Generalized Quantile Regression

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
M. Sc Mengmeng Guo

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Dr. Ulrich Kamecke

Gutachter:

1. Prof. Dr. Wolfgang Härdle
2. Prof. Ph.D Jianhua Huang

Tag des Kolloquiums: 01 August 2012

Abstract

Generalized quantile regressions, including the conditional quantiles and expectiles as special cases, are useful alternatives to the conditional means for characterizing a conditional distribution, especially when the interest lies in the tails. We denote $v_n(x)$ as the kernel smoothing estimator of the expectile curves. We prove the strong uniform consistency rate of $v_n(x)$ under general conditions. Moreover, using strong approximations of the empirical process and extreme value theory, we consider the asymptotic maximal deviation $\sup_{0\leqslant x\leqslant 1} |v_n(x) - v(x)|$. According to the asymptotic theory, we construct simultaneous confidence bands around the estimated expectile function. We develop a functional data analysis approach to jointly estimate a family of generalized quantile regressions. Our approach assumes that the generalized quantiles share some common features that can be summarized by a small number of principal components functions. The principal components are modeled as spline functions and are estimated by minimizing a penalized asymmetric loss measure. An iteratively reweighted least squares algorithm is developed for computation. While separate estimation of individual generalized quantile regressions usually suffers from large variability due to lack of sufficient data, by borrowing strength across data sets, our joint estimation approach significantly improves the estimation efficiency, which is demonstrated in a simulation study. The proposed method is applied to data from 150 weather stations in China to obtain the generalized quantile curves of the volatility of the temperature at these stations.

Keywords: Asymmetric loss function; Functional data analysis; Generalized quantile curve; Iteratively reweighted least squares; simultaneous confidence bands.

Zusammenfassung

Die generalisierte Quantilregression, einschließlich der Sonderfälle bedingter Quantile und Expektile, ist insbesondere dann eine nützliche Alternative zum bedingten Mittel bei der Charakterisierung einer bedingten Wahrscheinlichkeitsverteilung, wenn das Hauptinteresse in den Tails der Verteilung liegt. Wir bezeichnen mit $v_n(x)$ den Kerndichteschätzer der Expektilkurve und zeigen die stark gleichmßige Konsistenzrate von $v_n(x)$ unter allgemeinen Bedingungen. Unter Zuhilfenahme von Extremwerttheorie und starken Approximationen der empirischen Prozesse betrachten wir die asymptotischen maximalen Abweichungen $\sup_{0\leqslant x\leqslant 1}|v_n(x) - v(x)|$. Nach Vorbild der asymptotischen Theorie konstruieren wir simultane Konfidenzbänder um die geschätzte Expektilfunktion. Wir entwickeln einen funktionalen Datenanalyseansatz um eine Familie von generalisierten Quantilregressionen gemeinsam zu schätzen. Dabei gehen wir in unserem Ansatz davon aus, dass die generalisierten Quantile einige gemeinsame Merkmale teilen, welche durch eine geringe Anzahl von Hauptkomponenten zusammengefasst werden können. Die Hauptkomponenten sind als Splinefunktionen modelliert und werden durch Minimierung eines penalisierten asymmetrischen Verlustmaßes geschätzt. Zur Berechnung wird ein iterativ gewichteter Kleinste-Quadrate-Algorithmus entwickelt. Während die separate Schätzung von individuell generalisierten Quantilregressionen normalerweise unter großer Variablität durch fehlende Daten leidet, verbessert unser Ansatz der gemeinsamen Schätzung die Effizienz signifikant. Dies haben wir in einer Simulationsstudie demonstriert. Unsere vorgeschlagene Methode haben wir auf einen Datensatz von 150 Wetterstationen in China angewendet, um die generalisierten Quantilkurven der Volatilität der Temperatur von diesen Stationen zu erhalten.

Schlagwörter: asymmetrischen Verlustmaßes Funktionen; generalisierte Quantilregression; funktionalen Datenanalyseansatz; iterativ gewichteter Kleinste-Quadrate-Algorithmus; simultane Konfidenzbänder.

Acknowledgements

Firstly, I would like to express my deep gratitude and respect to my supervisor Professor Dr. Wolfgang Karl Härdle, who cautiously supervised me during my PhD period, and let me obtain systemic training. I also would like to thank my second supervisor Professor Ph.D Jianhua Huang, who gave me a lot of scientific advice on my project. Without the support, patience and guidance of them, this study would not have been completed.

I also would thank Prof. Dr. Brenda López Cabrera, Prof. Dr. Ostap OKhrin, Weining Wang and Prof. PhD Lan Zhou for their help in my research. Meanwhile, I also appreciate the scientific support and friendship from all of my colleagues in Ladislaus von Bortkiewicz Chair of Statistics.

I gratefully acknowledge the financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin and China Scholarship Council (CSC).

Last but certainly not the least I deeply indebted to my family for having been there whenever I needed them, especially the continues support from my grandparents, my parents, my sister and my brother. I would like also thank all my friends for their continuous encouragement to make me never give up.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The thesis includes three topics I was working on during my PhD studies.

Chapter 2 investigates the stochastic process of interest rate. As we know, the interest have been well studied by a lot of researcher, therefore there are already several methods to capture the stochastic process of interest rate. We recall several classical models, such as, Vasicek model, CIR model, or jump diffusion model. Interest rate modelling in an unstable macroeconomic context motivates interest rate models with time varying parameters. In this paper, the local parametric approach is introduced to adaptively estimate interest rate models. This method can be generally used in time varying coefficient parametric models. It is used not only to detect jumps and structural breaks, but also to choose the largest "time homogeneous" interval, in which the coefficients are statistically constant. We describe this adaptive approach in detail, illustrate it in simulations and apply it to real data. Using the three month treasury bill rate as a proxy of the short rate, we find that our method can detect both structural changes and stable intervals for homogeneous modelling of the interest rate process. In unstable macroeconomic periods, the time homogeneous interval can not last long. Furthermore, the proposed approach performs well in the long horizon forecasting.

Chapter 3 and 4 focus on the analysis of the tails of the distribution functions. We investigate both conditional quantile and expectile curves, which together are named as generalized quantile curves. In Chapter 3, the expectile curves are studied. We denote $v_n(x)$ as the kernel smoothing estimator of the expectile curves. We prove the strong uniform consistency rate of $v_n(x)$ under general conditions. Moreover, using strong approximations of the empirical process and extreme value theory, we consider the asymptotic maximal deviation $\sup_{0\leqslant x\leqslant 1}|v_n(x) - v(x)|$. According to the asymptotic theory, we construct simultaneous confidence bands around the estimated expectile function. Furthermore, we apply the confidence bands to the temperature analysis, i.e. we separately construct the confidence bands for the expecitle curves of the volatility of temperature in Berlin and Taipei,

we investigate the temperature risk drivers to these two cities.

In Chapter 4, We introduce the generalized quantile curves, which includes both quantile curves and expectile curves. They can transfer to each other via a transformation weight. Traditional generalized quantile regression focuses on a single curve, as described in Chapter 3. When a family of random curves are available, we can estimate the individual generalized quantile curves jointly by using the information from all subjects instead of estimating them separately. In reality, high dimensional data is always involves in the research. For example, in metrology, we might want to analyze the temperature fluctuations in China. As known, China has more than 150 weather stations allocated in different locations. Then the temperature data collected over time at each station are effectively producing a curve over the observed interval, with, say, 365 measurements made over 365 days of the year. We are interested in the extreme values of the temperature during some specific year, which in turn lead us to study the generalized quantile curves of the variation of the temperature. On one hand, to estimate the generalized quantile curves precisely, one direct approach is to collect all the information from all the stations. On the other hand, to avoid too many parameters to estimate and data sparsity, we apply a novel method – functional data analysis (FDA) combining least asymmetric weighted squares (LAWS), we estimate both the mean curve as the common factor curve and the individual departure curves of the generalized quantile curves via a penalized spline smoothing. We run both simulations and real data analysis to investigate the performance of the FDA method in comparison with the traditional single curve estimation method. Taking the temperature as an example, we estimate the generalized quantile curves for the volatility of the temperature in 150 weather stations in China to analyze the factors influencing the variation of the temperature in China.

Chapter 2

Adaptive Interest Rate Modeling

This chapter is based on the paper " Adaptive Interest Rate Modeling" by Guo and Härdle (2010).

2.1 Introduction

Interest rate is one of the key financial variables in any economy. As interest rates rise, bond prices fall and vice versa. Interest rate risk, caused by the variability of interest rates, is the risk borne by an interest-bearing asset, such as a loan or a bond. Moreover, interest rate risk causes credit risk, which in turn may induce further risks (for instance, financial crisis). For hedging purposes, it is crucial to price interest rate derivatives which of course depend on the dynamics of interest rates. As described in literature, if the macroeconomy is unstable, the volatility of interest rates will be larger, and vice versa. For instance, in 2002, bubbles existed in the US stock market, in 2003, the war in Iraq influenced the macroeconomy. In 2007, the macroeconomy has changed due to the subprime crisis. Correspondingly one finds the interest rate in these periods fluctuating strongly. On the other hand, changes in business cycle conditions or macroeconomic shocks may affect the dynamics of interest rates. These shocks or news are dominated by central bank announcements or from federal agencies, who release macroeconomic data at monthly or quarterly frequencies. They may contain a large, unanticipated component. However, the interest rates respond quickly to these unanticipated announcements. The corresponding findings are well documented in Jones et al. (1998) and Johannes (2004). Due to these unanticipated announcements and shocks in the macroeconomy, numerous empirical studies have demonstrated the dynamics of short rate processes not being stable. This may be reflected by time varying parameters in interest rate models. In this paper, we study the dynamics of the short rate, which can respond quickly to news and shocks. The short rate is the (annualized) interest

rate at which an entity can borrow money for an infinitesimally short period of time, it is the instantaneous-return rate of a risk-free investment.

Three main strands of literature exist to capture the instability of the dynamics of the short rate. In one line of literature, the described instability is modeled via structural breaks, captured by jump diffusion models. In this kind of models, it is assumed several unknown jumps exist in the dynamics of the short rate. For instance, Das (2002) incorporated jumps into the Vasicek (1977) model and found strong evidence of jumps in the daily federal funds rate. Johannes (2004) used a nonparametric diffusion model to study the secondary three month treasury bills and concluded that jumps are generally generated by the arrival of news about the macroeconomy. A general conclusion in the literature is that the dynamics vary significantly due to shocks and jumps, which is also well described in Lettau and Ludvigson (2001), Goyal and Welch (2003) and Paye and Timmermann (2006). Another strand of literature uses regime switching models to capture the business cycle character of interest rates, see Ang and Bekaert (2002), Bansal and Zhou (2002). They found that the interest rate has changed significantly and its volatility performs differently in expansion regimes and recession regimes. In a third kind of models, the process parameters (drift or volatility) are assumed to be functions of time. This is well documented in numerous studies, such as Hull and White (1990), Black and Karasinski (1991), A¨**i**t-Sahalia (1996), Stanton (1997), Fan et al. (2003) and Arapis and Gao (2006). For instance, using semi and nonparametric approaches, A¨**i**t-Sahalia (1996) found strong nonlinearity in the drift function of the interest rate model. Arapis and Gao (2006) applied nonparametric techniques to provide evidence that the specification of the drift has a considerable impact on the pricing of derivatives through its effect on the diffusion function. As a conclusion from these findings, one may say that the coefficients in the models, such as Vasicek (1977) model and Cox, Ingersoll and Ross (CIR) (1985) model, are time varying, especially with a nonlinear drift function. Thus, a short rate model with constant parameters may not be valid for a long time period.

In this paper, we introduce the time-varying CIR model, and estimate it from a novel point of view - the local parametric approach (LPA). Before we come to our method in detail, let us first review several methods to capture the time variation of the parameters. They are commonly used to estimate time-varying coefficients or to detect the breakpoints. We list out some, for instance, the wavelets method, see Fan and Wang (2007), or the kernel based estimation, see Cai (2007), as well as the taut string methods, see Davies and Kovac (2001), and the online checking change point used in Ombao et al. (2001). However, the local parametric approach can determine a homogeneous interval for each time point. What's more, by this method, we can detect jumps and structural break points in the stochastic process, which indicates that the commonly used regime switching models are

also included in our method as well. Based on the parameters inside the selected interval, one may distinguish expansion and recession stages of the macroeconomy. Moreover, the LPA has several nice properties. Firstly, it can describe the smooth time-varying parameters. The coefficients can arbitrarily depend on time, such as smooth time trend. Secondly, it allows for structural breaks and jumps in the parameter values. Thirdly, there is no requirement on the number of observations in a homogeneous interval before or after the break point.

The proposed approach can be applied to different real problems. Giacomini et al. (2009) considered time varying copulae estimation, C´ižek et al. (2009) applied it to compare the performance of global and time-varying ARCH and GARCH specifications, Härdle et al. (2011) applied this method to hierarchical Archimedean copulae, and found the LPA can be used to detect both adaptive copulae parameters and local dependency structures.

To assess the performance of the LPA, we do both simulations and empirical studies. In the simulation exercise, we show that the proposed LPA detects the structural breaks very well, and all the true parameters are located in the pointwise confidence intervals of the estimators. In the empirical study, we use the three month treasury bill rate as a proxy of the short rate and investigate the performance of the LPA to the time-varying CIR model by both in sample fitting and out of sample forecasting via comparing with moving window estimators.

The chapter is organized as follows. In Section 2.2, we give a short recall about standard interest rate models, later we explain the LPA in detail in Section 2.3. In Section 2.4, we present our simulation results. Empirical studies are presented in Section 4.5. We conclude in Section 4.6.

2.2 Interest Rate Models

In this section, we recall several standard short rate models. In general, the dynamics describing short rate processes should satisfy the properties:

1. Mean reversion (the interest rates always tend to return to an average level).

2. The interest rate is nonnegative.

Vasicek Model (1977)

$$dr(t) = a\{b - r(t)\}dt + \sigma dW_t$$

where a, b and σ are constants, W_t is a standard Brownian process. It is consistent with the mean reversion feature with a reversion speed a to the long run mean level b. However, in this model $r(t)$ can be negative.

Cox, Ingersoll and Ross (CIR) Model (1985)

$$dr(t) = a\{b - r(t)\}dt + \sigma\sqrt{r(t)}dW_t \tag{2.1}$$

The drift function $\mu\{r(t)\} = a\{b - r(t)\}$ is linear and possesses a mean reverting property, i.e. $r(t)$ moves in the direction of its long run mean b at speed a. The diffusion function $\sigma^2\{r(t)\} = r(t)\sigma^2$ is proportional to the interest rate $r(t)$ and ensures that the process stays positive. Moreover, here $r(t)$ has a positive impact on the standard deviation through (2.1).

Hull-White Model (1990)

$$dr(t) = \{\delta(t) - ar(t)\}dt + \sigma dW_t$$

This is an extended Vasicek model, where a and σ are constant, $\delta(t)$ is a deterministic function of time. Moreover, this model uses the time dependent reversion level $\delta(t)/a$ for the long run mean instead of the constant b in the Vasicek model.

Black-Karasinski Model (1991)

$$d\log r(t) = \delta(t)\{\log\mu(t) - \log r(t)\}dt + \sigma(t)dW_t$$

with $\delta(t)$, $\mu(t)$ and $\sigma(t)$ as a deterministic function of time, where $\mu(t)$ as the target interest rate. A drawback is that no closed form formula for valuing bonds in terms of $r(t)$ can be derived by this model.

2.3 Methodology

In the Vasicek model, the interest rate $r(t)$ can be negative, whereas the CIR model guarantees the interest rate to be nonnegative. In the Hull-White model, the volatility is a constant. The Black-Karasinski model assumes $\delta(t)$ and $\mu(t)$ are deterministic functions of time. Inherence of all these dynamics is that the coefficient functions can not arbitrarily depend on time. This property might be useful though in a changing macro setting. Thus, we introduce a time varying CIR model in this paper, which allows the coefficients arbitrarily functions of time. We introduce the LPA in detail, which is used to find the longest stable "time homogeneous" interval for each time point, where the parameters in the CIR model can be safely assumed to be constant.

The time varying CIR model is expressed as:

$$dr(t) = a_t\{b_t - r(t)\}dt + \sigma_t\sqrt{r(t)}dW_t \tag{2.2}$$

where, W_t is the standard Wiener Process. Denote the time varying parameters as $\theta_t = (a_t, b_t, \sigma_t)^\top$. This CIR model (2.2) includes all of the aforementioned parametric models, such as jump diffusion models, regime switching models, and also nonparametric specified time varying interest rate models.

The discrete version of (2.2) is:

$$Y_i = r_{t_{i+1}} - r_{t_i} = a_t\{b_t - r_{t_i}\}\Delta t + \sigma_t\sqrt{r_{t_i}}Z_i \tag{2.3}$$

Where $\{Z_i\}_{i=1}^T$ are normally distributed with zero mean and variance $\Delta t = t_{i+1} - t_i$, (more generally, Z_i can be a white noise process). The time unit can be one year, then $\Delta t = \frac{1}{250}$ for daily data, or for weekly data, $\Delta t = \frac{1}{52}$.

2.3.1 Likelihood Function of CIR Process

If a, b, σ are all positive, and $2ab \geq \sigma^2$ holds, then the CIR model is well defined and has a steady state distribution. Given r_t at time t, the density of $r_{t+\Delta t}$ at time point $t + \Delta t$ is:

$$p(r_{t+\Delta t}|r_t; \theta, \Delta t) = ce^{-u-v}(\frac{v}{u})^{\frac{q}{2}}I_q(a\sqrt{uv}) \tag{2.4}$$

where

$$\begin{aligned} c &= \frac{2a}{\sigma^2(1 - e^{-a\Delta t})} \\ u &= cr_t e^{-a\Delta t} \\ v &= cr_{t+\Delta t} \\ q &= \frac{2ab}{\sigma^2} \end{aligned}$$

and $I_q(2\sqrt{uv})$ is the modified Bessel function of the first kind with order q. The log likelihood function is given by:

$$L(\theta) = \sum_{i=1}^{T-1} \log p(r_{t_{i+1}}|r_{t_i}; \theta, \Delta t) \tag{2.5}$$

Fix now t, the MLE estimator $\tilde{\theta}_{I_k}$ in any interval $I_k = [t - m_k, t]$ is:

$$\tilde{\theta}_{I_k} = \arg\max L_{I_k}(\theta) = \arg\max \sum_{i \in I_k} \log p(r_{t_{i+1}}|r_{t_i}; \theta, \Delta t) \tag{$*$}$$

The accuracy of the estimation for a locally constant model with parameter θ_0 is measured via the log likelihood ratio $L_{I_k}(\tilde{\theta}_{I_k}, \theta_0) = L_{I_k}(\tilde{\theta}_{I_k}) - L_{I_k}(\theta_0)$. In C´**i**žek

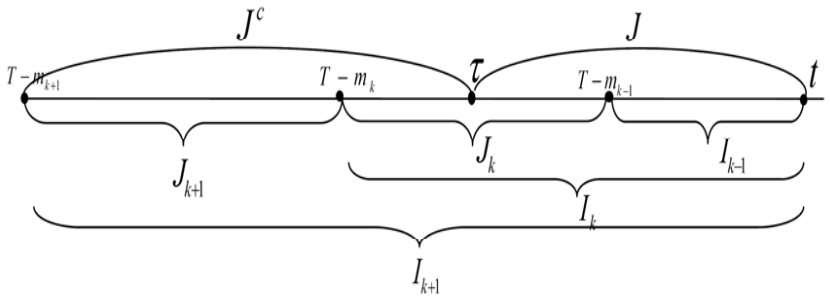

Figure 2.1: Construction of the Test Statistics for LPA: the involved interval I_k and J_k.

et al. (2009), it is proved that if Y_i follows a nonlinear process (2.2), then given I_k for any $r > 0$, there exists a constant $\Re_r(\theta_0)$, such that:

$$\mathsf{E}_{\theta_0} |L_{I_k}(\tilde{\theta}_{I_k}, \theta_0)|^r \leq \Re_r(\theta_0) \tag{2.6}$$

Thus, $\Re_r(\theta_0)$ can be treated as the parametric risk bound. It enables testing the parametric hypothesis on the basis of the fitted log likelihood $L_{I_k}(\tilde{\theta}_{I_k}, \theta_0)$.

2.3.2 Test of Homogeneous Intervals

Mercurio and Spokoiny (2004), C´**i**žek et al. (2009) and Spokoiny (2009) are informative references for the LPA. The general idea can be described as follows: suppose we have K (historical) candidate intervals with a starting interval I_0, i.e. $I_0 \subset I_1 \subset \cdots \subset I_K$, $I_k = [t - m_k, t]$ with $0 < m_k < t$. We increase the length from m_k to m_{k+1}, and test over the larger interval I_{k+1} whether $\tilde{\theta}_{k+1}$ is still consistent with $\tilde{\theta}_k$. To test an interval $I_k = [t - m_k, t]$, we fix the null hypothesis with a constant parameter $\theta_t \equiv \theta$. The alternative (a non constant θ_t) is given by an unknown change point τ in I_k, i.e. $Y_{t'}$ follows one process when $t' \in J = [\tau + 1, t]$ with parameter θ_J, and it follows another process when $t' \in J^c = [t - m_{k+1}, \tau]$ with parameter θ_{J^c}, where $\theta_J \neq \theta_{J^c}$. With this alternative, the log likelihood $(*)$ can be expressed as $L_J(\tilde{\theta}_J) + L_{J^c}(\tilde{\theta}_{J^c})$, giving the test statistics:

$$T_{I_{k+1},\tau} = L_J(\tilde{\theta}_J) + L_{J^c}(\tilde{\theta}_{J^c}) - L_{I_{k+1}}(\tilde{\theta}_{I_{k+1}}) \tag{2.7}$$

where $\tau \in J_k = I_k \backslash I_{k-1}$, see Figure 2.1. Since the change point $\tau \in I_k$ is unknown, we consider the maximum of the test statistics over J_k:

$$T_k = \max_{\tau \in J_k} T_{I_{k+1},\tau} \tag{2.8}$$

This statistics (2.8) is compared with critical values $\{\mathfrak{z}_k\}$, see below for more details.
The selected longest time homogeneous interval $I_{\hat{k}}$ satisfies

$$T_k \leq \mathfrak{z}_k, \qquad \text{for} \quad k \leq \hat{k} \tag{2.9}$$

and $T_{\hat{k}+1} > \mathfrak{z}_{\hat{k}+1}$. In interval $I_{\hat{k}}$ yields the adaptive estimator $\hat{\theta}_t = \hat{\theta}_{I_{\hat{k}}}$. The event $\{I_k \text{ is rejected}\}$ means that $T_\ell > \mathfrak{z}_\ell$ for some $\ell < k$, and hence a change point has been detected in the first k steps.

2.3.3 The Local Parametric Approach (LPA)

For any given t with intervals $I_0 \subset I_1 \subset \cdots \subset I_K$, the algorithm is described in four steps.

1. We estimate $\tilde{\theta}_{I_0}$ using the observations from the smallest interval $I_0 = [t - m_0, t]$, $\tilde{\theta}_{I_0}$ is always accepted.

2. We increase the interval to $I_k, (k \geq 1)$, get the estimator $\tilde{\theta}_{I_k}$ by MLE, and test homogeneity via (2.8), i.e. we test whether there is a change point in I_k. If (2.9) is fulfilled, we go on to step 3, otherwise we go to step 4.

3. Let $\hat{\theta}_{I_k} = \tilde{\theta}_{I_k}$, then further set $k = k + 1$, and go to step 2.

4. Accept as the longest time homogeneous interval $I_{\hat{k}} = I_{k-1}$, and define the local adaptive estimator as $\hat{\theta}_{I_{\hat{k}}} = \tilde{\theta}_{I_{k-1}}$. Additionally set $\hat{\theta}_{I_{\hat{k}}} = \hat{\theta}_{I_k} = \cdots = \hat{\theta}_{I_K}$ for all $k > \hat{k}$.

For a change point τ in I_k, we obtain $\hat{k} = k - 1$, and $I_{\hat{k}} = I_{k-1}$ is the selected longest time homogenous interval. We compare the test statistics with the critical value, if it is smaller than the critical value $\mathfrak{z}_k$ for interval I_k, we accept I_k as the time homogeneous interval, then we increase the interval to I_{k+1}, and do the test again. We sequentially repeat this procedure until we stop at some $k < K$ or we exhaust all the chosen intervals. For each time point t, we use the same algorithm, while we do not need to calculate the critical values a second time, since they depend on only the parametric specification and the length of interval m_k.

