}E[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]K_h^{(i)}(t-t_0)\frac{\mathbf{M}}{T}\int K_h^{(i)}(t-t_0)dt(1+o(1))$$
$$= \mu_i\mathbf{M}(1+o(1)).$$

Note that $T^{-1}vec(\mathbf{S}_{Ti}) = T^{-1}\sum_{t=1}^{T-s}\mathbf{Z}_t$, where $\mathbf{Z}_t = vec[(\mathbf{X}_{s+t}\mathbf{X}_{s+t}^T) \otimes I_k]K_h^{(i)}(s+t-t_0)$. It follows from stationarity that

$$Var(T^{-1}vec(\mathbf{S}_{Ti})) = \frac{T-s}{T^2}Var(\mathbf{Z}_1) + \frac{2(T-s)}{T^2}\sum_{l=1}^{T-s-1}(1-\frac{l}{T-s})cov(\mathbf{Z}_1, \mathbf{Z}_{l+1}) \qquad (3.21)$$

Let $d_T \longrightarrow \infty$ be a sequence of integers such that $d_T h \longrightarrow 0$. Define $J_1 = \sum_{l=1}^{d_T-1} \mid cov(\mathbf{Z}_1, \mathbf{Z}_{l+1}) \mid$ and $J_2 = \sum_{l=d_T}^{T-s-1} \mid cov(\mathbf{Z}_1, \mathbf{Z}_{l+1}) \mid$. Using the mixing condition (B6) and Davydov's lemma (see Hall and Heyde 1980, cor.A.2), we have: for components of $\mathbf{Z}_1$ and $\mathbf{Z}_{l+1}$,

$$| \, cov(\mathbf{Z}_{1,j}, \mathbf{Z}_{l+1,m}) \, | \leq C[\alpha(l)]^{1-2/\delta}[E|\mathbf{Z}_{1,j}|^{\delta}]^{1/\delta}[E|\mathbf{Z}_{l+1,m}|^{\delta}]^{1/\delta}.$$

where C is a generic constant. Directly calculating the mean and covariance, we establish that: $E|\mathbf{Z}_1|^{\delta} = O(h^{-\delta+1})$ and $|cov(\mathbf{Z}_1, \mathbf{Z}_{l+1})| = O(1)$, componentwise. Then $J_1 = O(d_T) = o(h^{-1})$ and $J_2 = O(h^{2/\delta-2}) \sum_{l=d_T}^{\infty}[\alpha(l)]^{1-2/\delta} = O(h^{2/\delta-2})$ $d_T^{-c} \sum_{l=d_T}^{\infty} l_{-c}[\alpha(l)]^{1-2/\delta} = o(h^{-1})$, if we set $h^{1-2/\delta}d_T^c = 1$, so that $d_T h \longrightarrow 0$ is satisfied. Thus,

$$\sum_{l=1}^{T-s-1} |cov(\mathbf{Z}_1, \mathbf{Z}_{l+1})| = J_1 + J_2 = o(h^{-1}) \tag{3.22}$$

Note that $var(\mathbf{Z}_1) = h^{-1}\nu_{2i}E[vec(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]^{\otimes 2}(1 + o(1))$. It follows from (3.21) and (3.22) that

$$Var(T^{-1}vec(\mathbf{S}_{Ti})) = \frac{1}{Th}\nu_{2i}E[vec(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]^{\otimes 2}(1 + o(1)) \tag{3.23}$$

By the Chebyshev inequality, we know:

$$T^{-1}\mathbf{S}_{Ti} = \mu_i\mathbf{M}(1 + o_p(1)) \tag{3.24}$$

Hence,

$$T^{-1} \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix} = U \otimes \mathbf{M}(1 + o_p(1)) \tag{3.25}$$

By (3.2) and (3.16), we have

$$\begin{bmatrix} vec(\hat{\mathbf{P}} - \mathbf{\Phi}(u)) \\ vec(h(\hat{\mathbf{Q}} - \mathbf{\Phi}'(u))) \end{bmatrix} = \mathbf{B}_T(u) + \mathbf{V}_T(u) \tag{3.26}$$

where

$$\mathbf{B}_T(u) = \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(t/T)\mathbf{X}_t K_h(t-t_0) \\ \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(t/T)\mathbf{X}_t K_h^{(1)}(t-t_0) \end{bmatrix} - \begin{bmatrix} vec(\mathbf{\Phi}(u)) \\ vec(h\mathbf{\Phi}'(u)) \end{bmatrix},$$

$$\mathbf{V}_T(u) = \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)\boldsymbol{\varepsilon}_t K_h(t-t_0) \\ \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)\boldsymbol{\varepsilon}_t K_h^{(1)}(t-t_0) \end{bmatrix} \equiv \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \mathbf{V}_{T0}^* \\ \mathbf{V}_{T1}^* \end{bmatrix}.$$

Thus $\mathbf{B}_T(u)$ and $\mathbf{V}_T(u)$ contribute to the bias and variance of the estimators, respectively. By the definition of $\mathbf{S}_{Ti}$ and (3.18), we have:

$$\begin{aligned} \mathbf{S}_{T0} vec(\mathbf{\Phi}(u)) &= \sum_{t=s+1}^{T} (\mathbf{X}_t^T \mathbf{X}_t) \otimes I_k vec(\mathbf{\Phi}(u)) K_h(t-t_0) \\ &= \sum_{t=s+1}^{T} \mathbf{X}_t^T \otimes (\mathbf{X}_t \otimes I_k) vec(\mathbf{\Phi}(u)) K_h(t-t_0) \end{aligned}$$

using (3.17), we obtain:

$$\begin{aligned} \mathbf{S}_{T0} vec(\mathbf{\Phi}(u)) &= \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(u)\mathbf{X}_t) K_h(t-t_0) \\ &= \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(u)\mathbf{X}_t K_h(t-t_0) \end{aligned}$$