To investigate the performance of the adaptive estimator, we introduce the small modeling bias (SMB). The SMB for interval I_k is:

$$\Delta_{I_k}(\theta) = \sum_{t \in I_k} \mathcal{K}\{r(t), r(t;\theta)\} \tag{2.10}$$

with $\mathcal{K}$ the Kullback-Leibler (KL) divergence,

$$\mathcal{K}\{r(t), r(t;\theta)\} = \mathsf{E} \log \frac{p\{r(t)\}}{p\{r(t;\theta)\}} \tag{2.11}$$

where $p(.)$ and $p(.;\theta)$ are the *pdf*s of $r(t)$ and $r(t;\theta)$ respectively. The SMB measures in terms of KL divergence the closeness of a constant parametric model with $p(.;\theta)$ to a time-varying nonparametric model with $p(.)$. Suppose now for a fixed $\Delta > 0$:

$$\mathsf{E}\, \Delta_{I_k}(\theta) \leq \Delta \tag{2.12}$$

Inequality (2.12) simply means that for some $\theta \in \Theta$, $\Delta_{I_k}(\theta)$ is bounded by a small constant , implying that the time varying model can be well approximated (over I_k) by a model with a fixed parameter θ.

Under the SMB condition (2.12) for some interval I_k and $\theta \in \Theta$, one has with a risk bound $\mathfrak{R}_r(\theta)$:

$$\mathsf{E} \log\{1 + \frac{|L_{I_k}(\tilde{\theta}_{I_k}, \theta)|^r}{\mathfrak{R}_r(\theta)}\} \leq 1 + \Delta \tag{2.13}$$

If Δ is not large, (2.13) extends the parametric risk bound $\mathfrak{R}_r(\theta)$ to the nonparametric situation, for details see C´**i**žek et al. (2009). An "oracle" choice I_{k^*} from the set $I_0, \cdots, I_K$ exists, which is defined as the largest interval satisfying (2.12). We denote the corresponding "oracle" parameter as $\theta_{I_{k^*}}$.

However, two types of errors occur in this algorithm: the first type is to reject the time homogeneous interval earlier than the "oracle" step, which means $\hat{k} \leq k^*$. The other type is to select a homogeneous interval larger than the "oracle", i.e. $\hat{k} > k^*$. The first type of error can be treated as a "false alarm", i.e. the algorithm stops earlier than the "oracle" interval I_{k^*}, which leads to selecting an estimate with a larger variation than $\theta_{I_{k^*}}$. The second type of the error arises if $\hat{k} > k^*$. Outside the oracle interval we are exploiting data which does not support the SMB condition. Both errors will be specified in a propagation and stability condition in the next section.

2.3.4 Choice of Critical Values

The accuracy of the estimator can be measured by the log likelihood ratio $L_{I_k}(\tilde{\theta}_{I_k}, \theta_0)$, which is stochastically bounded by the exponential moments (2.13). In general, $\tilde{\theta}_{I_k}$ differs from $\hat{\theta}_{I_k}$ only if a change point is detected at the first k steps. A small value of the likelihood ratio means that $\hat{\theta}_{I_k}$ belongs to the confidence set based on the estimate of $\tilde{\theta}_{I_k}$, i.e. statistically we "accept" $\hat{\theta}_{I_k} = \tilde{\theta}_{I_k}$. If the procedure stops at some $k \leq K$ by a false alarm, i.e. a change point is detected in I_k with the

adaptive estimator $\hat{\theta}_{I_k}$, then the accuracy of the estimator can be expressed via the "propagation" condition:

$$\mathop{\mathsf{E}}_{\theta_0} |L_{I_k}(\tilde{\theta}_{I_k}, \hat{\theta}_{I_k})|^r \leq \rho \Re_r(\theta_0) \tag{2.14}$$

In the parametric situation we can calculate the LHS of (2.14) and choose the critical value $\mathfrak{z}_l$ based on this inequality. The situation at the first k steps can be distinguished into two cases: There is a change point detected at some step $l \leq k$, or there is no change point in the first k intervals. We denote by $\mathfrak{B}_l$ the event of rejection at step l, that is,

$$\mathfrak{B}_l = \{T_1 \leq \mathfrak{z}_1, \cdots, T_{l-1} \leq \mathfrak{z}_{l-1}, T_l > \mathfrak{z}_l\} \tag{2.15}$$

and $\hat{\theta}_{I_k} = \tilde{\theta}_{I_{l-1}}$ on $\mathfrak{B}_l$, $l = 1, 2, \cdots, k$. Now choose $\mathfrak{z}_1$ by minimizing the following equation:

$$\max_{k=1,\cdots,K} \mathsf{E}_{\theta_0} |L(\tilde{\theta}_{I_k}, \tilde{\theta}_{I_0})|^r \mathbf{1}(\mathfrak{B}_1) = \rho \mathfrak{R}_r(\theta_0)/K \tag{2.16}$$

For $\mathfrak{z}_l, l \geq 2$, we use the same algorithm to calculate them. The event $\mathfrak{B}_l$ depends on $\mathfrak{z}_1, \cdots, \mathfrak{z}_l$. Since $\mathfrak{z}_1, \cdots, \mathfrak{z}_{l-1}$ have been fixed by previous steps, the event $\mathfrak{B}_l$ is controlled only by $\mathfrak{z}_l$. Hence, the minimal value of $\mathfrak{z}_l$ should ensure

$$\max_{k \geq l} \mathop{\mathsf{E}}_{\theta_0} |m_k \mathcal{K}(\tilde{\theta}_k, \tilde{\theta}_{l-1})|^r \mathbf{1}(\mathfrak{B}_l) = \rho \mathfrak{R}_r(\theta_0)/K \tag{2.17}$$

or we can express the criterion via the log likelihood ratio:

$$\max_{k \geq l} \mathop{\mathsf{E}}_{\theta_0} |L(\tilde{\theta}_{I_k}, \tilde{\theta}_{I_{l-1}})|^r \mathbf{1}(\mathfrak{B}_l) = \rho \mathfrak{R}_r(\theta_0)/K \tag{2.18}$$

where ρ and r are two global parameters, and m_k denotes the number of observations in I_k. The role of ρ is similar to the level of the test in hypothesis testing problems, while r describes the power of the loss function. We apply $r = 1/2$ in both the simulation and the real data analysis, since it makes the procedure more stable and robust against outliers. We also choose $\rho = 0.2$, however other values in the range $[0.1, 1]$ lead to similar results, Spokoiny (2009).

The critical value $\mathfrak{z}_l$ which satisfies (2.18) can be found numerically by Monte Carlo simulations from the parametric model. It is a decreasing function with respect to the log length of interval. When the interval is small, it is easier to accept it as the time homogeneous interval, since there can not be many jumps due to the short length, while if we increase the length of interval, as more observations are included, it will contain more uncertain information, especially when big jumps or visible structural changes can exist in the interval, therefore in this case, it tends

to reject the test statistics, and the corresponding critical values will decrease as well.

The length of the tested interval is assumed to geometrically increase with $m_k = [m_0 a^k]$. m_0 is the length of initial interval I_0, which is time homogeneous as default. a can be chosen from 1.1 to 1.3. However, the experiments reveal that the estimated results are not sensitive to the choice of a. In the time varying CIR model, to guarantee a reasonable quality of the estimation, we need a large sample size, since there are three parameters to be estimated. Therefore, we choose the length of the initial interval I_0 as $m_0 = 40$ and $a = 1.25$. As already discussed, interest rates are influenced by macroeconomic structures, and may also be subject to regime shifts. Therefore the longest interval we choose should cover one regime, while at least one change point exists between the expansion and recession regimes. Referring to a business cycle of around 4 years, we choose the number of intervals $K = 15$, so that $m_K = 1136$ is the longest tested time homogeneous interval used in both simulation and empirical exercises in this paper.

2.3.5 "Oracle" Property of The Estimators

In this section, we discuss the "oracle" properties of the LPA estimators. Recall that for the "oracle" choice k^*, (2.12) holds, and it also holds for every $k \leq k^*$, while it does not hold for any $k \geq k^*$. However, the "oracle" choice I_{k^*} and $\theta_{I_{k^*}}$ are of course unknown. The LPA algorithm tries to mimic these oracle values. In C´**i**žek et al. (2009), it is proved that under the SMB condition, i.e. when (2.12) holds, the "oracle" property of the LPA estimator $\hat{\theta}_{I_{\hat{k}}}$ satisfies the following property:

For $\theta \in \Theta$ and let $\max_{k \leq k^*} \mathsf{E}\,|L(\tilde{\theta}_{I_{k^*}}, \theta)|^r \mathbf{1}(\mathfrak{B}_1) \leq \mathfrak{R}_r(\theta)$, one has:

$$\mathsf{E} \log\{1 + \frac{|L_{I_{k^*}}(\tilde{\theta}_{I_{k^*}}, \theta)|^r}{\mathfrak{R}_r(\theta)}\} \leq 1 + \Delta \tag{2.19}$$

Further, one obtains:

$$\mathsf{E} \log\{1 + \frac{|L_{I_{k^*}}(\tilde{\theta}_{I_{k^*}}, \hat{\theta}_{I_{\hat{k}}})|^r}{\mathfrak{R}_r(\theta)}\} \leq \rho + \Delta \tag{2.20}$$

This property tells us that although the false alarm occurs before the "oracle" choice, i.e. $\hat{k} \leq k^*$, under the SMB condition, the adaptive estimator $\hat{\theta}_{I_{\hat{k}}}$ does not go far from the oracle value, which implies the LPA estimator does not introduce large errors into the estimation.

The SMB condition doesn't hold if $\hat{k} > k^*$, which means the detected interval is bigger than the "oracle" interval. However, the LPA estimator $\hat{\theta}_{I_{\hat{k}}}$ satisfies Theorem 4.3 in C´**i**žek et al. (2009):

t	a	b	σ
$t \in [1, 500]$	0.2	0.04	0.03
$t \in [501, 1000]$	0.5	0.06	0.1
$t \in [1001, 1500]$	0.8	0.01	0.07

Table 2.1: The parameter settings for simulations of the CIR process

Let $\mathsf{E}\,\Delta_{I_{k^*}(\theta)} \leq \Delta$ for $k^* \leq K$, then $L_{I_{k^*}}(\tilde{\theta}_{I_{k^*}}, \hat{\theta})\mathbf{1}(\hat{k} \geq k^*) \leq \mathfrak{z}_{k^*}$,

$$\mathsf{E}\log\{1 + \frac{|L_{I_{k^*}}(\tilde{\theta}_{I_{k^*}}, \hat{\theta}_{I_{\hat{k}}})|^r}{\mathfrak{R}_r(\theta)}\} \leq \rho + \Delta + \log\{1 + \frac{\mathfrak{z}_{k^*}^r}{\mathfrak{R}_r(\theta)}\} \tag{2.21}$$

This means that $\hat{\theta}_{I_{\hat{k}}}$ belongs with a high probability to the confidence interval of the oracle estimate $\tilde{\theta}_{I_{k^*}}$, i.e. it is still a reliable approximation for the oracle value $\theta_{I_{k^*}}$.

2.4 Simulation Study

We evaluate the performance of the LPA for the CIR model by simulations. We simultaneously change all three parameters $(a_t, b_t, \sigma_t)^\top$ and assume there are two change points for each parameter in the process. We tried several scenarios with change points at different times with satisfactory results. For brevity of presentation, we concentrate here on identical time for the change points of all the three parameters. We simulate the CIR path 100 times with the sample size $T = 1500$. Table 1 summarizes the parameter settings for simulations of the CIR model, the chosen values locate in the range of estimators from the global CIR model.

The estimators $\hat{a}$, $\hat{b}$, and $\hat{\sigma}$ are described in Figures 2.2 to 2.5. In each figure, the blue line respectively depicts the mean of the corresponding estimators from the 100 simulations, and the two dotted red lines are the 5%–95% pointwise confidence intervals for the estimators, the black line describes the respective real parameter. We use the first 250 data points as the training set referring to the moving window estimator, then we estimate the CIR model by the LPA from time point 251 to 1500. One can observe that for the mean reversion speed a, the LPA under the null contains the true parameter in Figure 2.2. Figure 2.3 presents the performance of the LPA estimator $\hat{b}$. Its performance is reasonable. It is obvious to detect there are two jump points, which respectively locate around time point 300 and 800. Taking the delayed time into consideration, the performance of $\hat{b}$ coincides with the true process. It is worth noting that the performance of the LPA estimator $\hat{\sigma}$ is preferable to that of both $\hat{a}$ and $\hat{b}$. The structural break points are evident in

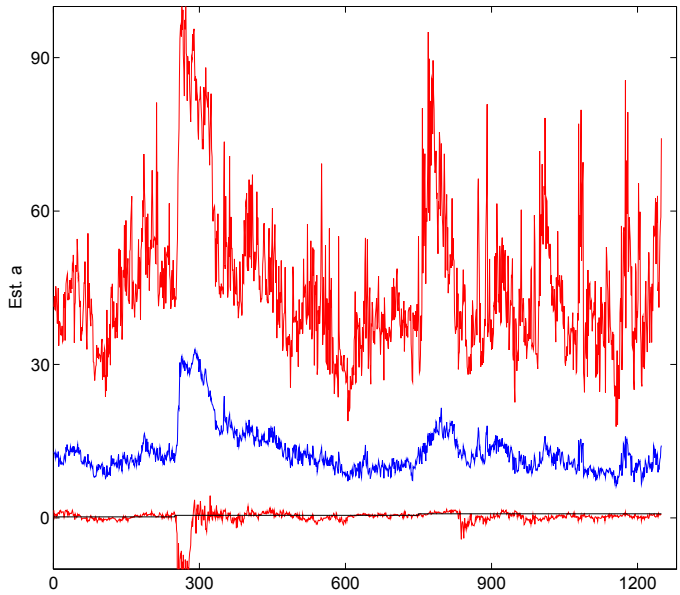

Figure 2.2: LPA estimator $\hat{a}$ with simulated CIR paths. The dotted red lines are the 5%–95% pointwise confidence intervals of $\hat{a}$, the blue line is the mean of $\hat{a}$, and the black line stands for the true process as set in Table 1.

Figure 2.4. Both the mean value and the confidence intervals of the estimator have the same trend as the true parameter path, which indicates the LPA can capture more precise information for volatilities.

Figure 2.5 depicts the selected longest time homogeneous interval for each time point. One can compare the selected homogeneous intervals with the LPA estimators in other figures, all of which provide consistent evidence for its performance. In the initial setting, we have two jumps respectively at 250, and 750. One can easily detect in Figure 2.5 that the two jump points locate respectively around 300 and 800, due to some delayed time. Further, both the 5%–95% pointwise confidence intervals and the mean of the length of the selected intervals coincide with the parameter settings.

2.5 Empirical Study

2.5.1 Data Description

We use the three month treasury bill rate from Federal Reserve Bank of St. Louis as a proxy for the short rate. It has been used frequently in the term structure

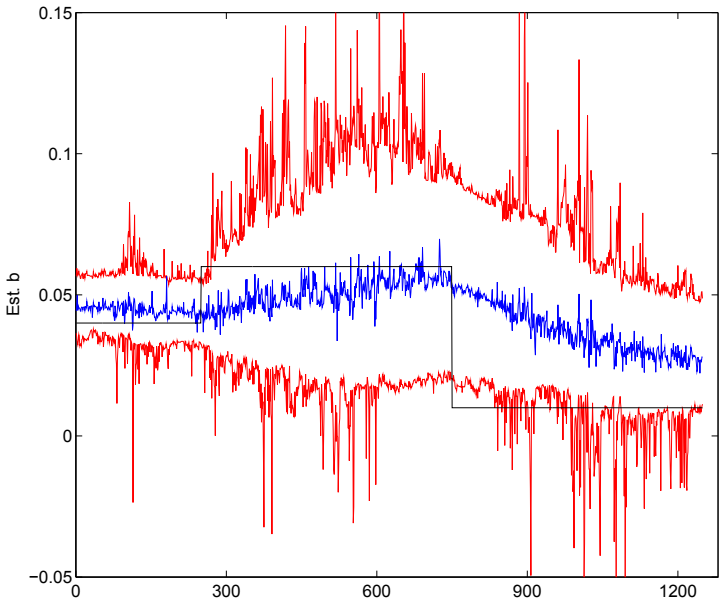

Figure 2.3: LPA estimator $\hat{b}$ with simulated CIR paths.The dotted red lines are the 5%–95% confidence interval of $\hat{b}$, the blue line is the mean of $\hat{b}$, and the black line stands for the true process as set in Table 1.

literature. The data consists of 2840 daily observations, ranging from 2 January, 1998 to 13 May, 2009. The summary statistics are shown in Table 2.2. The short rate and its daily change are displayed in Figure 2.6. Apparently, the volatility of the short rate is time varying. As described in the literature, there are several jumps and break points in the whole period; the short rate from 1999 to 2001 is little volatile, while from mid 2007 to 2009, its volatility is higher than that in other periods. On the basis of the phenomenon we observed from the plot the variation of the short rate is time varying, therefore, we fit the CIR model separately with three different scenarios, the first estimation is using the whole sample, another is with the observations from the beginning of 1998 to the end of July 2007, and the

	Mean	*SD*	*Skewness*	*Kurtosis*
r_t	0.0319	0.0176	-0.1159	-1.4104
dr_t	-1.764×10^{-5}	0.0006	-0.7467	34.4856

Table 2.2: Statistical summary of three month treasury bill rate (daily data) with the period from 2 January,1998 to 13 May, 2009

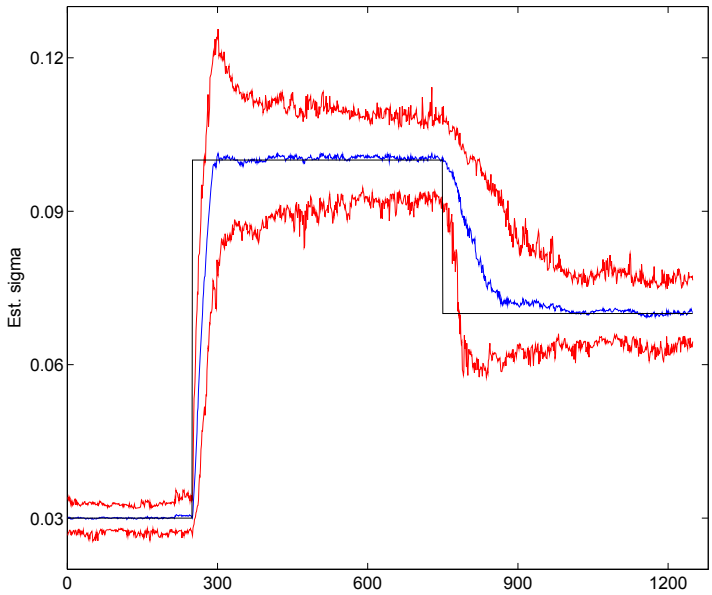

Figure 2.4: LPA estimator $\hat{\sigma}$ with simulated CIR paths. The dotted red lines are the 5%–95% confidence interval of $\hat{\sigma}$, the blue line is the mean of $\hat{\sigma}$, and the black line stands for true process as set in Table 1.

last estimated period is from August 2007 to May 2009. The results are presented in Table 2.3. All three parameters differ significantly during the three different periods. For instance, $\hat{a}$ is around 0.26 from the whole sample, and it changes to 0.14 when the observations range from 1998 to 2007, and in the last period, it jumps a relative high value 3.69. Similar performance can be detected for the long run mean $\hat{b}$. Interestingly, for the volatility, it is relatively low from 1998 to 2007, while it increases to 0.228 in the last period, which also can be verified by Figure 2.6. The volatility of the interest rate at that time is quite high.

Sample Size	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$
19980102–20090513	0.2657	0.0153	0.0944
19980102–20070731	0.1424	0.0252	0.0428
20070731–20090513	3.6792	0.0081	0.2280

Table 2.3: Estimated parameters of CIR model by MLE with three different time periods.

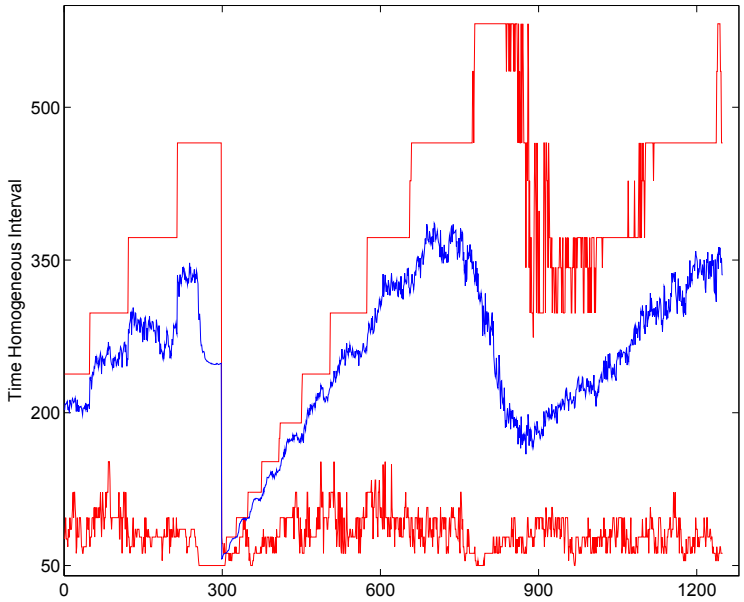

Figure 2.5: The length of time homogenous intervals for simulated CIR paths. The dotted red lines are the 5%–95% confidence interval, the blue lines is the mean of the estimators length of time homogeneous intervals.

2.5.2 Empirical Results

Firstly, we use the moving window estimation to investigate the stability of the coefficients in the CIR model. We specify three window sizes as $l = 250$, $l = 500$, $l = 750$, corresponding to one-year, two-year and three-year periods. Figure 2.7, 2.8 and 2.9 separately presents the moving window estimates $\hat{a}$, $\hat{b}$ and $\hat{\sigma}$. Quite similar performances are illustrated both in $\hat{a}$ and $\hat{b}$. One can find that large variations exist in the process. The moving window estimator $\hat{a}$ with a very large variation is shown in Figure 2.7. It is not surprising that $\hat{a}$ as in the simulation is very sensitive to the data and the length of interval, even for the window size $l = 750$, it still varies a lot. Similarly, big jumps exist in $\hat{b}$. It can be negative at some point, and always fluctuates a lot in different periods. However, the volatility $\hat{\sigma}$ performs in a much more stable way. It keeps almost the same value except in the last periods, where it jumps to a high volatility level.

The critical values are calculated from 500 Monte Carlo runs. We simulate the CIR paths with different combinations of $\hat{a}$, $\hat{b}$, $\hat{\sigma}$ which are chosen from the estimators using different subsamples of the real data. The performance of the critical values is described in Figure 2.10. One can notice, the critical value is a decreasing function with respect to the log length of intervals, which is consistent

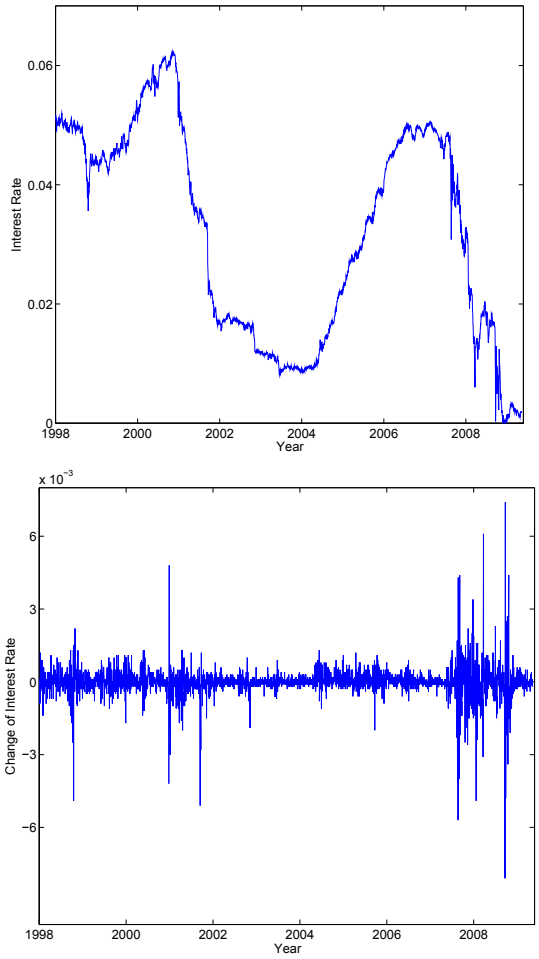

Figure 2.6: Three month treasury bill rate: 19980102—20090513. Top panel: Daily yields; Bottom panel: Changes of daily yields.

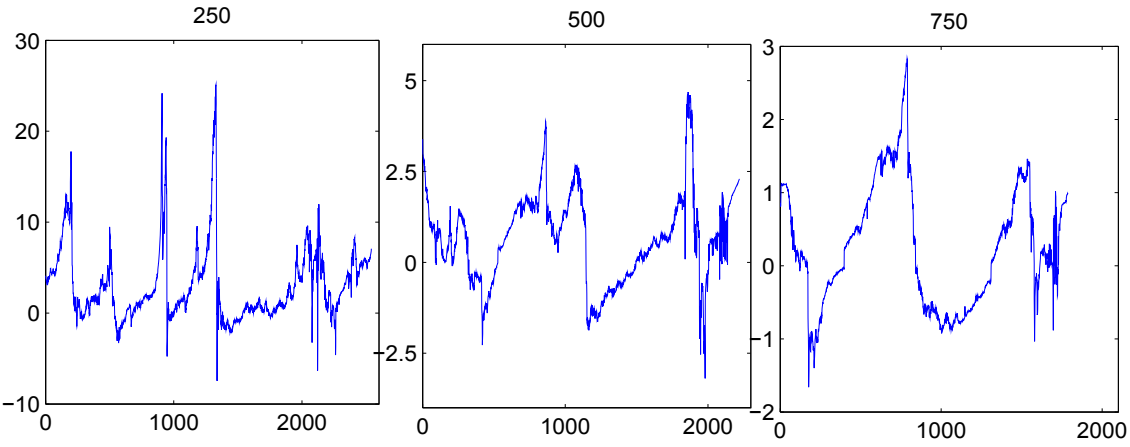

Figure 2.7: Moving window estimator $\hat{a}$ with window sizes 250, 500 and 750 (from left to right).

with the theory mentioned above. Moreover, although the parameter settings are different for the simulation, under the null, there are no very significant differences between the critical values. That is, the critical values are not sensitive to the parameters values we choose. We therefore choose the critical values based on the values estimated globally from data, i.e. $\theta_0^\top = (0.2657, 0.0153, 0.0944)^\top$.

The LPA results are shown from Figures 2.11 to 2.14. The performance of $\hat{a}$ from the LPA is very similarly like that of the moving window estimator $\hat{a}$. It varies a lot during the period, since the interest rate volatility is characterized by a fast mean reverting behaviour reflecting the impact of transient economic shocks, such as central bank announcements of base rate changes. $\hat{b}$ performs volatile in different periods, which is consistent with the behaviour of the length of selected time homogeneous interval described in Figure 2.14. It is relatively stable from 1999 to 2000, while its variation becomes larger in 2001 to 2003. From 2003 to 2007, it turns to be stable again, however in the last period, it reverts to a large variation again. $\hat{\sigma}$ performs relatively stable compared with the other two estimators in the CIR model during the whole time series. Whereas, we can still find three different regimes: from 2001 to 2003, the fluctuation of $\hat{\sigma}$ is increased; from mid 2007, the volatility jumps to a high level, which is also reflected in the length of the intervals $I_{\hat{k}}$ in Figure 2.14.

Figure 2.14 describes the selected time homogeneous interval for each time point t. Here we evaluate from 1999, and treat the first year as a time homogeneous interval. We can compare the performance of the LPA with that of the moving window method.

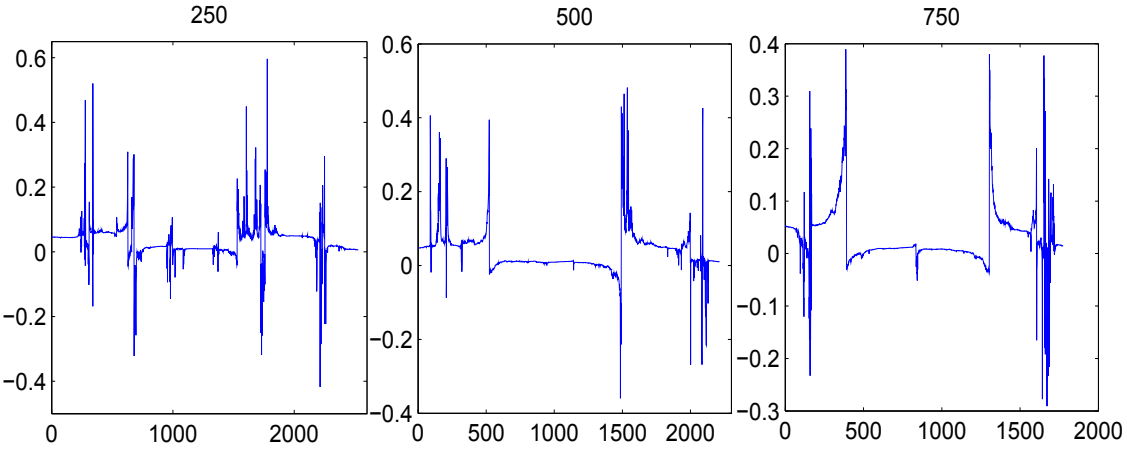

Figure 2.8: Moving window estimator $\hat{b}$ with window sizes 250, 500 and 750 (from left to right).

Please note that the interval $I_{\hat{k}}$ can drop rapidly when the LPA diagnoses a change point. After with a drop, the intervals increase slowly as the LPA gets more confidence into stability of parameters. Moreover, it is worth noting that the length of the selected time homogeneous interval has a close relationship with the regimes of the macroeconomy. On one hand, the recession regime induces shorter homogeneous intervals, and on the other hand, the length is extended in the blooming periods, where the macroeconomy is in a stable state. Let us first analyze the short rate before 2001. In that period, the economic activity continued to expand briskly, and the variation of the short rate was relatively small from the time series plot. We go on to compare the short rate in 2001-2003 with the corresponding selected time homogeneous interval. In this period, the US economy went into the recession period. It was influenced by the terrorist attack on 11 September, 2001, the stock market crash in 2002 and the war in Iraq in 2003, which induced a quite fluctuate macroeconomy: for example, increased oil prices, overstretched investment, too high productivity. All of these factors led to short selected homogeneous intervals. From 2004 to 2006, the economy headed towards a stable state again. The selected intervals lasted longer than before. From 2007, the situation reversed, another global recession came. Again it can be confirmed by the shorter length of the selected intervals, the same as in the period from 2001-2003.

Figure 2.15 depicts the performance of in-sample fitting. The real data is described by the black line, and the two red dashed lines stand for 10%–90% pointwise confidence intervals from the simulated data, where the parameters are

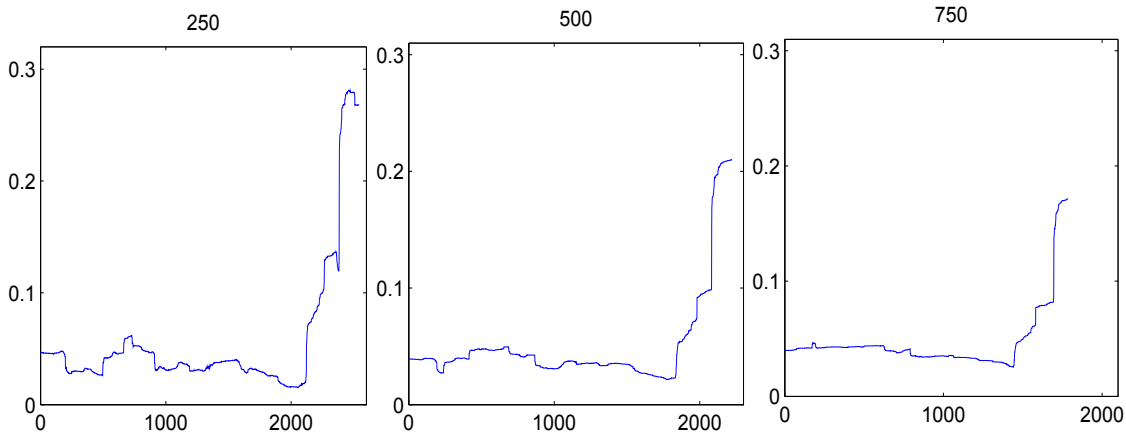

Figure 2.9: Moving window estimator $\hat{\sigma}$ with window sizes 250, 500 and 750 (from left to right).

set the same as that in calculating the critical values. The blue line is the in sample fitting path with the values estimated by the LPA, and the purple line is one randomly selected CIR path from the simulation. One can notice that the fitted sample path by the LPA estimator matches the real data path quite well, i.e. the LPA has an acceptable performance for in sample fitting. The structural break points from the fitted LPA path occur very closely to the real data path.

We further evaluate the out-of-sample forecasting performance of the LPA. We compare the forecasting result of the LPA with that of the moving window method by means of absolute prediction error (APE). It is defined over a prediction period horizon $\mathcal{H}$, $\text{APE}(t) = \sum_{h\in\mathcal{H}} |r_{t+h} - \hat{r}_{t+h|t}|/|\mathcal{H}|$, where $\hat{r}_{t+h|t}$ represents the interest rate prediction by a particular model. Both one-day and ten-day ahead forecasting are considered. Figures 2.16 to 2.18 present the comparison results. In each figure, the left panel stands for the ratio from the forecasting with horizon of one day, and the right panel presents the ten days ahead forecasting. It is clear to see that the LPA performs well especially in the long horizon forecasting.

First, let us consider the result from one-day ahead forecasting. One observes that, in general, the LPA is more preferable than the moving window estimation. Furthermore, as we increase the window size, the variation of the ratio becomes smaller, it is therefore obvious that the LPA performs relatively better, while when the economy is in an unstable state, the LPA for one-day ahead forecasting can not perform very precisely.

Next, let us discuss the prediction results with the horizon of ten days (i.e. 2 weeks). We observe that firstly in comparison with one-step forecasting, the

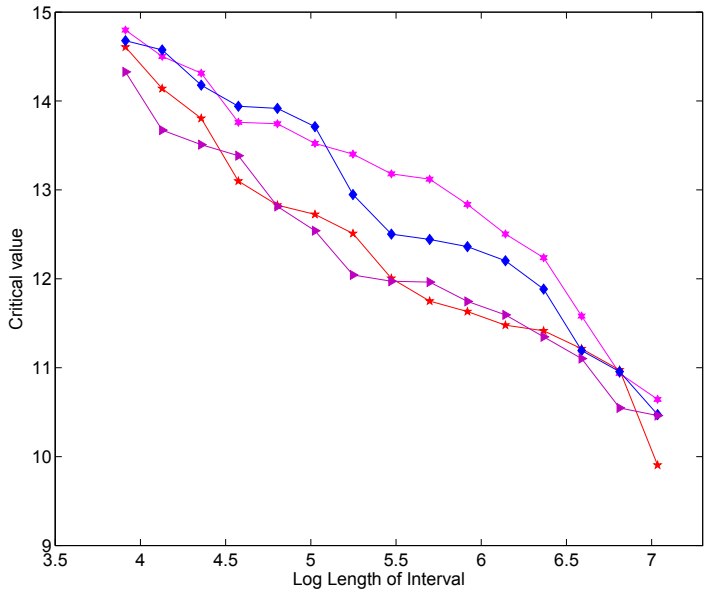

Figure 2.10: Critical values for four combinations of θ, changing one or two from a, b, σ with $m_0 = 40$, $K = 15$ and initial value $r_0 = 0.05$ referred from the real data.

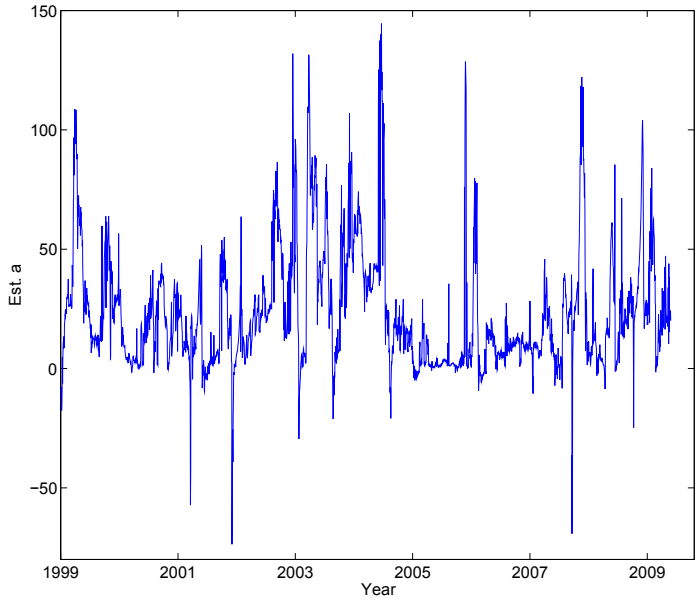

Figure 2.11: Estimated $\hat{a}$ for CIR model using three month treasure bill rate by the LPA.

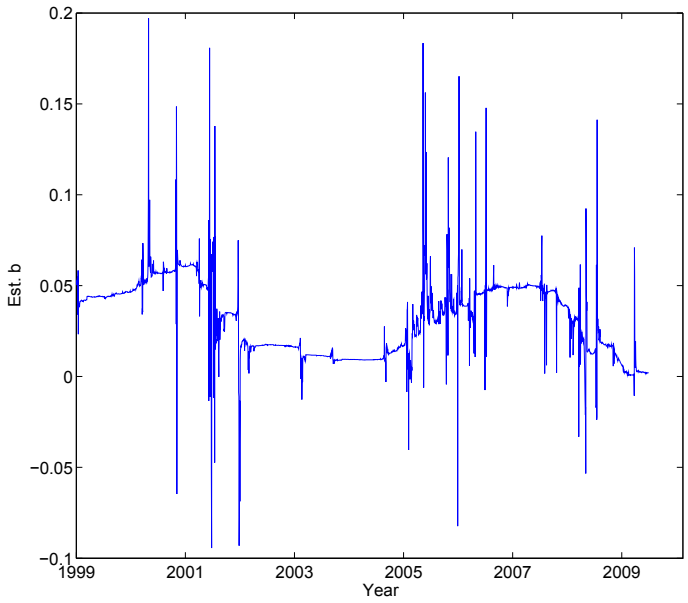

Figure 2.12: Estimated $\hat{b}$ for CIR model using three month treasure bill rate by the LPA.

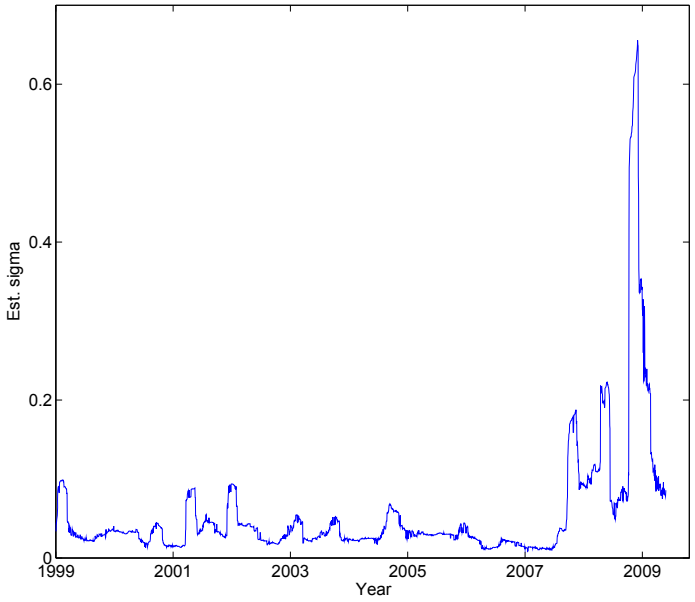

Figure 2.13: Estimated $\hat{\sigma}$ for CIR model using three month treasure bill rate by the LPA.

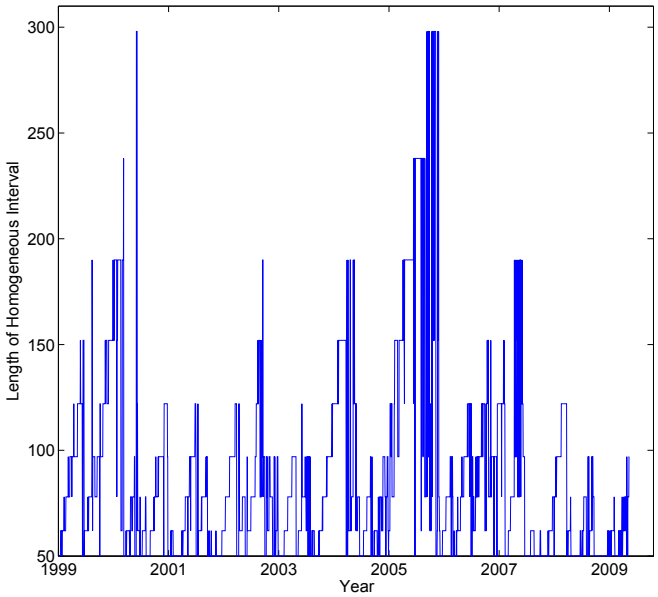

Figure 2.14: The selected longest time-homogeneous intervals using three month treasure bill rate with $\rho = 0.2$, and $r = 0.5$. The first reported time period is in 1999.

variation becomes smaller and the ratios are more stable. Secondly, the LPA shows a superior prediction performance. It is worth noting that generally for ten-day ahead forecasting, the LPA outperforms the moving window estimate in the whole period. Additionally, the LPA forecasting performance improves when we compare with longer moving window estimators, since it is not reasonable to assume all parameters remain the same in a long period. The prediction is clearly better no matter if it is in a stable state or in a volatile state. It is to say that the proposed LPA method shows advantages of forecasting. Additionally, we can confirm that the moving window estimations can not be valid in long horizon forecasting.

Table 2.4 summarizes the prediction performance for the LPA and the moving window (MW) estimations with the forecasting horizon of one day and ten days. We consider the mean of absolute forecasting errors (MAE) for each method. Note that for one-day ahead forecasting, there is no significant difference between the result from the LPA and that from the MW, although the MW estimators are slightly better than the estimators from the LPA, both of their MAEs are quite small. However, in ten-day ahead forecasting, the difference becomes huge. The LPA method performs much better than the MW method. The details are summarized in the table. The accuracy of the MW decreases a lot compared with

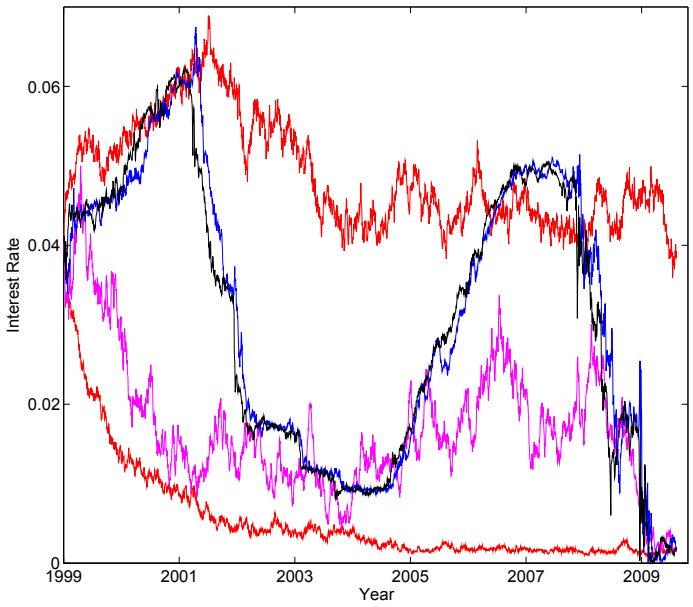

Figure 2.15: In-sample fitting for CIR model using three month treasure bill rate. The black line is the real data; The blue line is the fitted CIR path with the estimators by LPA; The two red lines are 10%–90% confidence intervals simulated with the global estimators; The purple line is a random selected CIR path.

the LPA, especially if we increase the window size, it is more obvious.

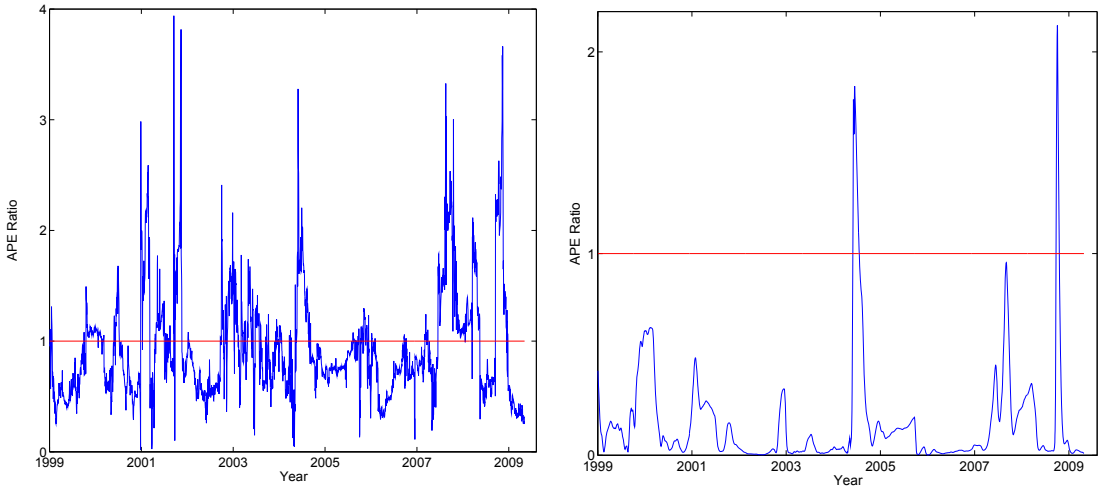

Figure 2.16: The ratio of the absolute prediction errors between the estimators by the LPA (numerator) and moving window estimator (denominator) with window size 250. The left panel: One-day ahead forecasting; The right panel: Ten-day ahead forecasting.

2.6 Conclusion

There are both considerable statistical evidence and economic reasons to believe that the short rate is not following a stable stochastic process. We apply a modern statistical method to describe the changing dynamics of the short rate. With the simple CIR model, and the LPA, we detect structural break points for the short rate process, which is consistent with the conclusion from the existing literature that the dynamics of interest rate is not stable, and also coincides with the reality. We obtain time homogenous intervals for each time point, which is useful to explain the regime switching point. We compare our results with moving window estimators, and the results show that the LPA performs better in both in-sample fitting and out-of-sample forecasting, independent of it being in a stable or unstable period.

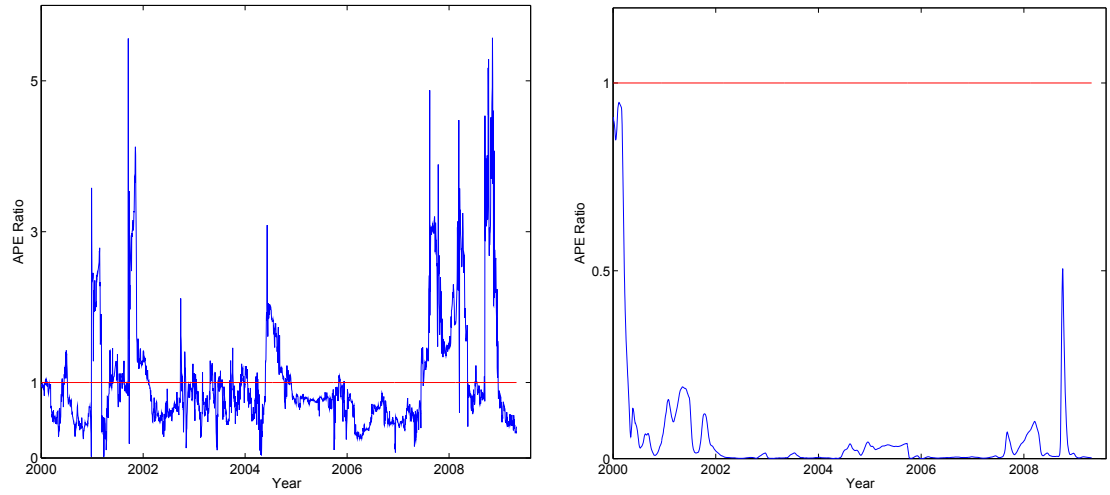

Figure 2.17: The ratio of the absolute prediction errors between the estimators by the LPA (numerator) and moving window estimator (denominator) with window size 500. The left panel: One-day ahead forecasting; The right panel: Ten-day ahead forecasting.