Similarly,

$\mathbf{S}_{T1} vec(\mathbf{\Phi}(u) = \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(u)\mathbf{X}_t) K_h^{(1)}(t-t_0)$

$\mathbf{S}_{T1} vec(h\mathbf{\Phi}'(u) = \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)h\mathbf{\Phi}'(u)\mathbf{X}_t) K_h^{(1)}(t-t_0)$

$\mathbf{S}_{T2} vec(h\mathbf{\Phi}'(u) = \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)h\mathbf{\Phi}'(u)\mathbf{X}_t) K_h^{(2)}(t-t_0)$

which gives that:

$$\begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix} \mathbf{B}_T(u) \equiv \begin{pmatrix} \mathbf{B}^*_{T0} \\ \mathbf{B}^*_{T1} \end{pmatrix} \tag{3.27}$$

where $\mathbf{B}^*_{Tj} = \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)[\mathbf{\Phi}(t/T) - \mathbf{\Phi}(u) - \mathbf{\Phi}'(u)\frac{(t-t_0)}{T}]\mathbf{X}_t K_h^{(j)}(t-t_0)$, for $j = 0, 1$. By the Taylor expansion, it can be shown that:

$$\begin{aligned} T^{-1}\mathbf{B}^*_{T0} &= \frac{h^2}{2T} \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)\mathbf{\Phi}\text{"}(u)\mathbf{X}_t K_h^{(2)}(t-t_0) + o_p(h^2) \\ &= \frac{h^2}{2T} \sum_{t=s+1}^{T} [(\mathbf{X}_t^T \otimes \mathbf{X}_t) \otimes I_k] vec(\mathbf{\Phi}\text{"}(u)) K_h^{(2)}(t-t_0) + o_p(h^2) \\ &= \frac{h^2}{2}\mu_2 \mathbf{M} vec(\mathbf{\Phi}\text{"}(u)) + o_p(h^2) \end{aligned}$$

Similarly, $T^{-1}\mathbf{B}^*_{T1} = \frac{h^2}{2}\mu_3 \mathbf{M} vec(\mathbf{\Phi}\text{"}(u)) + o_p(h^2)$. This combined with (3.25) leads to:

$$\mathbf{B}_T(u) = \frac{h^2}{2}(U^{-1} \otimes \mathbf{M}^{-1})[\omega \otimes (\mathbf{M} vec(\mathbf{\Phi}\text{"}(u)))](1+o_p(1)) \frac{h^2}{2}(U^{-1}\omega) \otimes vec(\mathbf{\Phi}\text{"}(u))(1+o_p(1)) \tag{3.28}$$

where the second equation comes from the identity:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}) \tag{3.29}$$

Let $\mathbf{V}^*_T = (\mathbf{V}^{T*}_{T0}, \mathbf{V}^{T*}_{T1})^T$. Using the same argument as for (3.23), we obtain:

$$Var(\sqrt{T^{-1}h}\mathbf{V}^*_T) = \mathbf{V} \otimes E[(\mathbf{X}_t \otimes I_k)\gamma_t^{*2}(\mathbf{X}_t \otimes I_k)^T](1 + o(1)) \tag{3.30}$$

For any unit vector $\mathbf{d} = (\mathbf{d}_1^T, \mathbf{d}_2^T)^T \in \mathcal{R}^{2d}$, where $\mathbf{d}_1$ and $\mathbf{d}_2$ are $d \times 1$ vectors, we get:

$$\sqrt{h/T}\mathbf{d}^T\mathbf{V}_T^* = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-s}\sqrt{h}[\mathbf{d}_1^T(\mathbf{X}_{s+t}\otimes I_k)K_h(t+s-t_0)$$
$$+\mathbf{d}_2^T(\mathbf{X}_{s+t}\otimes I_k)K_h^{(1)}(t+s-t_0)]\boldsymbol{\varepsilon}_{s+t}$$
$$\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T-s}\mathbf{R}_{T,t}$$

By (3.30), we get: $Var(\sqrt{h/T}\mathbf{d}^T\mathbf{V}_T^*) = \mathbf{d}^T[\mathbf{V}\otimes\mathbf{N}]\mathbf{d}(1+o(1)) \equiv \boldsymbol{\theta}^2(1+o(1))$. where $\mathbf{N}$ is defined earlier. Applying the "big-block" and "small-block" argument (see the proof of Theorem 6.3, Fan and Yao 2003), we have: $\sqrt{h/T}\mathbf{d}^T\mathbf{V}_T^* \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\boldsymbol{\theta}^2)$. Therefore, by (3.25) and the $Cramér-Wold$ device, $\sqrt{Th}\mathbf{V}_T(u)$ is asymptotically normal with mean zero and variance-covariance matrix $\boldsymbol{\Sigma} = (U\otimes\mathbf{M})^{-1}(V\otimes\mathbf{N})(U\otimes\mathbf{M})^{-1}$, using the identity (3.29), we have: $\boldsymbol{\Sigma} = (U^{-1}VU^{-1})\otimes(\mathbf{M}^{-1}\mathbf{N}\mathbf{M}^{-1})$. Then the result of the theorem follows. $\Box$.