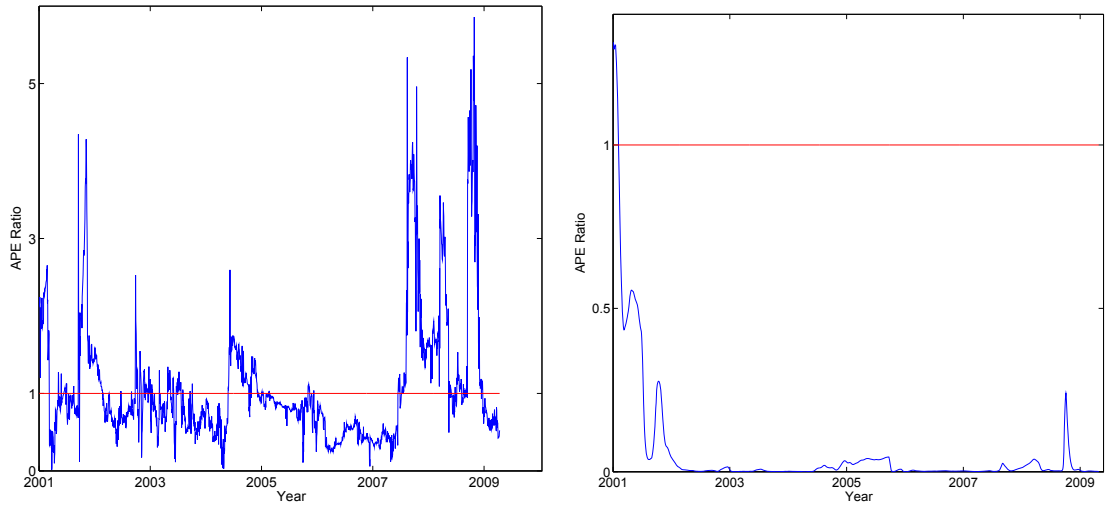

Figure 2.18: The ratio of the absolute prediction errors between the estimators by the LPA (numerator) and moving window estimator (denominator) with window size 750. The left panel: One-day ahead forecasting; The right panel: Ten-day ahead forecasting.

Forecasting Horizon		MAE		
		$l = 250$	$l = 500$	$l = 750$
One Day	LPA	4.7409×10^{-4}	4.8516×10^{-4}	4.9649×10^{-4}
	MW	4.7851×10^{-4}	4.4181×10^{-4}	4.1681×10^{-4}
Ten Days	LPA	**0.0201**	**0.0215**	**0.0232**
	MW	0.1868	1.0032	1.8054

Table 2.4: The table reports the forecast evaluation criteria for one day ahead and ten days ahead forecast of the short rate based on the LPA and moving window (MW) estimation. The first column refers to the forecasting horizon. The second column represents the mean absolute forecast errors according to different moving window sizes.

Chapter 3

Simultaneous Confidence Bands for Expectile Regression

This chapter is based on the paper " Simultaneous Confidence Bands for Expectile Regression" by Guo and Härdle (2011).

3.1 Introduction

In regression function estimation, most investigations are concerned with the conditional mean. Geometrically, the observations $\{(X_i, Y_i), i = 1, \ldots, n\}$ form a cloud of points in a Euclidean space. The mean regression function focuses on the center of the point-cloud, given the covariate X, see Efron (1991). However, more insights about the relation between Y and X can be gained by considering the tails of the conditional distribution.

Asymmetric least squares estimation provides a convenient and relatively efficient method of summarizing the conditional distribution of a dependent variable given the regressors. It turns out that similar to conditional percentiles, the conditional expectiles also characterize the distribution. Breckling and Chambers (1988) proposed M-quantiles, which extend this idea by a "quantile-like" generalization of regression based on asymmetric loss functions. Expectile regression, and more general M-quantile regression, can be used to characterize the relationship between a response variable and explanatory variables when the behaviour of "non-average" individuals is of interest. Jones (1994) described that expectiles and M-quantiles are related to means and quantiles are related to the median, and moreover expectiles are indeed quantiles of a transformed distribution. However, Koenker (2005) pointed out that expectiles have a more global dependence on the form of the distribution.

The expectile curves can be key aspects of inference in various economic prob-

lems and are of great interest in practice. Kuan et al. (2009) considered the conditional autoregressive expectile (CARE) model to calculate the VaR. Expectiles are also applied to calculate the expected shortfall in Taylor (2008). Moreover, Schnabel and Eilers (2009a) analyzed the relationship between gross domestic product per capita (GDP) and average life expectancy using expectile curves. Several well-developed methods already existed to estimate expectile curves. Schnabel and Eilers (2009b) combined asymmetric least square and P-splines to calculate a smooth expectile curve. In this paper, we apply the kernel smoothing techniques for the expectile curve, and construct the simultaneous confidence bands for the expectile curve, which describes a picture about the global variability of the estimator.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. rvs. We denote the joint probability density function (pdf) of the rvs is $f(x, y)$, $F(x, y)$ is the joint cumulative distribution function (cdf), conditional pdf is $f(y|x)$, $f(x|y)$ and conditional cdf $F(y|x)$, $F(x|y)$. Further, $x \in J$ with J a possibly infinite interval in $\mathbb{R}^d$ and $y \in \mathbb{R}$. In general, X may be a multivariate covariate.

From an optimization point of view, both quantile and expectile can be expressed as minimum contrast parameter estimators. Define $\rho_\tau(u) = |\mathbf{I}(u \leq 0) - \tau||u|$ for $0 < \tau < 1$, then the τ-th quantile is expressed as $\arg\min_\theta \mathsf{E}\,\rho_\tau(y - \theta)$, where

$$\mathsf{E}\,\rho_\tau(y-\theta) = (1-\tau)\int_{-\infty}^{\theta} |y-\theta| dF(y|x) + \tau \int_{\theta}^{\infty} |y-\theta| dF(y|x)$$

where θ is the estimator of the τ expectile, and define $\theta \in I$, where the compact set $I \subset \mathbb{R}$. With the interpretation of the contrast function $\rho_\tau(u)$ as the negative log likelihood of asymmetric Laplace distribution, we can see the τ-th quantile as a quasi maximum estimator in the location model. Changing the loss (contrast) function to

$$\rho_\tau(u) = |\,\mathbf{I}(u \leq 0) - \tau|u^2, \qquad \tau \in (0,1) \tag{3.1}$$

leads to expectile. Note that for $\tau = \frac{1}{2}$, we obtain the mean respective to the sample average. Putting this into a regression framework, we define the conditional expectile function (to level τ) as:

$$v(x) = \arg\min_\theta \mathsf{E}\{\rho_\tau(y-\theta)|X = x\} \tag{3.2}$$

Inserting (3.1) into (3.2), we obtain the expected loss function:

$$\mathsf{E}\{\rho_\tau(y-\theta)|X=x\} = (1-\tau)\int_{-\infty}^{\theta} (y-\theta)^2 dF(y|x) + \tau \int_{\theta}^{\infty} (y-\theta)^2 dF(y|x) \tag{3.3}$$

From now on, we silently assume τ is fixed therefore we suppress the explicit notion. Recall that the conditional quantile $l(x)$ at level τ can be considered as

$$l(x) = \inf\{y \in \mathbb{R} | F(y|x) \geq \tau\}$$

Therefore, the proposed estimate $l_n(x)$ can be expressed :

$$l_n(x) = \inf\{y \in \mathbb{R} | \widehat{F}(y|x) \geq \tau\}$$

where $\widehat{F}(y|x)$ is the kernel estimator of $F(y|x)$:

$$\widehat{F}(y|x) = \frac{\sum_{i=1}^n K_h(x - X_i)\mathbf{I}(Y_i \leq y)}{\sum_{i=1}^n K_h(x - X_i)}$$

In the same spirit, define $G_{Y|x}(\theta)$ as

$$G_{Y|x}(\theta) = \frac{\int_{-\infty}^{\theta} |y - \theta| dF(y|x)}{\int_{-\infty}^{\infty} |y - \theta| dF(y|x)}$$

Replacing θ by $v(x)$, we get

$$G_{Y|x}(v) = \frac{\int_{-\infty}^{v(x)} |y - v(x)| \, dF(y|x)}{\int_{-\infty}^{\infty} |y - v(x)| \, dF(y|x)} = \tau$$

so $v(x)$ can be equivalently seen as solving: $G_{Y|x}(\theta) - \tau = 0$ (w.r.t. θ). Therefore,

$$v(x) = G_{Y|x}^{-1}(\tau)$$

with the τth expectile curve kernel smoothing estimator:

$$v_n(x) = \hat{G}_{Y|x}^{-1}(\tau)$$

where the nonparametric estimate of $G_{Y|x}(v)$ is

$$\hat{G}_{Y|x}(\theta) = \frac{\sum_{i=1}^n K_h(x - X_i) \, \mathbf{I}(Y_i < y)|y - \theta|}{\sum_{i=1}^n K_h(x - X_i)|y - \theta|}$$

Quantiles and expectiles both characterize a distribution function although they are different in nature. As an illustration, Figure 3.1 plots curves of quantiles and expectiles of the standard normal $N(0, 1)$. Obviously, there is a one-to-one mapping between quantile and expectile, see Yao and Tong (1996). For fixed x, define $w(\tau)$ such that $v_{w(\tau)}(x) = l(x)$, then $w(\tau)$ is related to the τ-th quantile curve $l(x)$ via

$$w(\tau) = \frac{\tau l(x) - \int_{-\infty}^{l(x)} y dF(y|x)}{2\,\mathsf{E}(Y|x) - 2\int_{-\infty}^{l(x)} y dF(y|x) - (1 - 2\tau)l(x)} \tag{3.4}$$

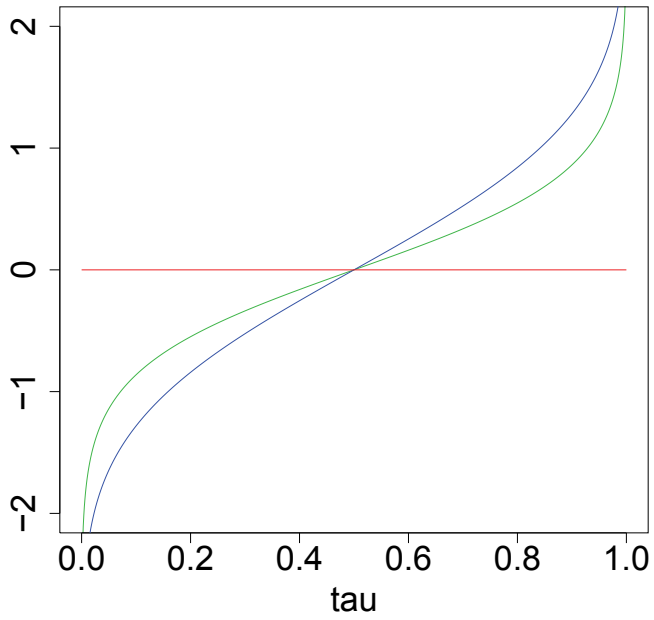

Figure 3.1: Quantile Curve(blue) and Expectile Curve(green) for Standard Normal Distribution (Color online).

$l(x)$ is an increasing function of τ, therefore, $w(\tau)$ is also a monotonically increasing function. Expectiles correspond to quantiles with this transformation w. However, it is not straightforward to apply (3.4), since it depends on the conditional distribution of the regressors. For very simple distributions, it is not hard to calculate the transformation $w(\tau)$, for example, $Y \sim U(-1, 1)$, then $w(\tau) = \tau^2/(2\tau^2 - 2\tau + 1)$. However, if the distribution is more complicated, even worse, the conditional distribution is unknown, it is hard to apply this transformation, see Jones (1994). Therefore, it is not feasible to calculate expectiles from the corresponding quantiles.

In the current paper, we apply the methodology to weather studies. Weather risk is an uncertainty caused by weather volatility. Energy companies take positions in weather risk if it is a source of financial uncertainty. However, weather is also a local phenomenon, since the location, the atmosphere, human activities and some other factors influence the temperature. We investigate whether such local factors exist. Taking two cities, Berlin and Taipei, as an example, we check whether the performance of high expectiles and low expectiles of temperature varies over time. To this end, we calculate the expectiles of trend and seasonality corrected temperature.

The structure of this paper is as follows. In Section 3.2, the stochastic fluc-

tuation of the process $\{v_n(x) - v(x)\}$ is studied and the simultaneous confidence bands are presented through the equivalence of several stochastic processes. We calculate the asymptotic distribution of $v_n(x)$, and the strong uniform consistency rate of $\{v_n(x) - v(x)\}$ is discussed in this section. In Section 4.4, a Monte Carlo study is to investigate the behaviour of $v_n(x)$ when the data is generated with the error terms standard normally distributed. Section 4.5 considers an application in the temperature of Berlin and Taipei. All proofs are attached in Section A.1.

3.2 Results

In light of the concepts of M-estimation as in Huber (1981), if we define $\psi(u)$ as:

$$\begin{aligned} \psi(u) &= \frac{\partial \rho(u)}{\partial u} \\ &= |\mathbf{I}(u \leq 0) - \tau| u \\ &= \{\tau - \mathbf{I}(u \leq 0)\}|u| \end{aligned}$$

$v_n(x)$ and $v(x)$ can be treated as a zero (w.r.t. θ) of the function:

$$H_n(\theta, x) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\psi(Y_i - \theta) \tag{3.5}$$

$$H(\theta, x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(x, y)\psi(y - \theta)dy \tag{3.6}$$

correspondingly.

Härdle (1989) has constructed the uniform confidence bands for general M-smoothers. Härdle and Song (2009) studied the uniform confidence bands for quantile curves. In our paper, we investigate expectile curves, one kind of M-smoother. The loss function for quantile regression is not differentiable, however it is differentiable for expectile when it is in the asymmetric quadratic form. Therefore, by employing similar methods as those developed in Härdle (1989), it is shown in this paper that

$$\begin{aligned} &\mathrm{P}\left[(2\delta \log n)^{1/2}\Big\{\sup_{x \in J} r(x)|v_n(x) - v(x)|/\lambda(K)^{1/2} - d_n\Big\} < z\right] \\ &\longrightarrow \exp\{-2\exp(-z)\}, \quad \text{as } n \to \infty. \end{aligned} \tag{3.7}$$

with some adjustment of $v_n(x)$, we can see that the supreme of $v_n(x) - v(x)$ follows the asymptotic Gumbel distribution, where $r(x)$, δ, $\lambda(K)$, d_n are suitable scaling parameters. The asymptotic result (3.7) therefore allows the construction of simultaneous confidence bands for $v(x)$ based on specifications of the stochastic

fluctuation of $v_n(x)$. The strong approximation with Brownian bridge techniques is applied in this paper to prove the asymptotic distribution of $v_n(x)$.

To construct the confidence bands, we make the following necessary assumptions about the distribution of (X, Y) and the score function $\psi(u)$ in addition to the existence of an initial estimator whose error is a.s. uniformly bounded.
$(A1)$ The kernel $K(\cdot)$ is positive, symmetric, has compact support $[-A, A]$ and is Lipschitz continuously differentiable with bounded derivatives;
$(A2)$ $(nh)^{-1/2}(\log n)^{3/2} \to 0$, $(n \log n)^{1/2} h^{5/2} \to 0$, $(nh^3)^{-1}(\log n)^2 \leqslant M$, M is a constant;
$(A3)$ $h^{-3}(\log n) \int_{|y|>a_n} f_Y(y)dy = \mathcal{O}(1)$, $f_Y(y)$ the marginal density of Y, $\{a_n\}_{n=1}^{\infty}$ a sequence of constants tending to infinity as $n \to \infty$;
$(A4)$ $\inf_{x \in J} |p(x)| \geqslant p_0 > 0$, where $p(x) = \partial\, \mathsf{E}\{\psi(Y-\theta)|x\}/\partial\theta|_{\theta=v(x)} \cdot f_X(x)$, where $f_X(x)$ is the marginal density of X;
$(A5)$ The expectile function $v(x)$ is Lipschitz twice continuously differentiable, for all $x \in J$.
$(A6)$ $0 < m_1 \leqslant f_X(x) \leqslant M_1 < \infty$, $x \in J$, and the conditional density $f(\cdot|y)$, $y \in \mathbb{R}$, is uniform locally Lipschitz continuous of order $\tilde{\alpha}$ (ulL-$\tilde{\alpha}$) on J, uniformly in $y \in \mathbb{R}$, with $0 < \tilde{\alpha} \leqslant 1$, and $\psi(x)$ is piecewise twice continuously differentiable.

Define also

$$\begin{aligned}
\sigma^2(x) &= \mathsf{E}[\psi^2\{Y - v(x)\}|x] \\
H_n(x) &= (nh)^{-1} \sum_{i=1}^{n} K\{(x - X_i)/h\}\psi\{Y_i - v(x)\} \\
D_n(x) &= (nh)^{-1} \frac{\partial \sum_{i=1}^{n} K\{(x - X_i)/h\}\psi\{Y_i - \theta\}}{\partial\theta}|_{\theta=v(x)}
\end{aligned}$$

and assume that $\sigma^2(x)$ and $f_X(x)$ are differentiable.

Assumption $(A1)$ on the compact support of the kernel could possibly be relaxed by introducing a cutoff technique as in Csörgö and Hall (1982) for density estimators. Assumption $(A2)$ has purely technical reasons: to keep the bias at a lower rate than the variance and to ensure the vanishing of some non-linear remainder terms. Assumption $(A3)$ appears in a somewhat modified form also in Johnston (1982). Assumption $(A4)$ guarantees that the first derivative of the loss function, i.e. $\psi(u)$ is differentiable. Assumptions $(A5)$ and $(A6)$ are common assumptions in robust estimation as in Huber (1981), Härdle et al. (1988) that are satisfied by exponential, and generalized hyperbolic distributions.

Zhang (1994) has proved the asymptotic normality of the nonparametric expectile. Under the Assumptions $(A1)$ to $(A4)$, we have:

$$\sqrt{nh}\{v_n(x) - v(x)\} \xrightarrow{\mathcal{L}} \mathrm{N}\big\{0, V(x)\big\} \tag{3.8}$$

with

$$V(x) = \lambda(K) f_X(x) \sigma^2(x) / p(x)^2$$

where we can denote

$$\lambda(K) = \int_{-A}^{A} K^2(u) du$$

$$\begin{aligned} \sigma^2(x) &= \mathsf{E}[\psi^2\{Y - v(x)\}|x] \\ &= \int \psi^2\{y - v(x)\} dF(y|x) \\ &= \tau^2 \int_{v(x)}^{\infty} \{y - v(x)\}^2 dF(y|x) + (1-\tau)^2 \int_{-\infty}^{v(x)} \{y - v(x)\}^2 dF(y|x) \end{aligned} \tag{3.9}$$

$$\begin{aligned} p(x) &= \mathsf{E}[\psi'\{Y - v(x)\}|x] \cdot f_X(x) \\ &= \{\tau \int_{v(x)}^{\infty} dF(y|x) + (1-\tau) \int_{-\infty}^{v(x)} dF(y|x)\} \cdot f_X(x) \end{aligned} \tag{3.10}$$

For the uniform strong consistency rate of $v_n(x) - v(x)$, we apply the result of Härdle et al. (1988) by taking $\beta(y) = \psi(y - \theta)$, $y \in \mathbb{R}$, for $\theta \in I$, $q_1 = q_2 = -1$, $\gamma_1(y) = \max\{0, -\psi(y-\theta)\}$, $\gamma_2(y) = \min\{0, -\psi(y-\theta)\}$ and $\lambda = \infty$ to satisfy the representations for the parameters there. We have the following lemma under some specified assumptions:

LEMMA 3.2.1 *Let $H_n(\theta, x)$ and $H(\theta, x)$ be given by (3.5) and (3.6). Under Assumption (A6) and $(nh/\log n)^{1/2} \to \infty$ through Assumption (A2), for some constant A^* not depending on n, we have a.s. as $n \to \infty$*

$$\sup_{\theta \in I} \sup_{x \in J} |H_n(\theta, x) - H(\theta, x)| \leq A^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{3.11}$$

For our result on $v_n(\cdot)$, we shall also require

$$\inf_{x \in J} \Big| \int \psi\{y - v(x) + \varepsilon\} dF(y|x) \Big| \geqslant \tilde{q}|\varepsilon|, \quad \text{for } |\varepsilon| \leqslant \delta_1, \tag{3.12}$$

where δ_1 and $\tilde{q}$ are some positive constants, see also Härdle and Luckhaus (1984). This assumption is satisfied if there exists a constant $\tilde{q}$ such that $f\{v(x)|x\} > \tilde{q}/p$, $x \in J$.

THEOREM 3.2.1 *Under the conditions of Lemma 3.2.1 and also assuming (3.12) holds, we have a.s. as $n \to \infty$*

$$\sup_{x \in J} |v_n(x) - v(x)| \leq B^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{3.13}$$

with $B^* = A^*/m_1\tilde{q}$ *not depending on* n *and* m_1 *a lower bound of* $f_X(x)$. *If additionally* $\tilde{\alpha} \geqslant \{\log(\sqrt{\log n}) - \log(\sqrt{nh})\}/\log h$, *it can be further simplified to*

$$\sup_{x\in J} |v_n(x) - v(x)| \leq B^*\{(nh/\log n)^{-1/2}\}.$$

THEOREM 3.2.2 *Let* $h = n^{-\delta}$, $\frac{1}{5} < \delta < \frac{1}{3}$ *with* $\lambda(K)$ *as defined before, and*

$$\begin{aligned} d_n &= (2\delta\log n)^{1/2} + (2\delta\log n)^{-1/2}[\log\{c_1(K)/\pi^{1/2}\} + \frac{1}{2}(\log\delta + \log\log n)], \\ &\quad \textit{if } c_1(K) = \{K^2(A) + K^2(-A)\}/\{2\lambda(K)\} > 0 \\ d_n &= (2\delta\log n)^{1/2} + (2\delta\log n)^{-1/2}\log\{c_2(K)/2\pi\} \\ &\quad \textit{otherwise with } c_2(K) = \int_{-A}^{A}\{K'(u)\}^2 du/\{2\lambda(K)\}. \end{aligned}$$

Then (3.7) holds with

$$r(x) = (nh)^{-\frac{1}{2}}p(x)\{\frac{f_X(x)}{\sigma^2(x)}\}^{\frac{1}{2}}$$

This theorem can be used to construct uniform confidence intervals for the regression function as stated in the following corollary.

COROLLARY 3.2.1 *Under the assumptions of the theorem above, an approximate* $(1-\alpha)\times 100\%$ *confidence band over* $[0,1]$ *is*

$$v_n(x) \pm (nh)^{-1/2}\{\hat{\sigma}^2(x)\lambda(K)/\hat{f}_X(x)\}^{1/2}\hat{p}^{-1}(x)\{d_n + c(\alpha)(2\delta\log n)^{-1/2}\}$$

where $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$ *and* $\hat{f}_X(x)$, $\hat{\sigma}^2(x)$ *and* $\hat{p}(x)$ *are consistent estimates for* $f_X(x)$, $\sigma^2(x)$ *and* $p(x)$.

With $\sqrt{V(x)}$ introduced, we can further write Corollary 3.2.1 as:

$$v_n(x) \pm (nh)^{-1/2}\{d_n + c(\alpha)(2\delta\log n)^{-1/2}\}\sqrt{\hat{V}(x)}$$

where $\hat{V}(x)$ is the nonparametric estimator of $V(x)$. Bandwidth selection is quite crucial in kernel smoothing. In this paper, we use the optimal bandwidth discussed in Zhang (1994), which has the following form

$$h_n^{opt} = \left(\frac{\sigma^2(x)\lambda(K)}{n[\Lambda\{v(x)|x\}]^2[\int\{y-v(x)\}^2K^2\{y-v(x)\}dF(y|x)]^2}\right)^{1/5} \tag{3.14}$$

where

$$\Lambda(\theta|x) = \frac{\partial^2\psi(\theta|x-u)}{\partial u^2}|_{u=0}$$

The proof is essentially based on a linearization argument after a Taylor series expansion. The leading linear term will then be approximated in a similar way as in Johnston (1982), Bickel and Rosenblatt (1973). The main idea behind the proof is a strong approximation of the empirical process of $\{(X_i, Y_i)_{i=1}^n\}$ by a sequence of Brownian bridges as proved by Tusnady (1977).

As $v_n(x)$ is the zero (w.r.t. θ) of $H_n(\theta, x)$, it follows by applying 2nd-order Taylor expansions to $H_n(\theta, x)$ around $v(x)$ that

$$v_n(x) - v(x) = \{H_n(x) - \mathsf{E}\, H_n(x)\}/p(x) + R_n(x) \tag{3.15}$$

where $\{H_n(x) - \mathsf{E}\, H_n(x)\}/p(x)$ is the leading linear term and the remainder term is written as:

$$\begin{aligned} R_n(x) &= H_n(x)\{p(x) - D_n(x)\}/\{D_n(x) \cdot p(x)\} + \mathsf{E}\, H_n(x)/p(x) \\ &\quad + \frac{1}{2}\{v_n(x) - v(x)\}^2 \cdot \{D_n(x)\}^{-1} \qquad (3.16) \\ &\quad \cdot (nh)^{-1} \sum_{i=1}^{n} K\{(x - X_i)/h\}\psi''\{Y_i - v(x) + r_n(x)\}, \qquad (3.17) \end{aligned}$$

$$|r_n(x)| < |v_n(x) - v(x)|.$$

We show in Section A.1 that (Lemma A.1.1) that $\|R_n\| = \sup_{x \in J} |R_n(x)| = \mathcal{O}_p\{(nh \log n)^{-1/2}\}$.

Furthermore, the rescaled linear part

$$Y_n(x) = (nh)^{1/2}\{\sigma^2(x) f_X(x)\}^{-1/2}\{H_n(x) - \mathsf{E}\, H_n(x)\}$$

is approximated by a sequence of Gaussian processes, leading finally to the Gaussian process

$$Y_{5,n}(x) = h^{-1/2} \int K\{(x - t)/h\} dW(x). \tag{3.18}$$

Drawing upon the result of Bickel and Rosenblatt (1973), we finally obtain asymptotically the Gumbel distribution.

We also need the Rosenblatt (1952) transformation,

$$T(x, y) = \{F_{X|y}(x|y), F_Y(y)\},$$

which transforms (X_i, Y_i) into $T(X_i, Y_i) = (X_i', Y_i')$ mutually independent uniform rv's. In the event that x is a d-dimension covariate, the transformation becomes:

$$\begin{aligned} T(x_1, x_2, \ldots, x_d, y) &= \{F_{X_1|y}(x_1|y), F_{X_2|y}(x_2|x_1, y), \ldots, \\ &\quad F_{X_k|x_{d-1},\ldots,x_1,y}(x_k|x_{d-1}, \ldots, x_1, y), F_Y(y)\}. \qquad (3.19) \end{aligned}$$

With the aid of this transformation, Theorem 1 of Tusnady (1977) may be applied to obtain the following lemma.

LEMMA 3.2.2 *On a suitable probability space a sequence of Brownian bridges B_n exists that*

$$\sup_{x\in J, y\in\mathbb{R}} |Z_n(x,y) - B_n\{T(x,y)\}| = \mathcal{O}\{n^{-1/2}(\log n)^2\} \quad a.s.,$$

where $Z_n(x,y) = n^{1/2}\{F_n(x,y) - F(x,y)\}$ denotes the empirical process of $\{(X_i, Y_i)\}_{i=1}^n$.

For $d > 2$, it is still an open problem which deserves further research.

Before we define the different approximating processes, let us first rewrite (3.18) as a stochastic integral w.r.t. the empirical process $Z_n(x,y)$,

$$Y_n(x) = \{hg'(x)\}^{-1/2} \iint K\{(x-t)/h\}\psi\{y - v(x)\}dZ_n(t,y),$$

$$g'(x) = \sigma^2(x) f_X(x).$$

The approximating processes are now:

$$Y_{0,n}(x) = \{hg(x)\}^{-1/2} \iint_{\Gamma_n} K\{(x-t)/h\}\psi\{y - v(x)\}dZ_n(t,y) \tag{3.20}$$

where $\Gamma_n = \{|y| \leqslant a_n\}, g(t) = \mathsf{E}[\psi^2\{y - v(x)\} \cdot \mathbf{I}(|y| \leqslant a_n)|X = x] \cdot f_X(x)$

$$Y_{1,n}(x) = \{hg(x)\}^{-1/2} \iint_{\Gamma_n} K\{(x-t)/h\}\psi\{y - v(x)\}dB_n\{T(t,y)\} \tag{3.21}$$

$\{B_n\}$ being the sequence of Brownian bridges from Lemma 3.2.2.

$$Y_{2,n}(x) = \{hg(x)\}^{-1/2} \iint_{\Gamma_n} K\{(x-t)/h\}\psi\{y - v(x)\}dW_n\{T(t,y)\} \tag{3.22}$$

$\{W_n\}$ being the sequence of Wiener processes satisfying

$$B_n(t', y') = W_n(t', y') - t'y'W_n(1,1)$$

$$Y_{3,n}(x) = \{hg(x)\}^{-1/2} \iint_{\Gamma_n} K\{(x-t)/h\}\psi\{y - v(t)\}dW_n\{T(t,y)\} \tag{3.23}$$

$$Y_{4,n}(x) = \{hg(x)\}^{-1/2} \int g(t)^{1/2} K\{(x-t)/h\}dW(t) \tag{3.24}$$

$$Y_{5,n}(x) = h^{-1/2} \int K\{(x-t)/h\}dW(t) \tag{3.25}$$

$\{W(\cdot)\}$ being the Wiener process.

Lemmas A.1.2 to A.1.7 ensure that all these processes have the same limit distributions. The result then follows from

LEMMA 3.2.3 *(Theorem 3.1 in Bickel and Rosenblatt (1973)) Let d_n, $\lambda(K)$, δ as in Theorem 3.2.2. Let*

$$Y_{5,n}(x) = h^{-1/2} \int K\{(x-t)/h\} dW(t).$$

Then, as $n \to \infty$, the supremum of $Y_{5,n}(x)$ has a Gumbel distribution.

$$\mathrm{P}\left\{(2\delta \log n)^{1/2}\Big[\sup_{x\in J} |Y_{5,n}(x)|/\{\lambda(K)\}^{1/2} - d_n\Big] < z\right\} \to \exp\{-2\exp(-z)\}.$$

Same as quantile, the supremum of a nonparametric expectile converge to its limit at a rate $(\log n)^{-1}$. We do not check the bootstrap confidence bands in this paper, which can be the future work. Instead, we point out several well documented literature related to this issue. For example, Claeskens and Keilegom (2003) discussed the bootstrap confidence bands for regression curves and their derivatives. Partial linear quantile regression and bootstrap confidence bands are well studied in Härdle et al. (2010). They proved that the convergence rate by bootstrap approximation to the distribution of the supremum of a quantile estimate has been improved from $(\log n)^{-1}$ to $n^{-2/5}$.

3.3 A Monte Carlo Study

In the design of the simulation, we generate bivariate random variables $\{(X_i, Y_i)\}_{i=1}^n$ with sample size $n = 50, n = 100, n = 200, n = 500$. The covariate X is uniformly distributed on $[0, 2]$

$$Y = 1.5X + 2\sin(\pi X) + \varepsilon \tag{3.26}$$

where $\varepsilon \sim \mathrm{N}(0, 1)$.

Obviously, the theoretical expectiles (fixed τ) are determined by

$$v(x) = 1.5x + 2\sin(\pi x) + v_N(\tau) \tag{3.27}$$

where $v_N(\tau)$ is the τth-expectile of the standard Normal distribution.

Figure 3.2 (in the left part) describes the simulated data (the grey points), together with the 0.5 estimated quantile and estimated expectile and theoretical expectile curves, which represents respectively the conditional median and conditional mean and the theoretical mean. The conditional mean and conditional median coincide with each other, since the error term is symmetrically distributed, which is obvious in Figure 3.2. In the right part of the figure, we consider the conditional 0.9 quantile and expectile curves. Via a transformation (3.4), there is a

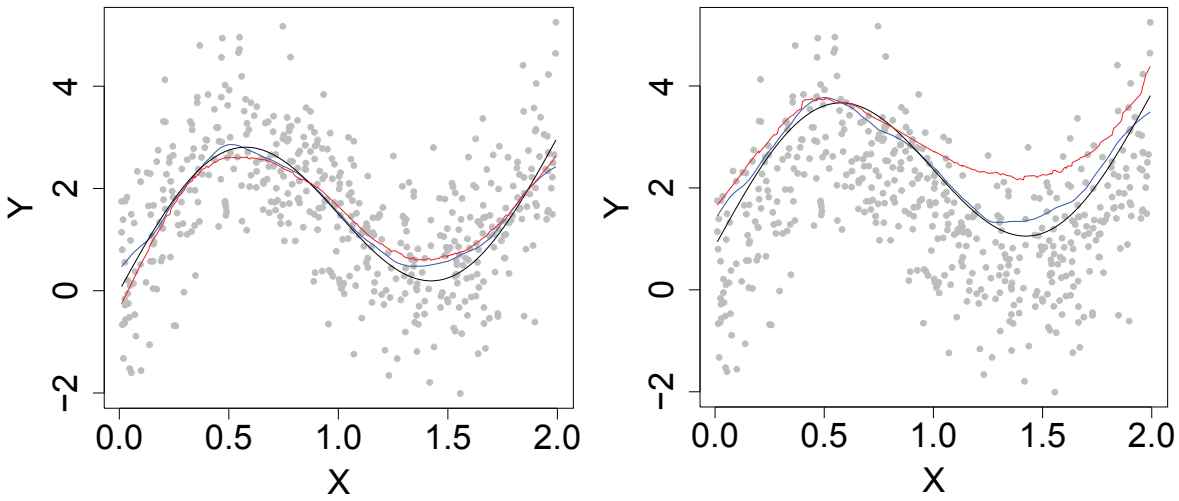

Figure 3.2: $\tau = 0.5$(left) and $\tau = 0.9$(right) Estimated Quantile and Expectile Plot. Quantile Curve, Theoretical Expectile Curve, Estimated Expectile Curve (Color online).

gap between the quantile curve and the expectile curve. By calculating $w(\tau)$ for the standard normal distribution, the 0.9 quantile can be expressed by the around 0.96 expectile. The estimated expectile curve is close to the theoretical one.

Figure 3.3 shows the 95% uniform confidence bands for expectile curve, which are represented by the two red dashed lines. We calculate both 0.1 (left) and 0.9 (right) expectile curves. The black lines stand for the corresponding 0.1 and 0.9 theoretical expectile curves, and the blue lines are the estimated expectile curves. Obviously, the theoretical expectile curves locate in the confidence bands.

To check the performance of the calculated confidence bands, we compare the simulated coverage probability with the nominal values for coverage probability 95% for different sample sizes. We apply this method to both 0.9 and 0.1 expectile. Table 3.1 and Table 3.2 present the corresponding results. We run the simulation 500 times in each scenario. Obviously, the coverage probabilities improve as we increase the sample size, and the width of the bands h becomes smaller for both 0.9 and 0.1 expectile. It is noteworthy that when the number of observation is large enough, for example $n = 500$, the coverage probability is very close to the nominal probability, especially for the 0.9 expectile.

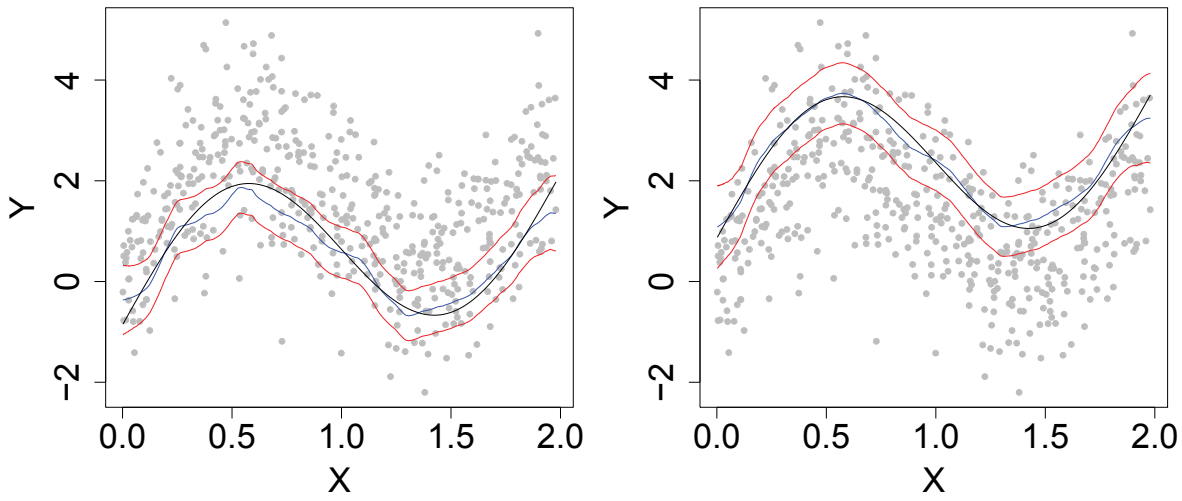

Figure 3.3: Uniform Confidence Bands for Expectile Curve for $\tau = 0.1$ (Left) and $\tau = 0.9$ (Right). Theoretical Expectile Curve, Estimated Expectile Curve and 95% Uniform Confidence Bands (Color online).

n	cp	h
50	0.526	1.279
100	0.684	1.093
200	0.742	0.897
500	0.920	0.747

Table 3.1: simulated coverage probabilities of 95% confidence bands for 0.9 expectile with 500 runs of simulation. cp stands for the coverage probability, and h is the width of the band.

n	cp	h
50	0.386	0.859
100	0.548	0.768
200	0.741	0.691
500	0.866	0.599

Table 3.2: simulated coverage probabilities of 95% confidence bands for 0.1 expectile with 500 runs of simulation. cp stands for the coverage probability, and h is the width of the band.

	Mean	*SD*	*Skewness*	*Kurtosis*	*Max*	*Min*
Berlin	9.66	7.89	-0.315	2.38	30.4	-18.5
Taipei	22.61	5.43	-0.349	2.13	33.0	6.5

Table 3.3: Statistical summary of the temperature in Berlin and Taipei

3.4 Application

In this part, we apply the expectile into the temperature study. We consider the daily temperature both of Berlin and Taipei, ranging from 19480101 to 20071231, together 21900 observations for each city. The statistical properties of the temperature are summarized in Table 3.3. The Berlin temperature data was obtained from Deutscher Wetterdienst, and the Taipei temperature data was obtained from the center for adaptive data analysis in National Central University in Taiwan.

Before proceeding to detailed modeling and forecasting results, it is useful to get an overall view of the daily average temperature data. Figure 3.4 displays the average temperature series of the sample from 2002 to 2007. The black line stands for the temperature in Taipei, and the blue line describes for the temperature in Berlin. The time series plots reveal strong and unsurprising seasonality in average temperature: in each city, the daily average temperature moves repeatedly and regularly through periods of high temperature (summer) and low temperature (winter). It is well documented that seasonal volatility in the regression residuals appears highest during the winter months where the temperature shows high volatility. Importantly, however, the seasonal fluctuations differ noticeably across cities both in terms of amplitude and detail of pattern.

Based on the observed pattern, we apply a stochastic model with seasonality and inter temporal autocorrelation, as in Benth et al. (2007). To understand the model clearly, let us introduce the time series decomposition of the temperature, with $t = 1, \cdots, 365$ days, and $j = 0, \cdots, J$ years:

$$\begin{aligned} X_{365j+t} &= T_{t,j} - \Lambda_t \\ X_{365j+t} &= \sum_{l=1}^{L} \beta_{lj} X_{365j+t-l} + \varepsilon_{t,j} \\ \Lambda_t &= a + bt + \sum_{m=1}^{M} c_l \cos\{\frac{2\pi(t-d_m)}{l \cdot 365}\} \end{aligned} \tag{3.28}$$

where $T_{t,j}$ is the temperature at day t in year j, and Λ_t denotes the seasonality effect. Motivation of this modeling approach can be found in Diebold and Inoue (2001). Further studies as Campbell and Diebold (2005) has provided evidence

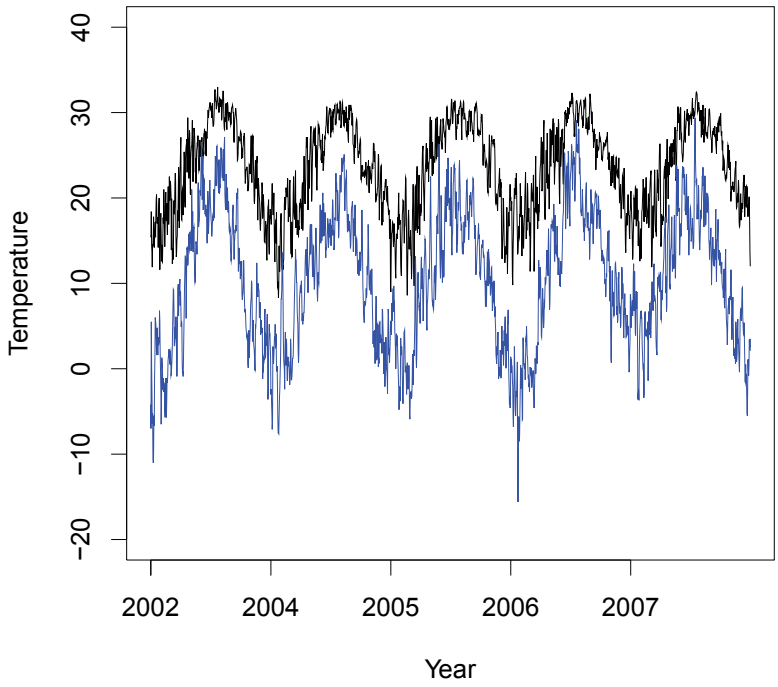

Figure 3.4: The time series plot of the temperature in Berlin and Taipei from 2002-2007. The black line stands for the temperature in Taipei, and the blue line is in Berlin (Color online).

that the parameters β_{lj} are likely to be j independent and hence estimated consistently from a global autoregressive process $AR(L_j)$ model with $L_j = L$. The analysis of the partial autocorrelations and Akaike's Information Criterion (AIC) suggests that a simple AR(3) model fits well the temperature evolution both in Berlin and Taipei.

In this paper, the risk factor of temperature, which is the residual $\hat{\varepsilon}_{t,j}$ from (3.28), is studied in the expectile regression. We intend to construct the confidence bands for the 0.01 and 0.9-expectile curves for the volatility of temperature. It is interesting to check whether the extreme values perform differently in different cities.

The left part of the figures describes the expectile curves for Berlin, and the right part is for Taipei. In each figure, the thick black line depicts the average expectile curve with the data from 1948 to 2007. The red line is the expectile for the residuals from (3.28) with the data of the first 20 years temperature, i.e. in the period from 1948 to 1967. The 0.9 expectile for the second 20 years (1968-1987) residuals is described by the green line, and the blue line stands for the expectile curve in the latest 20 years (1988-2007). The dotted lines are the 95% confidence bands corresponding to the expectile curve with the same color. Figures $3.5 - 3.7$ describe the 0.9 expectile curves for Berlin and Taipei, as well as their corresponding confidence bands. Obviously, the variance is higher in winter-earlier summer both in Berlin and Taipei.

Note that the behaviour of expectile curves in Berlin and Taipei is quite different. Firstly, the variation of the expectiles in Berlin is smaller than that of Taipei. All the expectile curves cross with each other in the last 100 observations of the year for Berlin, and the variance in this period is smaller. Moreover, all of these curves nearly locate in the corresponding three confidence bands. However, the performance of the expectile in Taipei is quite different from that of Berlin. The expectile curves for Taipei have similar trends for each 20 years. They have highest volatilities in January, and lowest volatility in July. More interestingly, the expectile curve for the latest 20 years does not locate in the confidence bands constructed using the data from the first 20 years and second 20 years, see Figure 3.5 and Figure 3.7. Similarly, the expectile curve for the first 20 years does not locate in the confidence bands constructed using the information from the latest 20 years.

Further, let us study low expectile for the residuals of the temperature in Berlin and Taipei. It is hard to calculate very small percentage of quantile curves, due to the sparsity of the data, expectiles though can overcome this drawback. One can calculate very low or very high expectiles, such as 0.01 and 0.99 expectile curves, even when there are not so many observations. Display of the 0.01 expecitles for the residuals and their corresponding confidence bands is given in Figures $3.8 -$

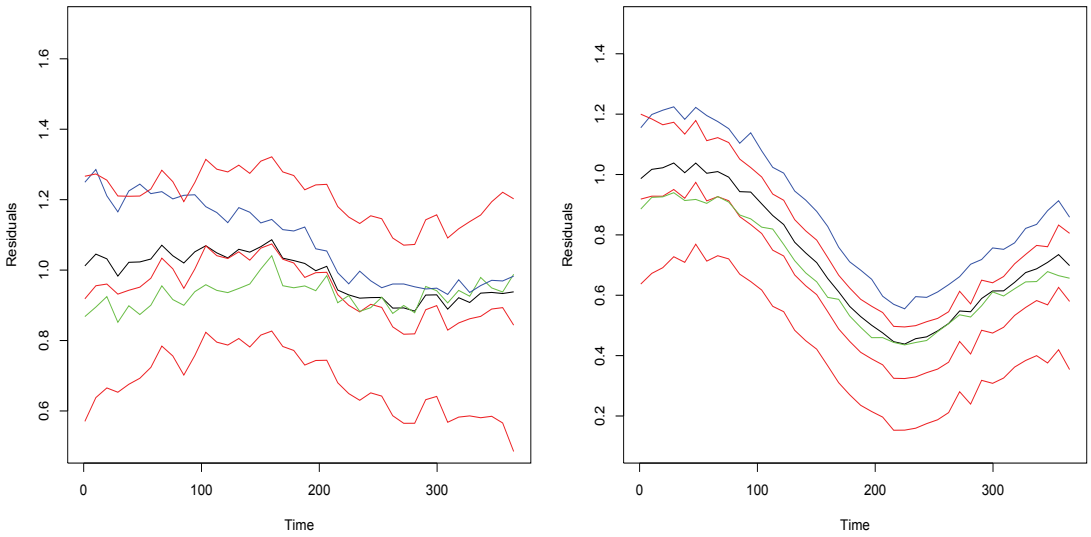

Figure 3.5: 0.9-expectile curves for Berlin (left) and Taipei (right) daily temperature residuals from 1948-2007 with the 95% uniform confidence bands for the first 20 years expectile.

3.10. One can detect that the shapes of the 0.01 expectile for Berlin and Taipei are different. It does not fluctuate a lot during the whole year in Berlin, while the variation in Taipei is much bigger. However, all the curves both for Berlin and Taipei locate in their corresponding confidence bands.

As depicted in the figures, the performance of the residuals are quite different from Berlin and Taipei, especially for high expectiles. The variation of the temperature in Taipei is more volatile. One interpretation is that in the last 60 years, Taiwan has been experiencing a fast developing period. Industrial expansion, burning of fossil fuel and deforestation and other sectors, could be an important factor for the bigger volatility in the temperature of Taipei. However, Germany is well-developed in this period, especially in Berlin, where there are no intensive industries. Therefore, one may say the residuals reveals the influence of the human activities, which induce the different performances of the residuals of temperature.

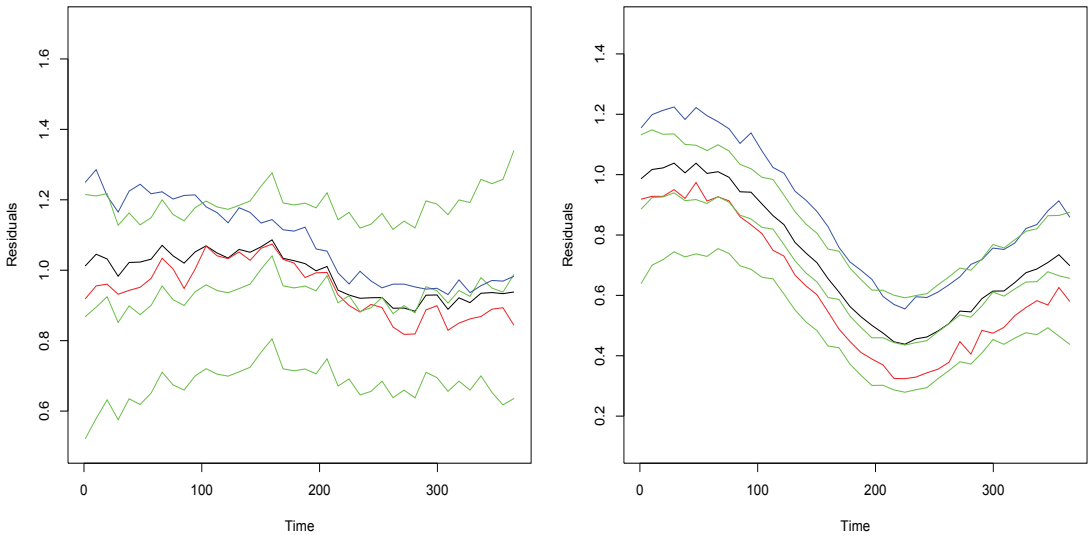

Figure 3.6: 0.9-expectile curves for Berlin (left) and Taipei (right) daily temperature residuals from 1948-2007 with the 95% uniform confidence bands for the second 20 years expectile.