Before moving forward to the detailed proofs of Theorem 3.3, we need the following definitions and lemmas.

Let $h_T = 1/\sqrt{Th}$, for each $j = 1, 2..., k$,

$\boldsymbol{\beta}_j(t_0)^T = (\boldsymbol{\Phi}_j^*(t_0/T), h\boldsymbol{\Phi}^{*'}(t_0/T))$ and $\mathbf{Z}_t(t_0) = (\mathbf{X}_t^T, \dfrac{t-t_0}{hT}\mathbf{X}_t^T)^T$. Define:

$\boldsymbol{\alpha}_{Tj}(t_0) = h_T^2\boldsymbol{\Gamma}^{-1}\sum_{t=1}^T \varepsilon_{tj}^*\mathbf{X}_tK(\dfrac{t-t_0}{hT})$,

$\mathbf{R}_{Tj}(t_0) = h_T^2\sum_{t=1}^T\boldsymbol{\Gamma}^{-1}[\boldsymbol{\Phi}_{0j}^*\mathbf{X}_t - \boldsymbol{\beta}_j(t_0)^T\mathbf{Z}_t(t_0)]\mathbf{X}_tK(\dfrac{t-t_0}{hT})$,

$R_{T1}^j = \sum_{t=1}^T\varepsilon_{tj}^*\mathbf{R}_{Tj}(t)^T\mathbf{X}_t$,

$R_{T2}^j = \sum_{t=1}^T\boldsymbol{\alpha}_{Tj}(t)^T\mathbf{X}_t\mathbf{X}_t^T\mathbf{R}_{Tj}(t)$,

$R_{T3}^j = \dfrac{1}{2}\sum_{t=1}^T\mathbf{R}_{Tj}^T(t)\mathbf{X}_t\mathbf{X}_t^T\mathbf{R}_{Tj}(t)$.

Definition 3.1: (Definition 1 in de Jong(1987)). For each $j = 1, ...k$, W_T^j is called clean if the conditional expectations of W_{stj} vanish:

$$E(W_{stj}|\mathbf{X}_s) = 0, a.s.$$

for all $s, t \leq T$.

Lemma 3.1: (Lemma 7.2 in Fan, Zhang and Zhang(2001)). Under Assumption (C), for each $j = 1, 2, ...k$, as $h \longrightarrow 0$, $Th \longrightarrow \infty$. We have:

$$R_{T1}^j = \sqrt{T}h^2 R_{T10}^j + O(hT^{-1/2}),$$
$$R_{T2}^j = \sqrt{T}h^2 R_{T20}^j + O(hT^{-1/2}),$$
$$R_{T3}^j = Th^4 R_{T30}^j + O(h^3).$$

Furthermore, for any $\delta > 0$, for $j = 1, 2...k$, there exists $M_j > 0$, s.t: $P(|\frac{R_{Ti}^j}{\sqrt{T}h^2}| > M_j) \leq \delta$, for $i = 1, 2$ and $P(|\frac{R_{T3}^j}{Th^4}| > M_j) \leq \delta$.

Using Lemma 3.1, we can easily show the following lemma.

Lemma 3.2: (Lemma 7.3 in Fan, Zhang and Zhang(2001)). Let $\hat{\Phi}(t/T)$ be the local linear estimator $\hat{P}$ we derived from Lemma 3.1. Let $\hat{\Phi}^*(t/T) \equiv Q\hat{\Phi}(t/T) = (\hat{\Phi}_1^{T*}(t/T), ..., \hat{\Phi}_k^{T*}(t/T))^T$, then under the Assumption (C), uniformly for $t_0 \in (0, T)$, for each $j = 1, 2...k$, we have:

$$\hat{\Phi}_j^*(t_0/T) - \Phi_j(t_0/T) = (\alpha_{Tj}(t_0) + R_{Tj}(t_0))(1 + o_p(1)),$$

where $\alpha_{Tj}(t_0)$ and $R_{Tj}(t_0)$ are defined earlier. Again, we define:
$U_{Tj} = h_T^2 \sum_{t,s=1}^T \varepsilon_{tj}^* \varepsilon_{sj}^* \mathbf{X}_s^T \Gamma^{-1} \mathbf{X}_t K(\frac{t-s}{hT})$,
$V_{Tj} = h_T^4 \sum_{t,s=1}^T \varepsilon_{tj}^* \varepsilon_{sj}^* \mathbf{X}_t^T [\sum_{l=1}^T \Gamma^{-1} \mathbf{X}_l \mathbf{X}_l^T \Gamma^{-1} K(\frac{t-l}{Th}) K(\frac{s-l}{Th})] \mathbf{X}_s$.