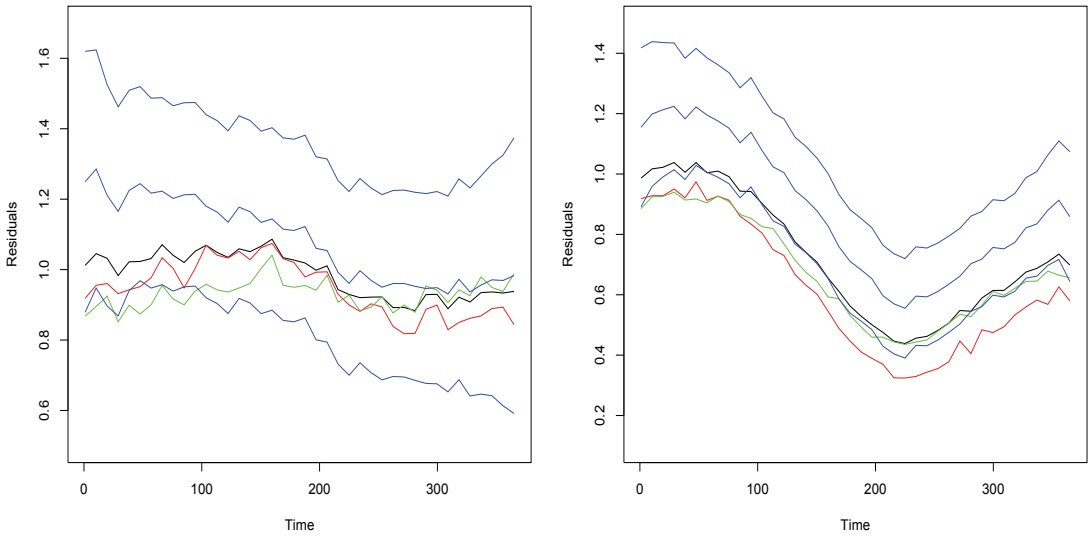

Figure 3.7: 0.9-expectile curves for Berlin (left) and Taipei (right) daily temperature residuals from 1948-2007 with the 95% uniform confidence bands for the latest 20 years expectile.

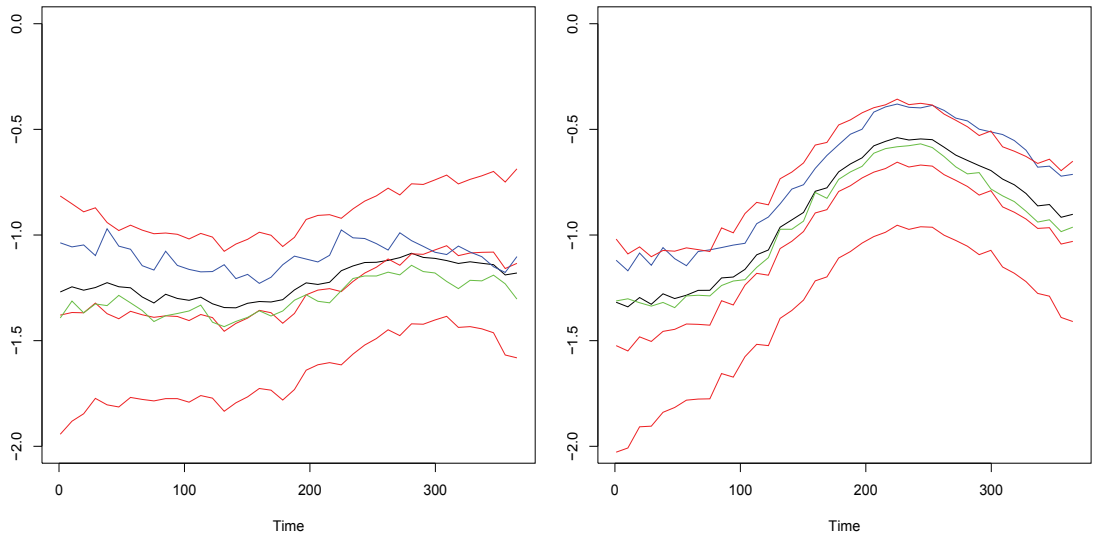

Figure 3.8: 0.01-expectile curves for Berlin (left) and Taipei (right) daily temperature residuals from 1948-2007 with the 95% uniform confidence bands for the first 20 years expectile.

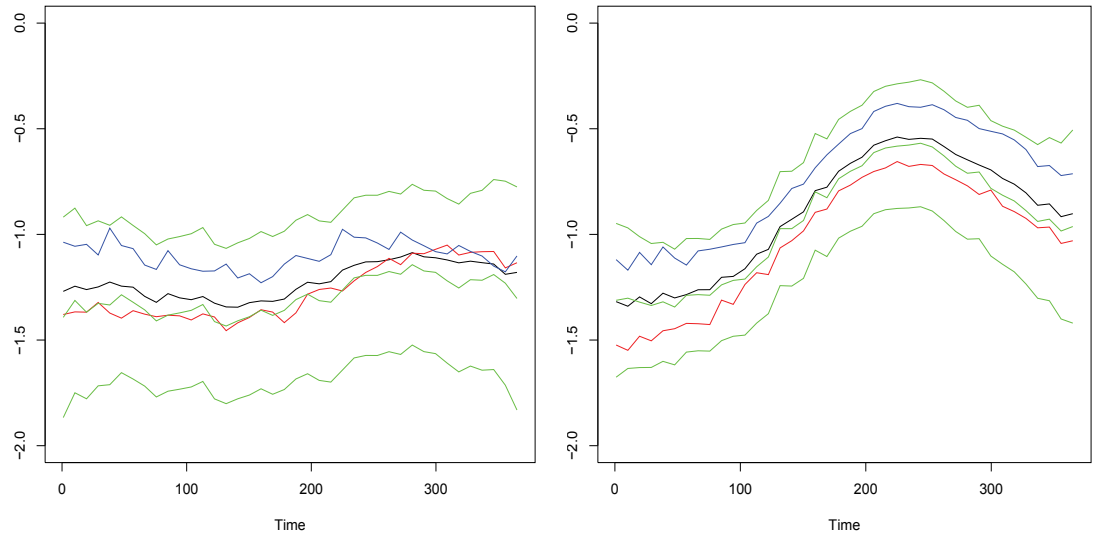

Figure 3.9: 0.01-expectile curves for Berlin (left) and Taipei (right) daily temperature residuals from 1948-2007 with the 95% uniform confidence bands for the second 20 years expectile.

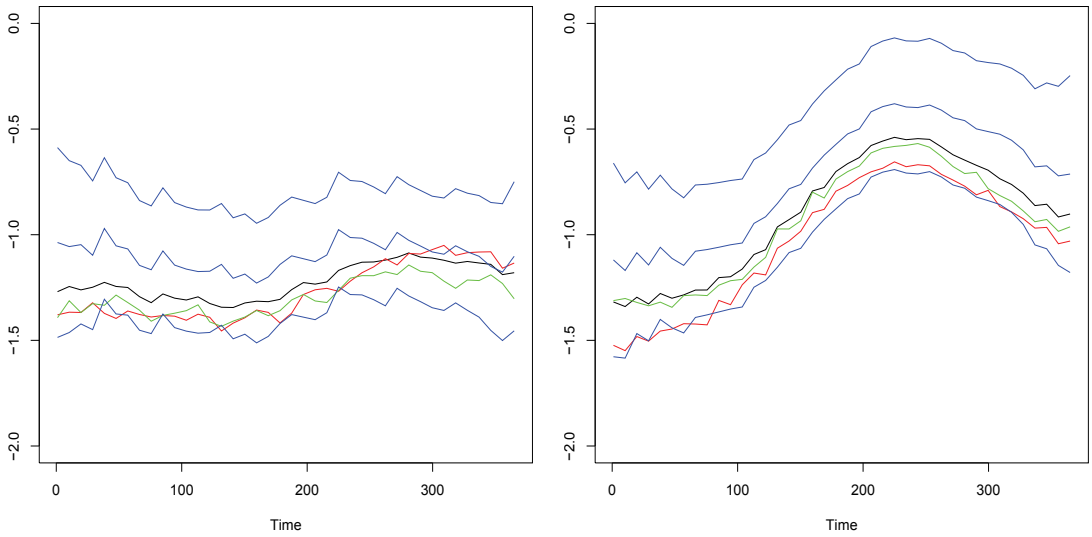

Figure 3.10: 0.01-expectile curves for Berlin (left) and Taipei (right) daily temperature residuals from 1948-2007 with the 95% uniform confidence bands for the latest 20 years expectile.

Chapter 4

Functional Data Analysis for Generalized Quantile Regression

4.1 Introduction

Conventional regression focuses on the conditional mean, typically in the center of the conditional distribution. In a wide spectrum of applications, other features like the variance or the tail behavior contingent on an explanatory variable is of interest. Increased interest focuses on the tail of the distribution, more precisely on conditional quantile or conditional expectile.

As known, quantile regression, introduced by Koenker and Bassett (1978) has been widely applied to capture the tail behaviour of a distribution, similarly, expectile regression, provides another convenient measure of describing the conditional distribution, Newey and Powell (1987). Based on the asymmetric contrast functions, Breckling and Chambers (1988) proposed M-quantiles, which includes quantile and expectile functions as special case. Jones (1994) described that expectiles are related to means the same as quantiles are related to the median, and moreover expectiles are indeed quantiles with a transformation function. Both quantile and expectile can be as the minimum of the asymmetric contrasts function. Quantile functions are to minimize the asymmetric absolute loss function, while expectile functions, defined in L_2 norm, are to minimize the asymmetric square loss function. For the generalized index of the loss function, we introduce a uniform name for both quantile regression and expectile regression as the generalized quantile regression.

The generalized quantile regression is widely used in financial market, demographic studies, and weather analysis, especially for the analysis of the extreme situations. One can apply the generalized quantile function to calculate Value at Risk (VaR) and expected shortfall (ES) in finance in Taylor (2008). It also can

be used to study the relationship between GDP and population in demography, Schnabel and Eilers (2009a), and Härdle and Song (2009) applied it to study the correlation between the wage and the level of education. It is also applied in weather study in Guo and Härdle (2011), where they investigated the volatility of temperature in Berlin and Taipei to detect the temperature risk drivers.

Extreme behaviours are calling more and more attention, especially in meteorology and agricultural economics. For instance, to avoid the loss induced by the extreme behaviour of weather, one is interested to study the fluctuations of temperature, price weather derivatives and further hedge weather risks, see Odening et al. (2008). Weather derivative is a financial instrument to hedge the weather risk. Temperature derivatives may be attractive for industries with temperature-dependent profits, such as tourism or the energy sector. Rainfall derivatives are crucial to the agriculture industry, since the productions, such as wheat, corn, rice, etc., are sensitive to the rainfall. One can rather say that the revenue of above mentioned sectors would have a strong correlation with the extreme fluctuations of weather. Therefore, to investigate the extreme behavior of weather motivates the pricing of weather derivatives and hedging the corresponding weather risk.

In reality, the temperature or rainfall data always involve high dimensional and complex structural data over space and time. For example, in a specific observed period, one always can collect the data from different locations, which can be treated as a longitudinal data. Traditional way to analyze the temperature extremes is to estimate the quantile or expectile functions of the volatility of temperature individually, Guo and Härdle (2011) and Anastasiadou and López-Cabrera (2012). While, the traditional estimation method may ignore the common structure of these generalized quantile curves. Moreover, when estimating very high percentage of quantile curves, due to few observations in the tail of the distribution, the estimate may show high variability.

To solve these issues, we introduce functional data analysis (FDA) to estimate the generalized quantile functions. FDA, a detailed summary of recent work in Ramsay and Silverman (2005), is to study the relationship between functional random variables or with other quantities. However, the traditional studies in functional data analysis focus on the average behavior, and numerous papers have well described solid theoretical foundation and methodology for FDA. However, to gain a informative description of the distribution, the percentages of the tails are motivated to study. Cardot et al. (2005) applied FDA to the quantile regressions. He assumed the explanatory variables have a functional covariate while the response is a scalar. Hence, the conditional quantile function is a function of the quantile index and the covariate. Instead, we fix the percentage of the generalized quantile, and allow the generalized quantile functions vary over individuals. We treat the longitudinal data as observed functional data, and apply FDA to

estimate the generalized quantile curves. We simultaneously obtain generalized quantile curves by applying the information from a collection of individuals, instead of the individual information. However, when more information are included into the regression, it induces more parameters to be estimated, i.e. the challenge of curse of dimensionality, see James et al. (2000).

The objective of this paper is to find several factors to explain the variation of the generalized quantile functions. We assume that the generalized quantile curves share some common features which can be summarized by a small number of principal component functions. In particular, the principle components and the common structure are modeled as penalized splines, and they can be estimated by minimizing the asymmetric contrast loss functions. To get the convergence of the estimators, the iteratively reweighted least squares are used for computation.

James et al. (2000) and Zhou et al. (2008) mentioned that the combination of FDA and penalized spline methods performs well for estimating the average behaviour when data is sparse. By simulation, we also show this method performs well for capturing the tail behaviours. When calculating very high percentage of the generalized quantile curves, by borrowing the common information among individuals makes a comparable result. The temperature data in a specific observed period at different weather stations can be treated as functional data. The generalized quantile curves for the volatility of temperature are estimated by our method. The common structure and individual departure of generalized quantile curves are obtained simultaneously. Analyzing the departures can be used to clarify the individual curves how the curves variate around the common structure curve. Further, the principal component functions contains the information that influences the volatility of temperature.

The structure of this paper is as follows. In Section 4.2, we briefly recall the unconditional and conditional generalized quantiles. Moreover, we explain the details to estimate the generalized quantile curve for single distributions. Functional data analysis (FDA) for generalized quantile curves is introduced in Section 4.3. Both the model construction and estimation algorithm will be discussed in detail. In Section 4.4, a Monte Carlo simulation is studied to investigate the performance of our method in comparison with the traditional single curve estimation. In Section 4.5, an application considers the temperature of 150 weather stations in China in 2010. The aim is to find the risk drivers to the temperature by calculating different percentages of expectile curves of the volatility of the temperature, we conclude in Section 4.6. The completed algorithm is attached in Appendix.

4.2 Generalized Regression Quantiles

Any random variable Y can be characterized by its cdf $F_Y(y) = P(Y \leq y)$, or equivalently, by its quantile function (qf)

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}, \qquad 0 < \tau < 1.$$

The τ-th quantile $Q_Y(\tau)$ minimizes the expected loss,

$$Q(\tau) = \arg\min_y \mathsf{E}\{\rho_\tau(Y - y)\}, \tag{4.1}$$

for the asymmetric loss function $\rho_\tau(Y - y)$ with

$$\rho_\tau(u) = u\{\tau - \mathbf{I}(u < 0)\}. \tag{4.2}$$

When Y is associated with a vector of covariates X, one is interested in studying the conditional (or regression) quantile $Q_{Y|X}(\tau|x) = F^{-1}_{Y|X=x}(\tau)$ as a function of x. Assuming linear dependence on covariates, the τ-th theoretical regression quantile is $Q_{Y|X}(\tau|x) = x^\top \beta^*$, where

$$\beta^* = \arg\min_\beta \mathsf{E}\{\rho_\tau(Y - X^\top \beta)|X = x\}. \tag{4.3}$$

Koenker and Bassett (1978) used this fact to define a minimum contrast estimator of regression quantiles. Since the loss function used in (4.1) and (4.3) can be interpreted as asymmetrically weighted absolute errors, it is natural to consider the asymmetrically weighted squared errors or other asymmetrically weighted loss functions. The expectile curves of Newey and Powell (1987) are the solutions of the optimization problem (4.3) with the loss function corresponding to

$$\rho_\tau(u) = u^2|\tau - \mathbf{I}(u < 0)|.$$

More general asymmetric loss functions have been considered by Breckling and Chambers (1988) to define their M-quantiles which include quantiles and expectiles as special cases.

We now restrict our attention to a univariate covariate but consider the more flexible nonparametric estimation. For fixed τ, the τ-th generalized regression quantile function is defined as

$$l_\tau(x) = \arg\min_\theta \mathsf{E}\{\rho_\tau(Y - \theta)|X = x\}, \tag{4.4}$$

where $\rho_\tau(Y - y)$ is an asymmetric loss function. In this paper we focus on the quantile and expectile curves, corresponding to

$$\rho_\tau(u) = |u|^\alpha|\tau - \mathbf{I}(u < 0)| \tag{4.5}$$

with $\alpha = 1, 2$, respectively, although with slight modifications our methodology is generally applicable for any $\alpha > 0$. According to Jones (1994), the expectiles can be interpreted as quantiles, not of the distribution $F(y|x)$ itself, but of a distribution related to $F(y|x)$. Specifically, write $H(y|x)$ for the conditional partial moment $\int_{-\infty}^{y} u\, F(du|x)$, and denote

$$G(y|x) = \frac{H(y|x) - yF(y|x)}{2\{H(y|x) - yF(y|x)\} + \{y - \mu(x)\}},$$

where $\mu(x) = H(\infty|x) = \int_{-\infty}^{\infty} u\, F(du|x)$ is the conditional mean function. The τ-th expectile function of the distribution $\mathcal{L}(Y|X = x)$ is the quantile of $G(y|x)$, that is, $l_\tau(x) = G^{-1}(\tau|x)$. When they are well-defined, both the conditional quantile function and the expectile function characterize the conditional distribution, and there is a one-to-one mapping between them, as described in Yao and Tong (1996). Quantiles are intuitive, but expectiles are easier to compute and more efficient to estimate, Schnabel and Eilers (2009b).

To estimate the generalized quantile functions, assume we have paired data (X_i, Y_i), $i = 1, \ldots, n$, an i.i.d. sample from the joint distribution of (X, Y). It follows from (4.4) that the generalized quantile function $l_\tau(\cdot)$ minimizes the unconditional expected loss,

$$l_\tau(\cdot) = \underset{f \in \mathcal{F}}{\arg\min}\, \mathsf{E}[\rho_\tau\{Y - f(X)\}], \tag{4.6}$$

where $\mathcal{F}$ is the collection of functions such that the expectation is well-defined. Using the method of penalized splines (Eilers and Marx, 1996; Ruppert et al., 2003), we represent $f(x) = b(x)^\top \gamma$, where $b(x) = \{b_1(x), \ldots, b_q(x)\}^\top$ is a vector of B-spline basis functions and γ is a q-vector of coefficients, and minimize the penalized avarage empirical loss,

$$\widehat{l}_\tau(\cdot) = \underset{f(\cdot) = b(\cdot)^\top \gamma}{\arg\min} \sum_{i=1}^{n} \rho_\tau\{Y_i - f(X_i)\} + \lambda \gamma^\top \Omega \gamma, \tag{4.7}$$

where Ω is a penalty matrix and λ is the penalty parameter. The penalty term is introduced to penalize the roughness of the fitted generalized quantile function $\widehat{l}_\tau(\cdot)$. When X_i's are evenly spaced, the penalty matrix Ω can be chosen such that $\gamma^\top \Omega \gamma = \sum_i (\gamma_{i+1} - 2\gamma_i + \gamma_{i-1})^2$ is the squared second difference penalty. In this case, $\Omega = D^\top D$ and D is the second-differential matrix such that $D\gamma$ creates the vector of second differences $\gamma_{i+1} - 2\gamma_i + \gamma_{i-1}$. In general, the penalty matrix Ω can be chosen to be $\int \ddot{b}(x)\ddot{b}(x)^\top dx$ such that $\gamma^\top \Omega \gamma = \int \{\ddot{b}(x)^\top \gamma\}^2 dx$, where $\ddot{b}(x) = \{\ddot{b}_1(x), \ldots, \ddot{b}_q(x)\}^\top$ denotes the vector of second derivatives of the

basis functions. The minimizing objective function in (4.7) can be viewed as the penalized negative log likelihood for the signal-plus-noise model

$$Y_i = l_\tau(X_i) + \varepsilon_i = b(X_i)^\top \gamma + \varepsilon_i, \tag{4.8}$$

where ε_i follows a distribution with a density proportional to $\exp\{-\rho_r(u)\}$, which corresponds respectively to the asymmetric Laplace distribution or the asymmetric Gaussian distribution for $\alpha = 1$ and $\alpha = 2$ (Koenker and Machado, 1999). Since these distributions are rather implausible for real-world data, their likelihood is better interpreted as a quasi-likelihood.

For expectiles ($\alpha = 2$ in the definition of loss function), Schnabel and Eilers (2009b) developed an iterative least asymmetrically weighted squares (LAWS) algorithm to solve the minimization problem (4.7), by extending an idea of Newey and Powell (1987). They rewrote the objective function in (4.7) as

$$\sum_{i=1}^{n} w_i(\tau)\{Y_i - b(X_i)^\top \gamma\}^2 + \lambda \gamma^\top \Omega \gamma, \tag{4.9}$$

where

$$w_i(\tau) = \begin{cases} \tau & \text{if } Y_i > b(X_i)^\top \gamma, \\ 1-\tau & \text{if } Y_i \le b(X_i)^\top \gamma. \end{cases} \tag{4.10}$$

For fixed weights $w_i(\tau)$'s, the minimizing $\widehat{\gamma}$ has a closed-form expression

$$\widehat{\gamma} = (B^\top W B + \lambda \Omega)^{-1} B^\top W Y, \tag{4.11}$$

where B is a matrix whose i-th row is $b(X_i)^\top$, W is the diagonal matrix whose ith diagonal entry is $w_i(\tau)$, and $Y = (Y_1, \ldots, Y_n)^\top$. Note that the weights $w_i(\tau)$'s depend on the spline coefficient vector γ. The LAWS algorithm iterates until convergence between computing (4.11) and updating W using (4.10) with γ being its current value obtained from (4.11).

With a slight modification, the LAWS algorithm can also be used to calculate the penalized spline estimator of conditional quantile functions, which correspond to $\alpha = 1$ in the asymmetric loss function. The weights for calculating the expectiles given in (4.10) need to be replaced by

$$w_i(\tau) = \begin{cases} \dfrac{\tau}{|Y_i - b(X_i)^\top \gamma| + \delta} & \text{if } Y_i > b(X_i)^\top \gamma, \\ \dfrac{1-\tau}{|Y_i - b(X_i)^\top \gamma| + \delta} & \text{if } Y_i \le b(X_i)^\top \gamma, \end{cases} \tag{4.12}$$

where $\delta > 0$ is a small constant used to avoid numerical problems when $Y_i = b(X_i)^\top \gamma$ is close to zero. In this case, the LAWS algorithm can be interpreted as

a variant of Majorization-Minimization (MM) algorithm and the convergence of the LAWS algorithm then follows from the general convergence theory of the MM algorithm; see Hunter and Lange (2000).

One advantage of expectiles is that they can always be calculated no matter how low or high of the generalized quantile level τ, while the empirical quantiles can be undefined at extreme tails of the data distribution. It is also known that estimation of expectiles is usually more efficient than that of quantiles since it makes more effective use of data Schnabel and Eilers (2009b). However, when τ is close to 0 or 1, we still can estimate expectiles, but quantiles exhibit high variability, because of sparsity of data in the tails of the distribution. In the next section, we will present a method for better quantile and expectile estimation when there is a need to estimate a collection of generalized quantile functions and, if these functions share some common features. We use functional data analysis techniques to improve the estimation efficiency by borrowing strength across data sets.

4.3 Functional data analysis for a collection of regression quantiles

4.3.1 Approach

When we are interested in a collection of generalized quantile curves, denoted as $l_i(t)$, $i = 1, \ldots, N$, we may treat them as functional data. Suppose $l_i(t)$'s are independent realizations of a stochastic process $l(t)$ defined on a compact interval $\mathcal{T}$ with the mean function $\mathsf{E}\{l(t)\} = \mu(t)$ and the covariance kernel $K(s,t) = \mathrm{Cov}\{l(s), l(t)\}, s, t \in \mathcal{T}$. (To emphasize the one-dimensional natural of the covariate, from now on we change notation for the covariate from x to t.) If $\int_{\mathrm{I}} K(t,t)dt < \infty$, then Mercer's Lemma states that there exists an orthonormal sequence of eigen-functions (ψ_j) and a non-increasing and non-negative sequence of eigenvalues (κ_j) such that

$$(K\psi_j)(s) \stackrel{\text{def}}{=} K(s,t)\psi_j(t)dt = \kappa_j\psi_j(s),$$

$$K(s,t) = \sum_{j=1}^{\infty} \kappa_j\psi_j(s)\psi_j(t),$$

and

$$\sum_{j=1}^{\infty} \kappa_j = \int_{\mathrm{I}} K(t,t)dt.$$

Moreover, we have the following Karhunen-Loève expansion

$$l(t) = \mu(t) + \sum_{j=1}^{\infty} \sqrt{\kappa_j}\xi_j\psi_j(t), \tag{4.13}$$

where $\xi_j \stackrel{\text{def}}{=} \frac{1}{\sqrt{\kappa_j}} \int\ l(t)\psi_j(s)ds$, $\mathsf{E}(\xi_j) = 0$, $\mathsf{E}(\xi_j\xi_k) = \delta_{j,k}$, $j, k \in \mathbb{N}$, and $\delta_{j,k}$ is the Kronecker delta.

Usually statistical estimation demands a parsimonious model for estimation efficiency and thus the terms associated with small eigenvalues in (4.13) can be neglected. As a result, we obtain the following factor model

$$l_i(t) = \mu(t) + \sum_{k=1}^{K} f_k(t)^\top \alpha_{ik} = \mu(t) + f(t)^\top \alpha_i, \tag{4.14}$$

where f_k is the k-th factor with $f(t) = \{f_1(t), \cdots, f_K(t)\}^\top$, $\alpha_i = (\alpha_{i1}, \cdots, \alpha_{iK})^\top$ is the vector of scores, and K is the number of factors to be used in the model. The function μ can be interpreted as the mean function, and the factors f_k's can be interpreted as the functional principal components James et al. (2000); Zhou et al. (2008). Since the factor model (4.14) indicates that the collection of generalized quantile curves share the same mean function and the same set of principal components, it opens the door for borrowing information across data sets to improve the estimation efficiency.

Accepting the parametrizations in (4.14), estimation of the generalized quantile functions l_i's is reduced to the estimation of the mean and principal components functions. Using the method of penalized splines again, we represent these functions in the form of basis expansions

$$\begin{aligned} \mu(t) &= b(t)^\top \theta_\mu, \\ f(t)^\top &= b(t)^\top \Theta_f, \end{aligned} \tag{4.15}$$

where $b(t) = \{b_1(t), \cdots, b_q(t)\}^\top$ is a q-vector of B-splines, θ_μ is a q-vector and $\Theta_f = \{\theta_{f,1}, \cdots, \theta_{f,K}\}^\top$ is a $q \times K$ matrix of spline coefficients. The B-splines are normalized so that

$$\int b(t)b(t)^\top dt = \mathbf{I}_q.$$

Thus the estimation problem is further reduced to the estimation of spline coefficients. For identifiability, we impose the following restriction

$$\Theta_f^\top \Theta_f = \mathbf{I}_K.$$

The above two equations imply the usual orthogonality requirements of the principal component curves:

$$\int f(t)f(t)^\top dt = \Theta_f^\top \int b(t)b(t)^\top dt \Theta_f = \mathbf{I}_K.$$

Denote the observations as $\{Y_{ij}\}$ with $i = 1, \cdots, N$, $j = 1, \cdots, T_i$. Combining (4.14) and (4.15) yields the following data model

$$l_{ij} \stackrel{\text{def}}{=} l_i(t_{ij}) = b(t_{ij})^\top \theta_\mu + b(t_{ij})^\top \Theta_f \alpha_i. \tag{4.16}$$

Here, the scores α_i's are treated as fixed effects instead of random effects for convenience in applying the asymmetric loss minimization and, for identifiability, their average is assumed to be 0. The empirical loss function for generalized quantile estimation is

$$S = \sum_{i=1}^{N} \sum_{j=1}^{T_i} \rho_\tau\{Y_{ij} - b(t_{ij})^\top \theta_\mu - b(t_{ij})^\top \Theta_f \alpha_i\}, \tag{4.17}$$

where $\rho_\tau(u)$ is the asymmetric loss function defined in (4.5). To ensure the smoothness of the estimates of the mean curve and the principal components curves, we use a moderate number of knots and apply a roughness penalty to regularize the fitted curves. The squared second derivative penalties for the mean and principal components curves are given by

$$M_\mu = \theta_\mu^\top \int \ddot{b}(t)\ddot{b}(t)^\top \, dt \, \theta_\mu = \theta_\mu^\top \Omega \, \theta_\mu,$$
$$M_f = \sum_{k=1}^{K} \theta_{f,k}^\top \int \ddot{b}(t)\ddot{b}(t)^\top \, dt \, \theta_{f,k} = \sum_{k=1}^{K} \theta_{f,k}^\top \Omega \, \theta_{f,k}.$$

The penalized empirical loss function is then

$$S^* = S + \lambda_\mu M_\mu + \lambda_f M_f, \tag{4.18}$$

where λ_μ and λ_f are nonnegative penalty parameters. Note that we use the same penalty parameter for all principal components curves for the sake of simplicity. We propose to minimize the penalized loss (4.18) to estimate the parameters θ_μ, Θ_f, and α_i's. The choice of the penalty parameters will be discussed later in the paper.

Define the vector $L_i = \{l_{i1}, \cdots, l_{iT_i}\}^\top$ and the matrix $B_i = \{b(t_{i1}), \cdots, b(t_{iT_i})\}^\top$. The data model can be written in matrix form as

$$L_i = B_i\theta_\mu + B_i\Theta_f\alpha_i \tag{4.19}$$

Writing $Y_i = (Y_{i1}, \ldots, Y_{iT_i})^\top$, the data have the following signal-plus-noise representation

$$Y_i = L_i + \varepsilon_i = B_i\theta_\mu + B_i\Theta_f\alpha_i + \varepsilon_i \tag{4.20}$$

where ε_i is the random error vector whose components follow some asymmetric distribution as in (4.8), corresponding to the asymmetric loss minimization for the generalized quantile regression. Equation (4.20) has also been used in Zhou et al. (2008) for a random effects model of functional principal components, where both α_i and ε_i are multivariate normally distributed. Since the signal-plus-noise model (4.20) for generalized quantile regression is not a plausible data generating model but rather an equivalent representation of the asymmetric loss minimization, the EM-algorithm used in Zhou et al. (2008) can not be simply extended and justified in the current context.

4.3.2 Algorithm

This subsection develops an iterative penalized least asymmetrically weighted squares (PLAWS) algorithm for minimizing the penalized loss function defined in (4.18), by defining weights in a similar manner as in (4.10) and (4.12).

We fix the quantile level $\tau \in (0, 1)$. To estimate the expectile curves, for $i = 1, \cdots, N$ and $j = 1, \cdots, T_i$, define the weights

$$w_{ij} = \begin{cases} \tau & \text{if } Y_{ij} > l_{ij}, \\ 1 - \tau & \text{if } Y_{ij} \leq l_{ij}. \end{cases} \tag{4.21}$$

where $l_{ij} = b(t_{ij})^\top\theta_\mu - b(t_{ij})^\top\Theta_f\alpha_i$ is a function of the parameters. To estimate the quantile curves, define the weights

$$w_{ij} = \begin{cases} \dfrac{\tau}{|Y_{ij} - l_{ij}| + \delta} & \text{if } Y_{ij} > l_{ij}, \\ \dfrac{1 - \tau}{|Y_{ij} - l_{ij}| + \delta} & \text{if } Y_{ij} \leq l_{ij}, \end{cases} \tag{4.22}$$

where l_{ij} is defined as in (4.21) and δ is a small positive constant. Using these weights, the asymmetric loss function in (4.17) can be written as the following weighted sum of squares

$$S = \sum_{i=1}^{N}\sum_{j=1}^{T_i} w_{ij}\{Y_{ij} - b(t_{ij})^\top\theta_\mu - b(t_{ij})^\top\Theta_f\alpha_i\}^2, \tag{4.23}$$

and the penalized loss function (4.18) becomes the following penalized weighted

least squares criterion

$$S^* = \sum_{i=1}^{N} (Y_i - B_i\theta_\mu - B_i\Theta_f\alpha_i)^\top W_i (Y_i - B_i\theta_\mu - B_i\Theta_f\alpha_i) + \lambda_\mu\theta_\mu^\top \Omega\, \theta_\mu + \lambda_f \sum_{k=1}^{K} \theta_{f,k}\Omega\, \theta_{f,k}, \tag{4.24}$$

where $W_i = \text{diag}\{w_{i1}, \ldots, w_{iT_i}\}$. Since the weights depend on the parameters, the PLAWS algorithm iterates until convergence between minimizing (4.24) and updating the weights using (4.21) and (4.22).

To minimize (4.24) for fixed weights, we alternate minimization with respect to θ_μ, Θ_f, and α_i. Such minimizations have close-form solutions

$$\begin{aligned}
\widehat{\theta}_\mu &= \left\{\sum_{i=1}^{N} B_i^\top W_i B_i + \lambda_\mu\Omega\right\}^{-1}\left\{\sum_{i=1}^{N} B_i^\top W_i(Y_i - B_i\widehat{\Theta}_f\widehat{\alpha}_i)\right\}, \qquad (4.25)\\
\widehat{\theta}_{f,l} &= \left\{\sum_{i=1}^{N} \widehat{\alpha}_{il}^2 B_i^\top W_i B_i + \lambda_f\Omega\right\}^{-1}\left\{\sum_{i=1}^{N} \widehat{\alpha}_{il} B_i^\top W_i(Y_i - B_i\widehat{\theta}_\mu - B_i Q_{il})\right\},\\
\widehat{\alpha}_i &= (\widehat{\Theta}_f^\top B_i^\top W_i B_i \widehat{\Theta}_f)^{-1}\left\{\widehat{\Theta}_f^\top B_i^\top W_i(Y_i - B_i\widehat{\theta}_\mu)\right\},
\end{aligned}$$

where

$$Q_{il} = \sum_{k\neq l} \widehat{\theta}_{f,k}\widehat{\alpha}_{ik},$$

and $i = 1, \cdots, N$, $k, l = 1, \cdots, K$, $\widehat{\theta}_{f,k}$ is the k-th column of $\widehat{\Theta}_f$.

Any iterative algorithm needs a method of obtaining initial values, which is referred in Appendix. Moreover, one can find the details of the algorithm in Appendix A.2.

4.3.3 Choice of Auxiliary Parameters

In the paper, for simplicity, we use equally spaced knots for the B-splines. The choice of the number of knots to be used is not critical, as long as it is moderately large, since the smoothness of the fitted curves is mainly controlled by the roughness penalty term. For typical sparse functional datasets, 10-20 knots is often sufficient; see Zhou et al. (2008). The optimal choice of the penalty parameter for the single curve estimation used in initialization follows the method in Schnabel and Eilers (2009b). There are several well developed methods for choosing

the auxiliary parameters in the FDA framework, such as, AIC, BIC and cross-validation (CV). In this paper, all the auxiliary parameters, such as the number of principal components/factors to be included, and the penalty parameters λ_μ and λ_f, will be chosen via the 5-fold cross-validation by minimizing the cross-validated asymmetric loss function.

4.4 Simulation

In this section, simulations are set up to investigate the performance of the FDA method to estimate the generalized quantile curves. We compare the estimation results from individual estimation of the expectile curves and the FDA jointly estimation method. The general setup is written as

$$Y_{ij} = \mu(t_j) + f_1(t_j)\alpha_{1i} + f_2(t_j)\alpha_{2i} + e_{ij} \tag{4.26}$$

We also assume that for $i = 1, \cdots, N$, t_j is equidistant in $[0, 1]$, and $j = 1, \cdots, T$. The mean curve is $\mu(t) = 1 + t + \exp\{-(x - 0.6)^2/0.05\}$ and we set $f_1(t) = \sin(2\pi x)/\sqrt{0.5}$ and $f_2(t) = \cos(2\pi x)/\sqrt{0.5}$. Let $\alpha_{1i} \sim \mathrm{N}(0, 36)$, $\alpha_{2i} \sim \mathrm{N}(0, 9)$ and $e_{it} \sim (0, \sigma^2)$.

Different scenarios are built up by changing the sample size and the distribution of the error terms, that is we have large sample size and small sample size. While, we also assume the error terms to be normally distributed or time varying normally distributed, or it can be t distributed. The distributions of the error terms are designed:

- $e_{it} \sim N(0, 0.5)$
- $e_{it} \sim N(0, \mu(t) \times 0.5)$
- $e_{it} \sim t(5)$
- small sample: $N = 20, T = 100$
- large sample: $N = 40, T = 150$

The theoretical τ-th quantile and expectile, which we consider as a reference for the estimators:

$$l_{ij} = l_i(t_j) = \mu(t_j) + f_1(t_j)\alpha_{1i} + f_2(t_j)\alpha_{2i} + \varepsilon_{i\tau}$$

where $i = 1, \cdots, N$, $t_j = 1, \cdots, T$ and ε_τ represents the corresponding τ-th theoretical quantile and expectile of e_{it}. To design the criteria for the comparison, we define

- The individual curve:

$$\begin{aligned} \text{Theoretical curve:} \quad l_i &= \mu + \sum_{k=1}^{K} f_k \alpha_{ik} \\ \text{FDA estimated:} \quad \widehat{l}_{i,fda} &= B_i \widehat{\theta}_\mu + B_i \widehat{\Theta}_f \widehat{\alpha}_i \\ \text{Individually estimated:} \quad \widehat{l}_{i,in} &: \text{ Referred to Section 4.2.} \end{aligned}$$

- The mean curve:

$$\begin{aligned} \text{Theoretical curve:} \quad m &= \mu(t) + e_\tau \\ \text{FDA estimated:} \quad m_{fda} &= \frac{1}{N} \sum_{i=1}^{N} B_i \widehat{\theta}_\mu \\ \text{Individually estimated:} \quad m_{in} &= \frac{1}{N} \sum_{i=1}^{N} \widehat{l}_{i,in} \end{aligned}$$

One can evaluate the jointly estimation method (FDA) via comparing its performance with other existing individual estimation methods, taking the method in Schnabel and Eilers (2009b) as an example. Two aspects of comparison are considered in the paper. We firstly compare the theoretical mean curve of the estimated mean curve by the FDA method and the estimated mean curve from single curve estimation. Meanwhile, we are also interested in how far the estimated individual expectile curves by the two methods are from the theoretical expectile curves. We evaluate the estimators in each scenario in terms of the MSE (mean square errors).

Table 4.1 and **??** show the results of different methods based on 200 runs of simulations in each scenario. The first table summarize the results for 95% expectile regressions. We compare separately the common shape curve and the individual curves in each scenario. Moreover, the MSE from the small sample size and the large sample size are shown in each table. Obviously, when we increase the sample size, i.e. to increase the number of individuals as well as the observations in each individual, the MSE from FDA and single estimation method are getting smaller. It is worth noting that when the error term is normally distributed, both methods perform well, that is to say, they both can provide reliable results when the error term is not very volatile. While our FDA method provides even better result, especially when the sample size is large. However, when the volatility of the random errors becomes larger, for instance, as we change the error term to the time-varying with the mean of the error term as designed in scenario 2 or we change the error term to be t distributed in scenario 3, summarized in Table 4.1, the MSE from both methods becomes larger than that from the normally distributed. However,

	Individual		*Mean*	
Sample Size	*FDA*	*Single*	*FDA*	*Single*
$N = 20, T = 100$	0.0815	0.1407	0.0491	0.0288
$N = 40, T = 150$	0.0189	0.0709	0.0028	0.0063
$N = 20, T = 100$	0.1571	0.2957	0.0272	0.0377
$N = 40, T = 150$	0.1002	0.2197	0.0118	0.0172
$N = 20, T = 100$	0.2859	0.5194	0.0454	0.0556
$N = 40, T = 150$	0.1531	0.4087	0.0181	0.0242

Table 4.1: The mean squared errors (MSE) of the FDA and the single curve estimation for 95% expectile curves when the error term is normally distributed with mean 0 and variance 0.5 (Top), with variance $\mu(t) \times 0.5$ (Middle) and $t(5)$ distribution (Bottom).

if we increase the sample size, the difference of the between estimated common structure curves by FDA with the theoretical one is quite small comparing with that from the individual estimation method. One may detect that the estimated individual curves from FDA method also outperform the single curve estimation method. More details is summarized in the corresponding tables. Table **??** summarized the result for quantile curves when the error term is normally distributed and $t(5)$ distributed. Similarly, the FDA method has a better performance than single curve estimation in both large sample size and small sample size. The FDA method has a quite smaller MSE in comparison with the individual estimation, especially when the error term with volatile distributed. One can say that the performance of FDA is similar for the expectile curves. However, the MSE becomes slightly larger than that of the expectile regressions, since for the very high quantile, the results may contain a larger bias due to not many observations in the tail.

Figure 4.1 shows the estimated common structure curve $\mu(t)$ by FDA method together with the designed common shape curves for the 95% expectile curves with the error term normally distributed, together with the 95% pointwise confidence intervals. The blue line is the estimated common shape curve, and the black solid one is the designed line respectively in both plots. The red dashed lines represent the 95% confidence intervals. The left plot displays the result of the small sample size, and the right one is for the large sample size. It is obvious that both estimated common structure curves fit the respective true ones well, while the one from the large data set fits slightly better, moreover, the confidence intervals are smaller than that of the small sample size. Figure 4.2 tells us the estimated factor curves and the real factor curves and corresponding confidence

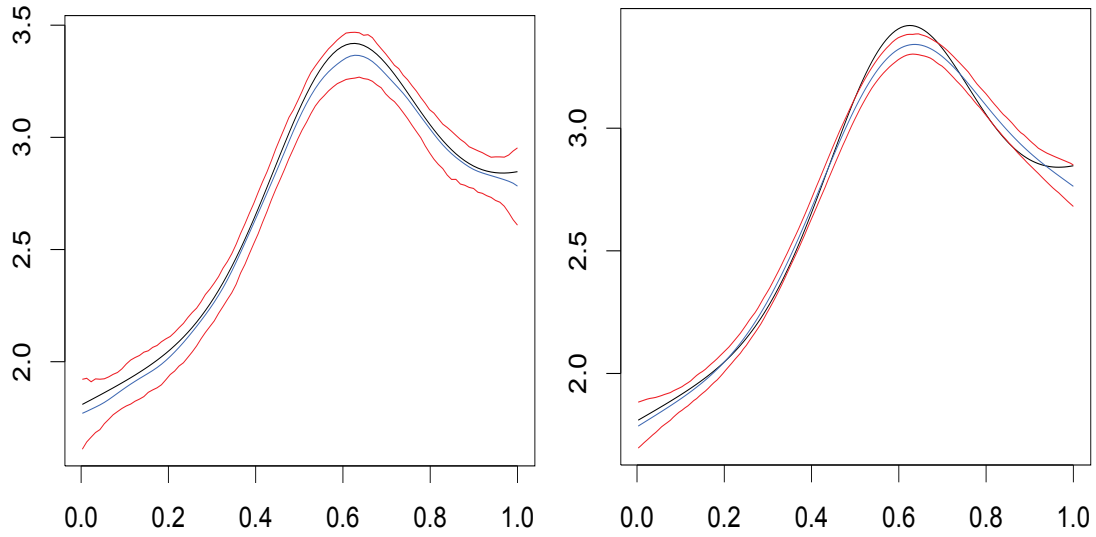

Figure 4.1: The estimated μ (blue), the real μ (black) and the $5\% - 95\%$ pointwise confidence intervals for 95% expectile curves when the error term is normally distributed with mean 0 and variance 0.5. The sample size is $N = 20, M = 100$ (Left) and $N = 40, M = 150$ (Right).

intervals for the 95% expectile curves. The solid black lines stand for the first and the second factor, the solid blue lines are the corresponding estimated factors by FDA method. The dashed red lines are the respective confidence interval. The left figure describes the result from the small sample size, and the right one captures the performance for the large sample size. Similarly, the confidence intervals from large sample size become smaller, i.e. the estimation can be more accurate. Figure 4.3 shows the estimated common structure curves of 95% quantile curves for the simulation when the error term is normally distributed with mean 0 and variance 0.5 for both small sample size and large sample size. Figure 4.4 describes the corresponding estimated factor curves and the real factor curves. It is no doubt that the performance of FDA method for quantile regression is comparable with the expectile regression.

4.5 Application

In this section, we apply the proposed FDA method to study the daily temperature of 150 weather stations in China in year 2010. The locations of these weather stations are shown in Figure 4.5. The data is obtained from China Meteorological Administration.

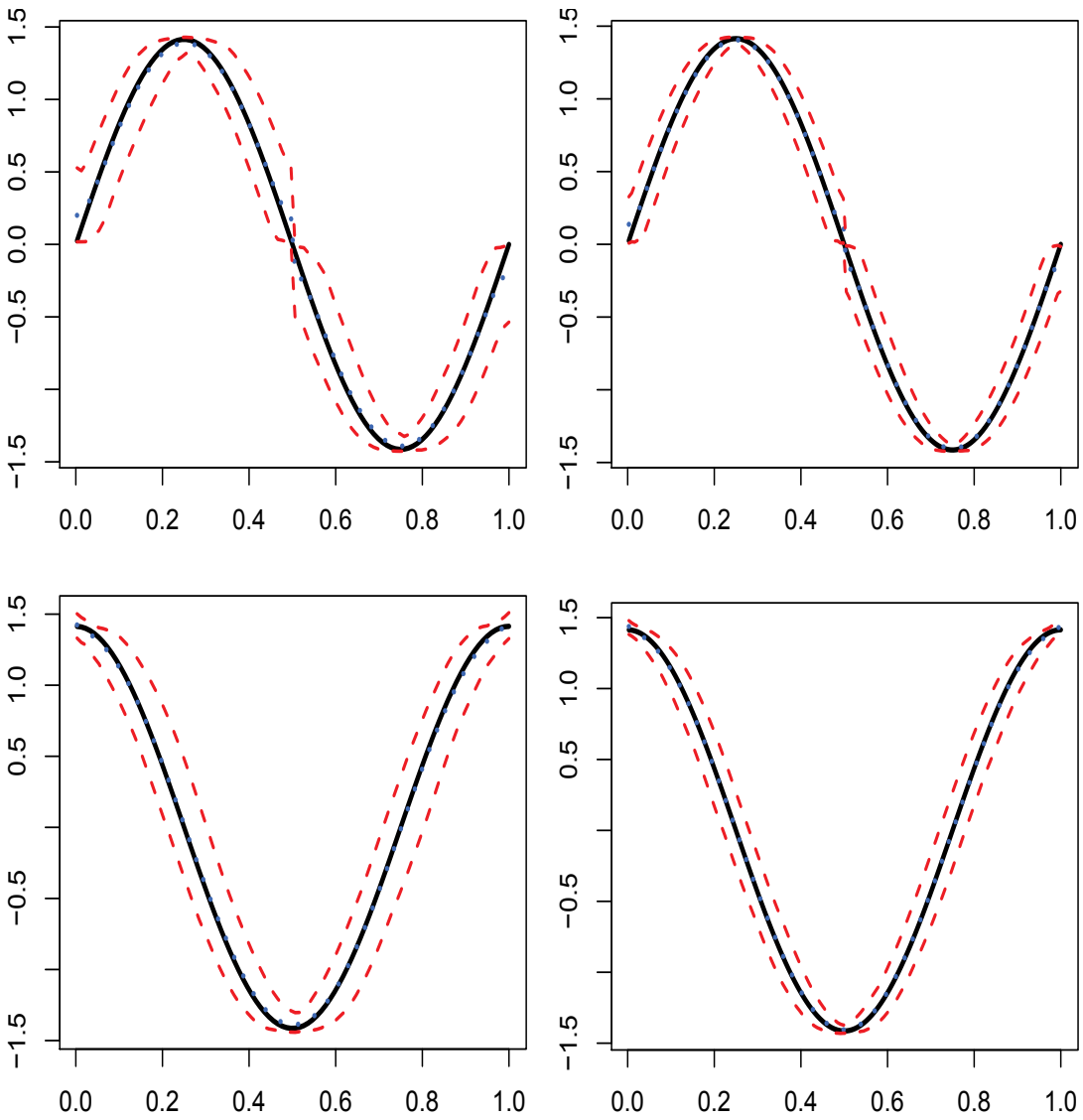

Figure 4.2: The estimated first factor f_1 (blue), the real f_1 (black) and the $5\% - 95\%$ pointwise confidence intervals for 95% expectile curves (Top). The estimated second factor f_2 (blue), the real f_2 (black) and the $5\% - 95\%$ pointwise confidence intervals for 95% expectile curves (Bottom). The error term is normally distributed with mean 0 and variance 0.5. The sample size is $N = 20, M = 100$ (Left) and $N = 40, M = 150$ (Right).