Lemma 3.3: (Lemma 7.4 in Fan ,Zhang and Zhang(2001)). Under Assumption (C), assume $\varepsilon_t \sim \mathcal{N}(0, \Sigma_0)$, where $Q\Sigma_0 Q^T \equiv (\sigma_{ij}^2)_{i,j=1}^k$. As $h \longrightarrow 0$, $Th^{3/2} \longrightarrow \infty$, for $j = 1...k$, we have:

$U_{Tj} = \dfrac{D}{h}K(0)\sigma_{jj}^2 + \dfrac{1}{T}\sum_{s\neq t}^T \varepsilon_{sj}^*\varepsilon_{tj}^*\mathbf{X}_t^T\Gamma^{-1}\mathbf{X}_s K_h(s-t) + o_p(h^{-1/2})$,

$V_{Tj} = \dfrac{D}{h}\nu_0\sigma_{jj}^2 + \dfrac{2}{Th}\sum_{s<t}^T \varepsilon_{sj}^*\varepsilon_{tj}^*\mathbf{X}_s^T\Gamma^{-1}K * K(\dfrac{s-t}{hT})\mathbf{X}_t + o_p(h^{-1/2})$

with $K_h(\cdot) = \dfrac{1}{h}K\left(\dfrac{\cdot}{hT}\right)$.

Proof of Theorem 3.3: Firstly, we show that:

$$\begin{aligned}
\frac{\widetilde{RSS}_a}{T} &= \frac{1}{T}\sum_{t=1}^T e_{t1}^T\Sigma^{-1}e_{t1} \\
&= trace(\frac{\sum_{t=1}^T e_{t1}^T\Sigma^{-1}e_{t1}}{T}) \\
&= \frac{1}{T}\sum_{t=1}^T trace(e_{t1}^T e_{t1}^T\Sigma^{-1}) \\
&= \frac{1}{T}trace([\sum_{t=1}^T e_{t1}e_{t1}^T]\Sigma^{-1}) \\
&= \frac{1}{T}trace([T-1]\hat{\Sigma}_0\Sigma^{-1}) + o_p(1) \\
&= \frac{T-1}{T}trace(\hat{\Sigma}_0\Sigma^{-1}) + o_p(1) \\
&= trace(\hat{\Sigma}_0\Sigma^{-1}) + o_p(1).
\end{aligned}$$

Knowing that $\hat{\Sigma}_0 = \dfrac{1}{T-1}\sum_{t=1}^T e_{t1}e_{t1}^T$.

Secondly, by the definition, we obtain: for each $j = 1...k$,

$$\begin{aligned}
-\lambda_{Tj}(\Phi_0)\sigma_{jj}^2 = &-h_T^2\sum_{l=1}^T \varepsilon_{lj}^*[\sum_{t=1}^T \varepsilon_{tj}^*\mathbf{X}_t^T\Gamma^{-1}\mathbf{X}_l K(\frac{t-t_0}{hT})] \\
&+ \frac{1}{2}h_T^4\sum_{l=1}^T\sum_{t=1}^T\sum_{s=1}^T \varepsilon_{tj}^*\varepsilon_{sj}^*\mathbf{X}_t^T\Gamma^{-1}\mathbf{X}_l\mathbf{X}_l^T\mathbf{X}_s\Gamma^{-1}K(\frac{t-l}{hT})K(\frac{s-l}{hT})
\end{aligned}$$

$$- R_{T1}^j + R_{T2}^j + R_{T3}^j + O_p(\frac{1}{Th^2}).$$

Thus, we apply Lemma 3.1, Lemma 3.2 and Lemma 3.3. We find out:

$$-\lambda_{Tj}(\Phi_0) = \mu_T + d_{1Tj} - \frac{1}{2\sqrt{h}} W_T^j + o_p(h^{-1/2}). \tag{3.31}$$

where $W_T^j = \dfrac{\sqrt{h}}{T\sigma_{jj}^2} \sum_{s \neq t}^T \varepsilon_{tj}^* \varepsilon_{sj}^* [2K_h(s-t) - K_h * K_h(s-t)] \mathbf{X}_s^T \Gamma^{-1} \mathbf{X}_t$.

Now, we need to show that for all $j = 1, ...k$,

$$W_T^j \xrightarrow{L} N(0, w)$$

where $w = 2D\|2K - K * K\|_2^2$. Define:

$$W_{stj} = \frac{\sqrt{h}}{T} c_T(s,t) \varepsilon_{sj}^* \varepsilon_{tj}^* / \sigma_{jj}^2,$$

for $1 \leq s < t \leq T$ with $c_T(s,t)$ can be written as:

$$c_T(s,t) = b_1(s,t) + b_2(s,t) - b_3(s,t) - b_4(s,t)$$

where $b_1(s,t) = 2K_h(s-t)\mathbf{X}_s^T \Gamma^{-1} \mathbf{X}_t$, $b_2(s,t) = b_1(t,s)$;

$b_3(s,t) = K_h * K_h(s-t)\mathbf{X}_s^T \Gamma^{-1} \mathbf{X}_t$, $b_4(s,t) = b_3(t,s)$.

Hence $W_T^j = \sum_{s<t}^T W_{stj}$, for $j = 1...k$.

In order to employ Proposition 3.2 in de Jong(1987), we need to check the following conditions:

(1) W_T^j is clean.

(2) $var(W_T^j) \longrightarrow w.$ as $T \longrightarrow \infty$.

(3) G_I^j is of smaller order than $var(W_T^j)$.

(4) G_{II}^j is of smaller order than $var(W_T^j)$.

(5) G^j_{IV} is of smaller order than $var(W^j_T)$.

where

$G^j_I = \sum_{1\le s<t\le T} E(W^4_{stj})$,

$G^j_{II} = \sum_{1\le s<t<l\le T}[E(W^2_{stj}W^2_{slj}) + E(W^2_{tsj}W^2_{tlj}) + E(W^2_{lsj}W^2_{ltj})]$,

$G^j_{IV} = \sum_{1\le s<t<l<u\le T}[E(W_{stj}W_{slj}W_{utj}W_{ulj}) + E(W_{stj}W_{suj}W_{ltj}W_{luj}) + E(W_{slj}W_{suj}$ $W_{tlj}W_{tuj})]$.

Now we check each of the conditions above.

Condition (1) follows straightforwardly from the definition.

To verify (2), we notice that: $var(W^j_T) = \sum^T_{s<t} E(W^2_{stj})$. Denote:

$K(v, m) = K * \cdots * K(v)$ as the m-th convolution of $K(\cdot)$ at v for $m = 1, 2....$ Therefore it follows that:

$$E[c^2_T(s,t)\varepsilon^{*2}_{sj}\varepsilon^{*2}_{tj}] = \frac{D\sigma^4_{jj}}{h}[16K(0,2) - 16K(0,3) + 4K(0,4)](1 + O(h))$$

which leads to: $w = 2D\int[2K(x) - K * K(x)]^2dx = 2D\|2K - K * K\|^2_2$.

Condition (3) is satisfied by noting that for each $j = 1, 2...k$,

$$E[b_1(1,2)\varepsilon^*_{1j}\varepsilon^*_{2j}]^4 = O(h^{-3})b_3(1,2)\varepsilon^*_{1j}\varepsilon^*_{2j}]^4 = O(h^{-2}).$$

which implies that $E(W^4_{12j}) = \frac{h^2}{T^4}O(h^{-3})$. Thus, $G^j_I = O(T^{-2}h^{-1}) = o(1)$.

Condition (4) is verified by the following calculation:

$$E(W^2_{12j}W^2_{13j}) = O(E(W^4_{12j})) = O(T^{-4}h^{-1})$$

which gives that: $G^j_{II} = O(T^{-1}h^{-1}) = o(1)$ for all $j = 1, 2...k$.

To prove condition (5), it suffices to compute the term $E(W_{12j}W_{23j}W_{34j}W_{41j})$.

By direct calculations:

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_1(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_1(1,2)b_1(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_1(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_3(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

and similarly for the other terms. Hence,

$$E(W_{12j}W_{23j}W_{34j}W_{41j}) = T^{-4}h^2O(h^{-1}) = O(T^{-4}h)$$

which leads to: $G_{IV}^j = O(T^4T^{-4}h) = O(h) = o(1)$.

By now, we have shown for each $j = 1, 2...k$, we have, under H_0:

$$\sigma_T^{-1}(\lambda_{Tj}(\Phi_0) - \mu_T + d_{1Tj}) \xrightarrow{L} \mathcal{N}(0,1)$$

where $\sigma_T^2 = \dfrac{D}{2h}\int(2K(x) - K * K(x))^2dx$,
$\mu_T = \dfrac{D}{2h}(2K(0) - \int K^2(x)dx)$,
$d_{1Tj} = \sigma^{-2}[Th^4R_{T30}^j - T^{1/2}h^2(R_{T10}^j - R_{T20}^j)] = O_p(Th^4 + \sqrt{T}h^2)$.

From the definition, we get our GLR test statistic, under H_0:

$$\begin{aligned}
\lambda_T(\Phi_0) &\approx \frac{T}{2}\frac{R\tilde{S}S_0 - R\tilde{S}S_a}{R\tilde{S}S_a} \\
&\approx \frac{R\tilde{S}S_0 - R\tilde{S}S_a}{2trace(\hat{\Sigma}_0\Sigma^{-1})} \\
&= \frac{1}{2trace(\hat{\Sigma}_0\Sigma^{-1})}\sum_{j=1}^{k}\lambda_j(RSS_0 - RSS_a)(j) \\
&= \frac{1}{trace(\hat{\Sigma}_0\Sigma^{-1})}\sum_{j=1}^{k}\lambda_j\sigma_{jj}^2\lambda_{Tj}(\Phi_0)
\end{aligned}$$

We already shown that $var(\lambda_{Tj}(\Phi_0)) = \sigma_T^2(1 + O(1))$.

Let

$$\mu_T^* = \frac{\mu_T}{trace(\hat{\Sigma}_0\Sigma^{-1})}\sum_{j=1}^{k}\lambda_j\sigma_{jj}^2,$$
$$d_{1T}^* = \frac{1}{trace(\hat{\Sigma}_0\Sigma^{-1})}\sum_{j=1}^{k}\lambda_j\sigma_{jj}^2 d_{1Tj}$$

Now, we focus on the variance of $\lambda_T(\Phi_0)$, we have:

$$var[\lambda_T(\Phi_0)] = var[\frac{1}{trace(\hat{\Sigma}_0\Sigma^{-1})}\sum_{j=1}^{k}\lambda_j\sigma_{jj}^2\lambda_{Tj}(\Phi_0)] = \frac{\sigma_T^2}{trace(\hat{\Sigma}_0\Sigma^{-1})^2}\sum_{j=1}^{k}\lambda_j^2\sigma_{jj}^4$$
$$+ \frac{2}{trace(\hat{\Sigma}_0\Sigma^{-1})^2}\sum_{1\leq i<j\leq k} cov(\lambda_j\sigma_{jj}^2\lambda_{Tj}(\Phi_0), \lambda_i\sigma_{ii}^2\lambda_{Ti}(\Phi_0))$$

Since $-\lambda_{Tj}(\Phi_0) = \mu_T + d_{1Tj} - \frac{1}{2\sqrt{h}}W_T^j + o_p(h^{-1/2})$, for $i < j$, we obtain:

$$cov(\lambda_i\sigma_{ii}^2\lambda_{Ti}(\Phi_0), \lambda_j\sigma_{jj}^2\lambda_{Tj}(\Phi_0)) = \frac{\lambda_i\lambda_j\sigma_{ii}^2\lambda_{jj}^2}{4h}cov(W_T^i, W_T^j)$$
$$= \frac{\lambda_{ij}\sigma_{ii}^2\lambda_{jj}^2}{4h}E(W_T^i W_T^j),$$

Similar with the calculation of $var(W_T^j)$, we obtain:

$$E(W_T^i W_T^j) = E[(\sum_{s<t} W_{ti})(\sum_{s<t} W_{stj})]$$
$$= \sum_{1\leq s<t\leq T} E(W_{sti}W_{stj})$$

$$= \sum_{1 \leq s < t \leq T} E[\frac{hc_T^2(s,t)}{T^2 \sigma_{ii}^2 \sigma_{jj}^2} \varepsilon_{si}^* \varepsilon_{ti}^* \varepsilon_{sj}^* \varepsilon_{tj}^*]$$

We note that:

$$E(c_T^2(s,t) \varepsilon_{si}^* \varepsilon_{ti}^* \varepsilon_{sj}^* \varepsilon_{tj}^*) = \frac{D\sigma_{ij}^4}{h}[16K(0,2) - 16K(0,3) + 4K(0,4)](1 + O(h)).$$

Therefore, we have:

$$var(\lambda_T(\Phi_0) = \frac{\sigma_T^2}{trace(\hat{\Sigma}_0 \Sigma^{-1})^2}[\sum_{j=1}^{k} {}_j^2 \sigma_{jj}^4 + 2\sum_{i<j}^{k} \lambda_i \lambda_j \sigma_{ij}^4](1 + O(1)).$$

Denote: $\sigma^{*2} \equiv \dfrac{\sigma_T^2}{trace(\hat{\Sigma}_0 \Sigma^{-1})^2}[\sum_{j=1}^k \lambda_j^2 \sigma_{jj}^4 + 2\sum_{i<j}^k \lambda_{ij} \sigma_{ij}^4]$. Then:

$$var(\lambda_T(\Phi_0)) \longrightarrow \sigma^{*2}.$$

where $\sigma_T^2 = \dfrac{2D}{h} \int (K(x) - \dfrac{1}{2} K * K(x))^2 dx.$

Notice that, $\lambda_T(\Phi_0) - \mu_T^* + d_{1T}^*$ is clean. In order to apply Proposition 3.2 in de Jong (1987) again, it remains to check:

(1') F_I^{ij} is of smaller order than $E(W_T^i W_T^j)$

(2') F_{II}^{ij} is of smaller order than $E(W_T^i W_T^j)$

(3') F_{IV}^{ij} is of smaller order than $E(W_T^i W_T^j)$

where

$F_I^{ij} = \sum_{1 \leq s < t \leq T} E(W_{sti}^2 W_{stj}^2)$,

$F_{II}^{ij} = \sum_{1 \leq s < t < l \leq T} [E(W_{sti} W_{sli} W_{stj} W_{slj}) + E(W_{tsi} W_{tli} W_{tsj} W_{tlj}) + E(W_{lsi} W_{lti} W_{lsj} W_{ltj})]$,

$F_{IV}^{ij} = \sum_{1 \leq s < t < l < u \leq T} [E(W_{sti} W_{slj} W_{utj} W_{uli}) + E(W_{sti} W_{suj} W_{ltj} W_{lui}) + E(W_{sli} W_{suj}$
$W_{ltj} W_{lui}) + E(W_{sli} W_{suj} W_{tlj} W_{tui})]$.

Condition (1') holds because $E(W_{12i}^2 W_{12j}^2) = O(E(W_{12i}^4)) = O(T^{-4} h^{-1})$, Hence,

$F_I^{ij} = O(T^{-2}h^{-1}) = o(1)$.

To prove (2'), note that:

$$
\begin{aligned}
E(W_{12i}W_{13i}W_{12j}W_{13j}) &= O(E(W_{12i}^2 W_{13i}^2)) \\
&= O(E(W_{12i}^4)) = O(T^{-4}h^{-1})
\end{aligned}
$$

Thus, $F_{II}^{ij} = O(T^{-1}h^{-1}) = O(1)$.

To prove (3'), it suffices to calculate the term $E(W_{12i}W_{23j}W_{34i}W_{41j})$. By straightforward calculations:

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_1(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_1(1,2)b_1(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_1(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_3(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$

Then, $E(W_{12i}W_{23j}W_{34i}W_{41j}) = \dfrac{h^2}{T^4}O(h^{-1}) = O(T^{-4}h)$. yielding

$$F_{IV}^{ij} = O(h) = o(1),$$

Therefore, we have shown that $var(\lambda_T(\Phi_0))$ has been dominated by σ^{*2}. Hence,

$$\sigma^{*-1}(\lambda_T(\Phi_0) - \mu_T^* + d_{1T}^*) \xrightarrow{D} \mathcal{N}(0,1)$$

This finishes the proof. $\square$.

Proof of Theorem 3.4: Theorem 3.4 is one special case of Theorem 3.3 when Φ_0^* under H_0 is a vector of constants. So, with the same notation as in the proof of Theorem 3.3, we have for each $j = 1, ...k$, $\Phi_0^{*"} = 0$. Hence, $R_{T10}^j = R_{T20}^j = R_{T30}^j = 0$

which leads to each $d_{1Tj} = 0$ and $d_{1T}^* = 0$. The rest of the proof is the same as the proof Theorem 3.3. □.

Proof of Theorem 3.5: Under H_a and Assumption C, applying Theorem 3.3, we have: for each $j = 1, ...k$,

$-\lambda_{Tj}(\Phi_0) = -\mu_T + \nu_{Tj} - d_{2T} - [\frac{1}{2\sqrt{h}} W_T^j + \sum_{t=1}^{T} G^T(u)(I_k \otimes \mathbf{X}_t)\varepsilon_t/(\sigma_{jj}^2)] + o_p(h^{-1/2})$.

where $\nu_{Tj} = \frac{Th^4}{8\sigma_{jj}^2} E\{\mathbf{\Delta}^{"T}(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{\Delta}^{"}(u)\}\omega_0$, $d_{2T} = \frac{T}{2} E\{\mathbf{G}^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}(u)\}$ with μ_T and W_T^j defined in the proof of Theorem 3.3. The rest of the proof is similar to the proof of Theorem 3.3. The details are omitted. □.

CHAPTER 4: CONCLUSION

In this dissertation, first of all,I define the spatial QR and study spatial quantile regression estimation of multivariate threshold time series models. I derive asymptotic normality of the proposed estimator. I conduct simulations and analyze a real example to show the performance of the proposed estimator.

Furthermore, I extend the multivariate threshold time series model to multivariate time-varying coefficient model. I also get an explicit representation of the estimator of the time-varying coefficient using local linear technique. Asymptotic normality is established as well.

Last but not the least, I propose the new test statistic which is built based on the comparison of the likelihood under between null and alternative hypotheses. I give the theoretical asymptotic null and alternative distributions. Monte Carlo simulations are conducted to illustrate the power of the proposed test procedure and an application to a real data set is presented too.

There are still many interesting research topics related to this dissertation which deserve further investigation. First, one may release the stationary or mixing conditions. I only focus on the asymptotic result under stationary time series data setting. Secondly, the generalized quasi-likelihood ratio test statistic can be extended to other models. For example, additive models, predictive regression models and so on. Last but not the least,few paper available in literature about multivariate time-varying coefficient models under nonstationary time series setting due to the difficulty of deriving explicit representation for the nonstationary data. All of the above issues can should be given a lot of attention as a future research topic.

REFERENCES

[1] Bronchitic, E. 1985, Robust Model Selection in Regression. Statistics &Probability Letters, 3, 21-23.

[2] Barnes, M and Hughes, A.2002. A Quantile Regression Analysis of the Cross Section of Stock Market Returns, working paper, Federal Reserve Bank of Boston.

[3] Bassett, G and Chen, H.2001. Portfolio Style: Return-Based Attribution Using Quantile Regression, Empirical Economics, Springer-Verlag, pp. 1405C1441.

[4] Bollerslev, T. 1990. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. Review of Economics and Statistics, 72, 498-505.

[5] Chakraborty, B. 2003. On multivariate quantile regression. Journal of Statistical Planning and Inference, 110, 109-132.

[6] Chaudhuri, P. 1996. On a geometric notion of quantiles for multivariate data. Journal of American Statistical Association, 91, 862-871.

[7] Chen, R. and Tsay, R. S. 1993. Functional coefficient autoregressive models. Journal of American Statistical Association, 88, 298-308.

[8] Cai, Z., Fan, J. and Li, R. 2000. Efficient estimation and inferences for varying-coefficient models. J. Amer. Statist. Assoc. 95 888C902. MR1804446.

[9] Cai, Z. 2002a, Regression quantile for time series. Econometric Theory 18, 169-192.

[10] Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941-956.

[11] Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for Dynamic smooth coefficient models. Journal of the American Statistical Association. 103, 1595-1608.

[12] Cai, Z., Das, M., Xiong, H.,Wu, X.,2006. Functional coefficient instrumental variables models. Journal of Econometrics. 133, 207-241.

[13] Chaudhuri, P., 1991. Nonparametric quantile estimates of regression quantiles and their local Bahadur representation. The Annals of Statistics 19, 760-777.

[14] Davis, R. A. and Dunsmuir, W. T. M, 1997. Least absolute deviation estimation for regression with ARMA errors, J. Theor. Prob, 10, 481-497.

[15] Engle, R. andManganelli, S, 2004. CAVaR: Conditional Autoregressive Value at Risk by Regression Quantiles, Journal of Business and Economic Statistics.22(4), 367-381

[16] El Banter, F. and M. Hallin, 1999, L_1 Estimation in linear models with heterogeneous white noise, Statistics and Probability Letters, 45, 305-315.

[17] Fan, J., Yao,Q., and Tong. H. 1996, Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems, Biometrika,83,189-206.

[18] Fan, J., Zhang, C., Zhang, J., 2001. Generalized Likelihood Ratio Statistics and Wilks Phenomenon. The Annuals of Statistics.29,153-193.

[19] Fan, J. and Gijbels, I, 1996. Local Polynomial Modeling and Its Applications. Chapman and Hall, London.

[20] Fan,J. and Jiang,J, 2005. Nonparametric inferences for additive models. Journal of the American Statistical Association, 100,891-907

[21] Gourieroux, C., Monfort, A. and Renault, E. 1987, Consistent M-Estimators in a Semi- Parametric Model, Document de Travail de lINSEE, 8706.

[22] Horowitz, J. L. and Spokpiny, V. G. 2002. An adaptive, rate-optimal test of linearity for median regression models. J. Amer. Statist. Assoc. 97 822C835. MR1941412

[23] He, X. and Zhu, L.X. 2003. A lack-of-fit test for quantile regression. J. Amer. Statist. Assoc. 98 1013C1022. MR2041489

[24] Hastie, T. J. and Tibshirani, R. J. 1993. Varying-coefficient models. J. Roy. Statist. Soc. B. 55 757C796.

[25] He, X. Shi,P., 1996. Bivariate tensor-product B-splines in a partly linear model. Journal of Multivariate Analysis 58, 162-181.

[26] He,X., Ng, P., and Portnoy, S. 1998. Bivariate Quantile Smoothing Splines. Journal of the Royal Statistical Society. Ser. B, 60. 537-550.

[27] Hencricks, W. and Koenker, R. 1992. Hierarchical spline models for conditional quantiles and the demand for electricity. J. Amer. Statist. Assoc. 87 58C68.

[28] Honda,T. 2004. Quantlie Regression in Varying Coedfficient Models. Journal of Statistical Planning and Inference, 121, 113-125.

[29] Harrison and Rubinfeld, 1978, Hedonic prices and the demand for clean air, Journal of Environmental Economics and Management, Vol. 5, pp. 81-102.

[30] Horowitz, J. L.,and Lee, S. 2005. Nonparametric Estimation of an Additive Quantile Regression Model.Journal of the American Statistical Association, 100, 1238-1249.

[31] Jiang, J., Zhao, Q. and Hui, Y. V, 2001. Robust modlling of ARCH models. Journal of Forecasting, 20, 111-133.

[32] Koenker, R. and G. Bassett 1982a Robust Tests for Heteroscedasticity Baxed on Regression Quantiles, Econometrica, 50, 43-61.

[33] Koenker, R. and G. Bassett. 1982b. Test of Linear Hypotheses and L_1 Estimation, Econometrica, 50, 1577-1584.

[34] Koenker, R., 2004. Quantreg: An R Package for Quantile Regression and Related Methods. http://cran.r-project.org.

[35] Koenker, R., 2005. Quantile Regression. Econometric Society Monograph Series, Cambridge University Press, New York.

[36] Koenker, R., Bassett, G.W.,1978. Regression Quantiles. Econometrica 46, 33-50.

[37] Koenker, R., and Xiao, Z, 2004, Unit Root Quantile Autoregression Inference, Journal of the American Statistical Association,99,775-787.

[38] Komunjer, I., 2005. Quasi-maximum likelihood estimation for conditional quantiles. Journal of Econometrics 128, 137-164.

[39] Koenker, R. and Bassett, G. 1978. Regression quantiles. Econometrica, 46, 33-50.

[40] Koenker, R. and J. Machado. 1999, Goodness of Fit and Related Inference Processes for quantile Regression. Journal of the Royal Statistical Society, Series B,66, 145-163

[41] Koltchinskii, V. 1997. M-estimation, convexity and quantiles. The Annals of Statistics, 25, 435-477.

[42] Koul, H. L. and Saleh, A. K. Md. E. 1995. Autoregression quantiles and related rank scores processes. Ann. Statist, 23, 670-689.

[43] ao Pan, J. and Yao, Q, 2008. Modelling multiple time series via common factors. Biometrika, 95, 365-379.

[44] Peng, L. and Yao, Q. 2003, Least absolute deviation estimation for ARCH and GARCH models. Biometrika, 90, 967-975.

[45] Rose, S.J. 1992. Social Stratification in the United States. The New Press, New York.

[46] Serfling, R, 2004. Nonparametric multivariate descriptive measures based on spatial quantiles. J. Statist. Plann. Inference, 123, 259-278.

[47] Tsay, R. S, 1989, Testing and modeling threshold autoregressive processes. Jour. Amer. Statist. Assoc, 84, 231-240.

[48] Tsay, R. S, 1998, Testing and modeling multivariate threshold models. Jour. Amer. Statist. Assoc, 93, 1188-1202.

[49] Tong, H. 1983. Threshold models in non-linear time series analysis, New York: Springer.

[50] Tong, H 1990, Nonlinear Time Series: A Dynamical System Approach Oxford, U.K.: Oxford University Press.

[51] Tong, H. and Lim, K.S, 1980, Threshold autoregressive, limit cycles and cyclical data. J. Roy. Stat. Soc. B, 42, 245-292.

[52] Tsay, R. S, 1989. Testing and modeling threshold autoregressive processes. Journal of the American Statistical Association 84, 231-240.

[53] Wu, T., Yu, K. and Yu,Y. 2010. Single-index quantile regression. Journal of Multivariate Analysis. 101, 1607-1621.

[54] Zhang, C. 2000, Topics in Generalized Likelihood Ratio Test. Phd dissertation, The University of North Carolina at Chapel Hill.