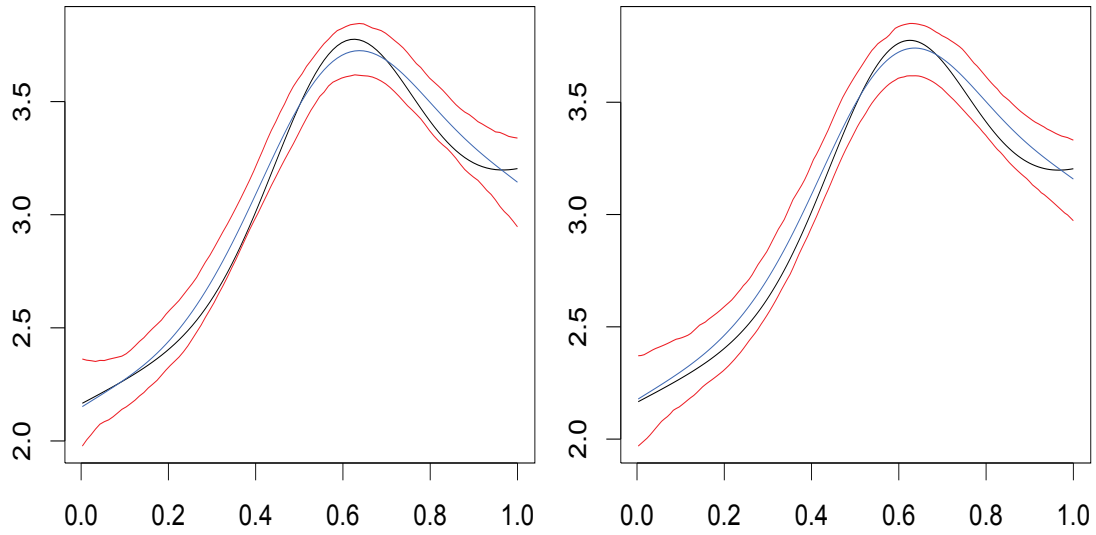

Figure 4.3: The estimated μ (blue), the real μ (black) and the $5\% - 95\%$ pointwise confidence intervals for 95% quantile curves with error term normally distributed with mean 0 and variance 0.5. The sample size is $N = 20, M = 100$ (Left) and $N = 40, M = 150$ (Right).

Recently, a market for trading on temperature events, or more popularly, buying and selling temperature, has emerged. Financial contracts, whose value depends on certain temperature events are called temperature derivatives, and provide a financial protection against undesirable weather events or a tool for speculating in future temperature levels. Temperature derivatives may be attractive for industries with temperature-dependent profits, such as tourism or the energy sector.

The volatility of temperature is crucial to the weather derivatives pricing, and it also provides evidence to crop insurance, especially in China. Therefore, to study the volatility of temperature motivates us to hedge the corresponding weather risk. Further, we analyze the functional factors to explain these curves and the corresponding fixed effects for all the weather stations. It would be meaningful to see the factors affecting the variation of the temperature.

The temperature in all 150 weather stations shows a clear seasonable pattern-lower in winter and higher in summer. The temperature record for each station also has strong autocorrelation. Therefore, the volatility of the temperature would be the residuals after removing the seasonal effect and the autoregressive effect. The technology has already been well documented in the literature. Campbell and Diebold (2005) introduced the AR-GARCH model to capture the dynamics of the average day temperature. And more detailed description is written in Härdle and

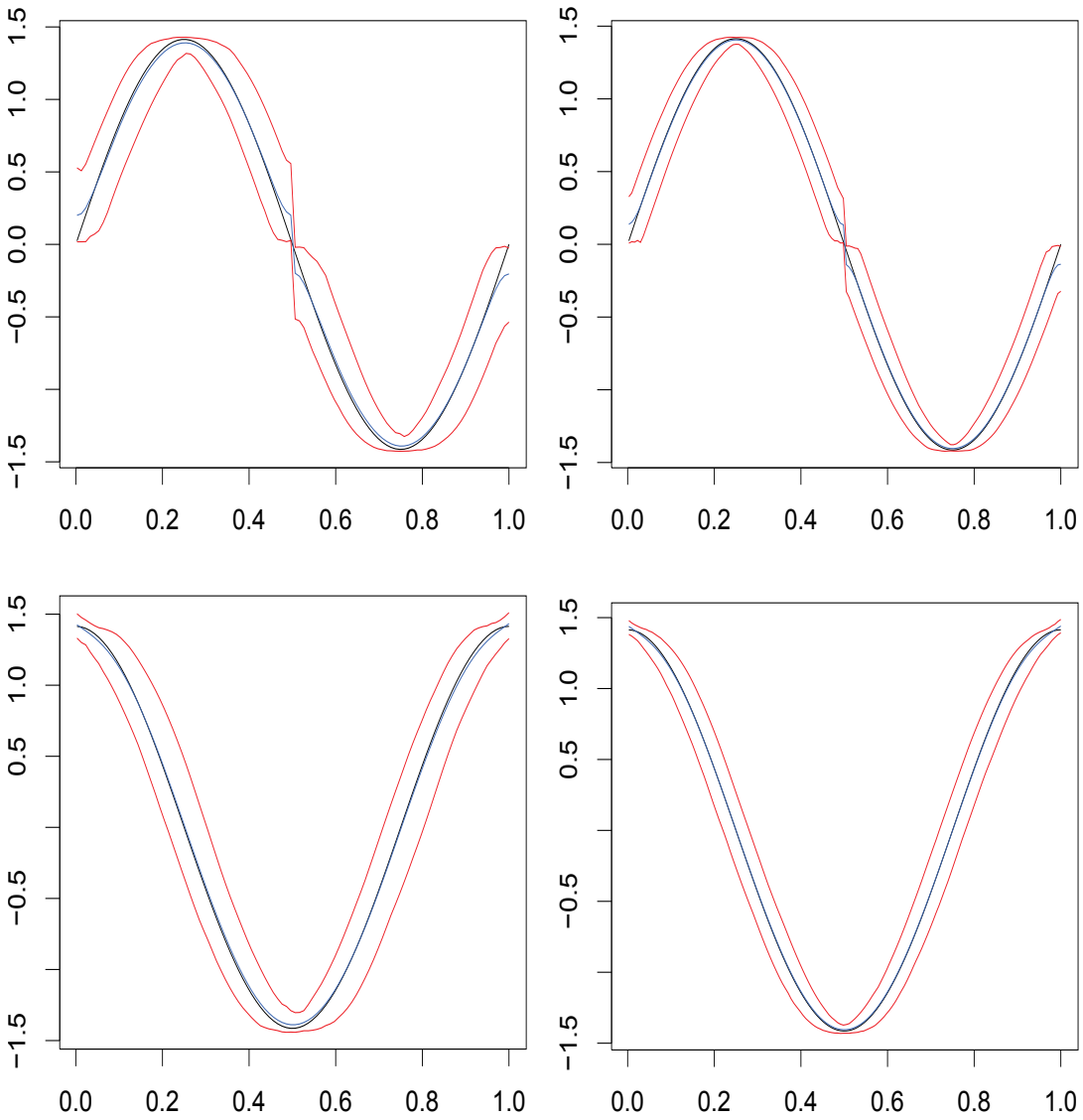

Figure 4.4: The estimated first factor f_1 (blue), the real f_1 (black) and the 5%−95% pointwise confidence intervals for 95% quantile curves (Top). The estimated second factor f_2 (blue), the real f_2 (black) and the 5%−95% pointwise confidence intervals for 95% quantile curves (Bottom). The error term is normally distributed with mean 0 and variance 0.5. The sample size is $N = 20, M = 100$ (Left) and $N = 40, M = 150$ (Right).

Figure 4.5: 150 Weather Stations in China

López-Cabrera (2011). The temperature on day t for city i is to decompose the the temperature into two parts, the seasonal part and the random part.

- The temperature at T_{it} on day t:

$$T_{it} = X_{it} + \Lambda_{it}$$

- The seasonal effect Λ_{it}:

$$\Lambda_{it} = a_i + b_i t + \sum_{m=1}^{M} c_{im} \cos\{\frac{2\pi(t - d_{im})}{m \cdot 365}\}$$

- X_{it} follows an $AR(p_i)$ process:

$$\begin{aligned} X_{it} &= \sum_{j=1}^{p_i} \beta_{ij} X_{i,t-j} + \varepsilon_{it} \qquad (4.27) \\ \widehat{\varepsilon}_{it} &= X_{it} - \sum_{j=1}^{p_i} \hat{\beta}_{ij} X_{t-j} \end{aligned}$$

The variation of the temperature is expressed as $\widehat{\varepsilon}_{it}$ in (4.27). To understand its performance, we investigate different percentages of expectile curves, such as, the

95%, 75%, 50% and the 25% expectile curves for each weather station. Following the aforementioned algorithm, we choose all the auxiliary parameters by 5-fold CV. In the paper, we choose $K = 3$ principal components for each percentage of expectile curves to explain the large enough variance for each station. To note that, we use different smoothing parameters λ's for different expectiles.

Figure 4.6 describes the estimated expectile curves for these 150 weather stations in 2010, respectively, 25%, 50%, 75% and 95%. The grey lines in each plot are the estimated individual expectile curves by our FDA method. The dashed red lines are the pointwise intervals that cover 95% of the mass of the distributions for the individual curves. It is obvious that these four expectile curves perform slightly differently. The 25%,75% and 95% expectile curves vary in a relative larger range comparing with the 50% expectile, especially the 95% expectile curves. Note that in these four plots, for the upper tails and the mean curve, i.e, 75%, 95% and 50% expectile curves, they have low part around observation 200, to say it is around the beginning of July, the summer time, and two peaks appear around observation 100 and 300, which are respectively around the beginning of April and the end of November, i.e. the spring and the fall. Hence, one can conclude that during summer time, the volatility of the temperature in China does not change a lot, while in spring and fall, it is relatively a little more volatile. While, for the lower part of the distribution of the volatility, i.e. 25% expectile curves, we only have one significant peak at position 300.

The estimated factor curves for the expectile curves are shown in Figure 4.7. Generally, it is known that there are several factors that influence temperature, such as latitude, altitude, distance from large bodies of water, ocean currents, mountain barriers, air masses, prevailing wind system, and human behaviors. The first factor has similar pattern in all these 4 expectiles, which are very flat, except for the 50% expectile curves, further, they vary little around 1. While, the first factor for the 50% expectile looks a little more volatile, which displays a U shape. Therefore, the first factor for the lower and upper tails of the volatility of temperature is time-invariant, which would be the geographical factors, such as the latitude and the attitude and so on. The second factor for each expectile has similar pattern, except for 95% expectile and they are more volatile comparing with the first factor. The second factor is time varying, which can be explained by the seasonal factor, while the 95% performs the opposite. The third factor shows similar trend for the expectile, except for the 75% expectile curves. They are decreasing with time. In sum, one may say that the volatility of all these 4 different percentages of the expectile curves is influenced by similar factors according to the shape of the estimated factors.

Figure 4.8 to 4.10 describe respectively to the three estimated fixed effect α_i's for all the expectile curves, which are also projected to the China map. We draw

each point with the heat colors. One can notice that there is a clear pattern for α_1 in Figure 4.8. In the low percentage of the expectile, i.e. 25% and 50% expectile curves, in the north and northeast part, α_1 are negative values, while in the south part, the values are positive. Oppositely, for the 75% and 95% expectile curves, the scores are positive in the north part and negative in the south part of China. This is consistent with the geographical knowledge. For the extreme values of temperature always show in the north. The temperature generally has larger change than that from the south part. The second mapping plot also shows different allocations of the scores. For the lower expectile, the middle part show larger values than the north and south parts, which are opposite with the 50%, 75% and 95% expectile curves. The third mapping does not show very clear pattern as the first two has. Moreover, please note that the scale of all α_i for 50% expectile curves is smaller than others.

4.6 Conclusion

In this paper, we provide a novel methodology to estimate the generalized quantile curves when a family of random curves are available. Further, we deduce the close form solutions for the generalized quantile curves. We use all the information from the observed data, and estimate both the common shape curve and the departure curves for each individual expectile curve. We found our method outperforms the individually estimate method for generalized quantile regression curves, which is verified in the simulations. In the application, we investigated different factors to influence the volatility of the temperature. Roughly, the risk drivers for different percentages of expectiles are quite similar. The factors which influence both the temperature and the volatility of the temperature, can be expressed as three factors. One is time invariant, and the other two are time varying.

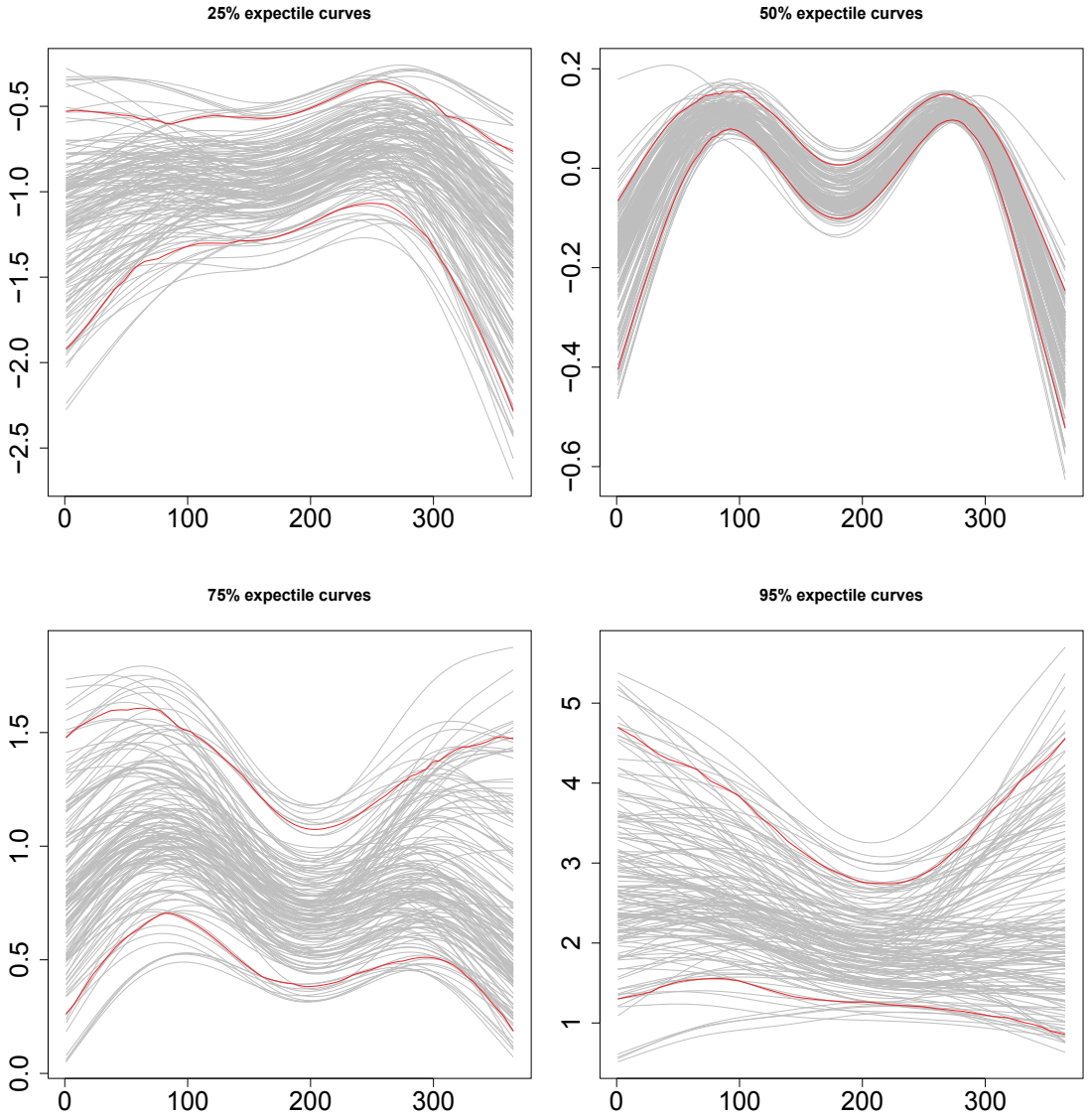

Figure 4.6: The estimated expectile curves of the volatility of the temperature for 150 weather stations cities in China in 2010 for the $25\%, 50\%, 75\%, 95\%$ expectiles. The grey lines stand for the individual expectile curves estimated by the FDA method.

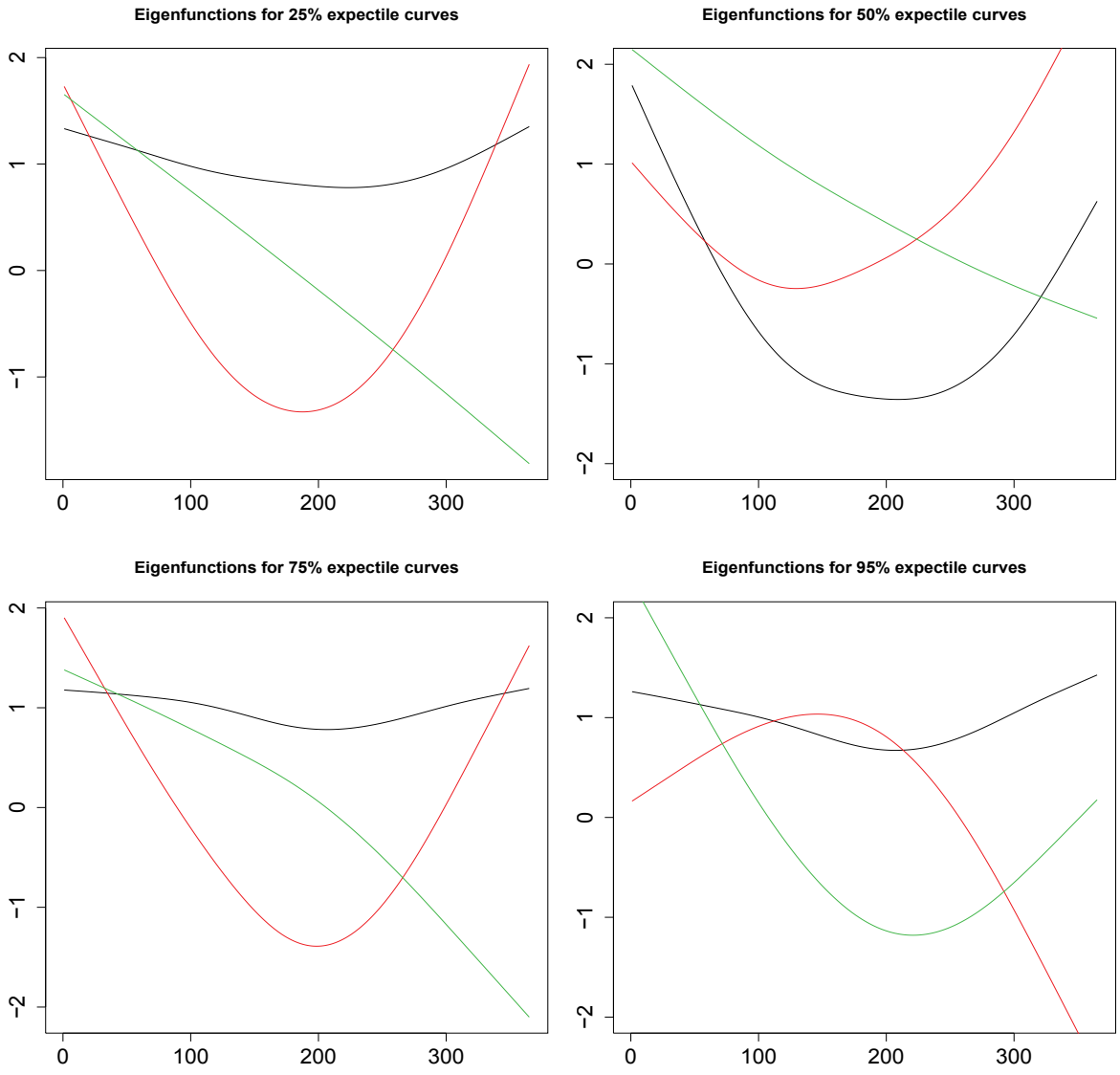

Figure 4.7: The estimated three factors for the $25\%, 50\%, 75\%, 95\%$ expectiles (from left to right) curves of the volatility of the temperature of China in 2010 with the data from 150 weather stations. The black solid curve is the first factor, the red dashed curve is the second and the green dotted curve represents the third factor.

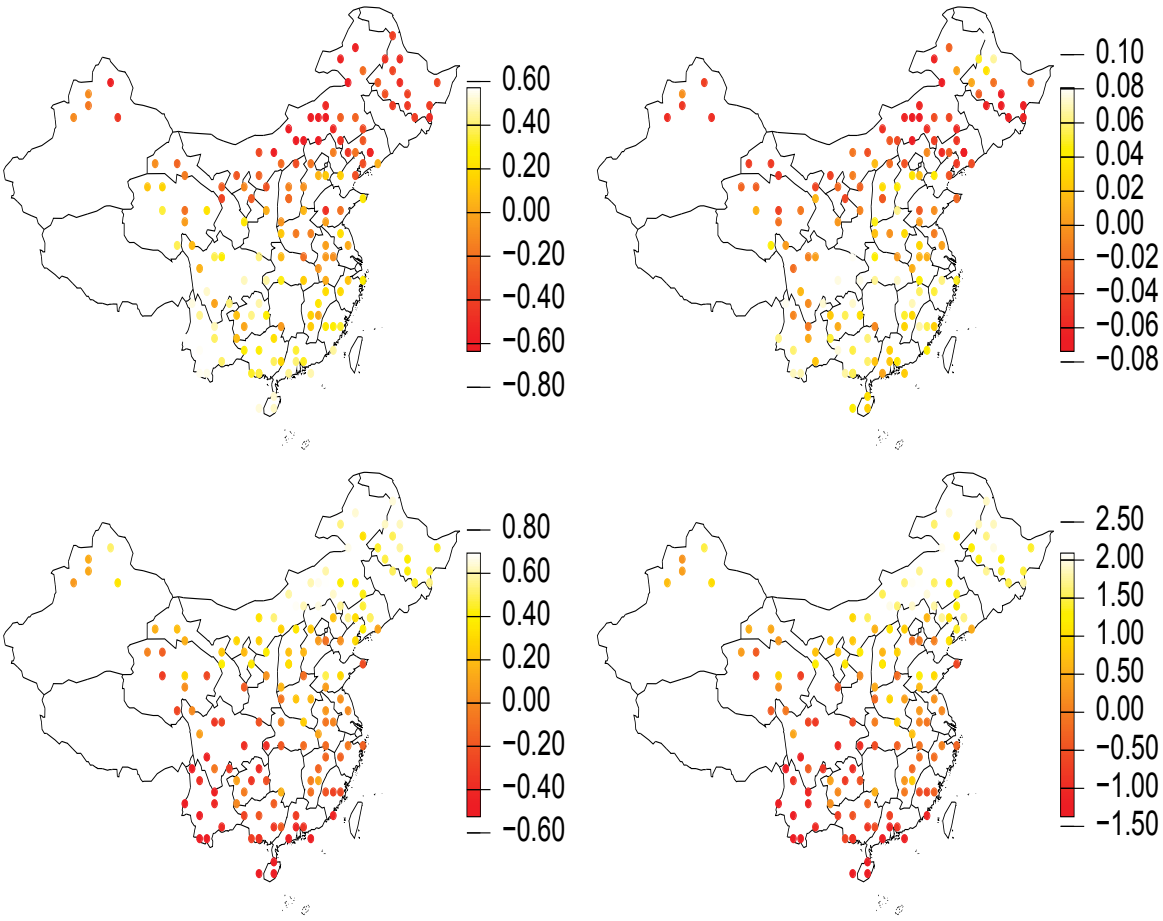

Figure 4.8: The estimated fixed effect α_1 for the 25%, 50%, 75% and 95% expectile curves of the temperature variation.

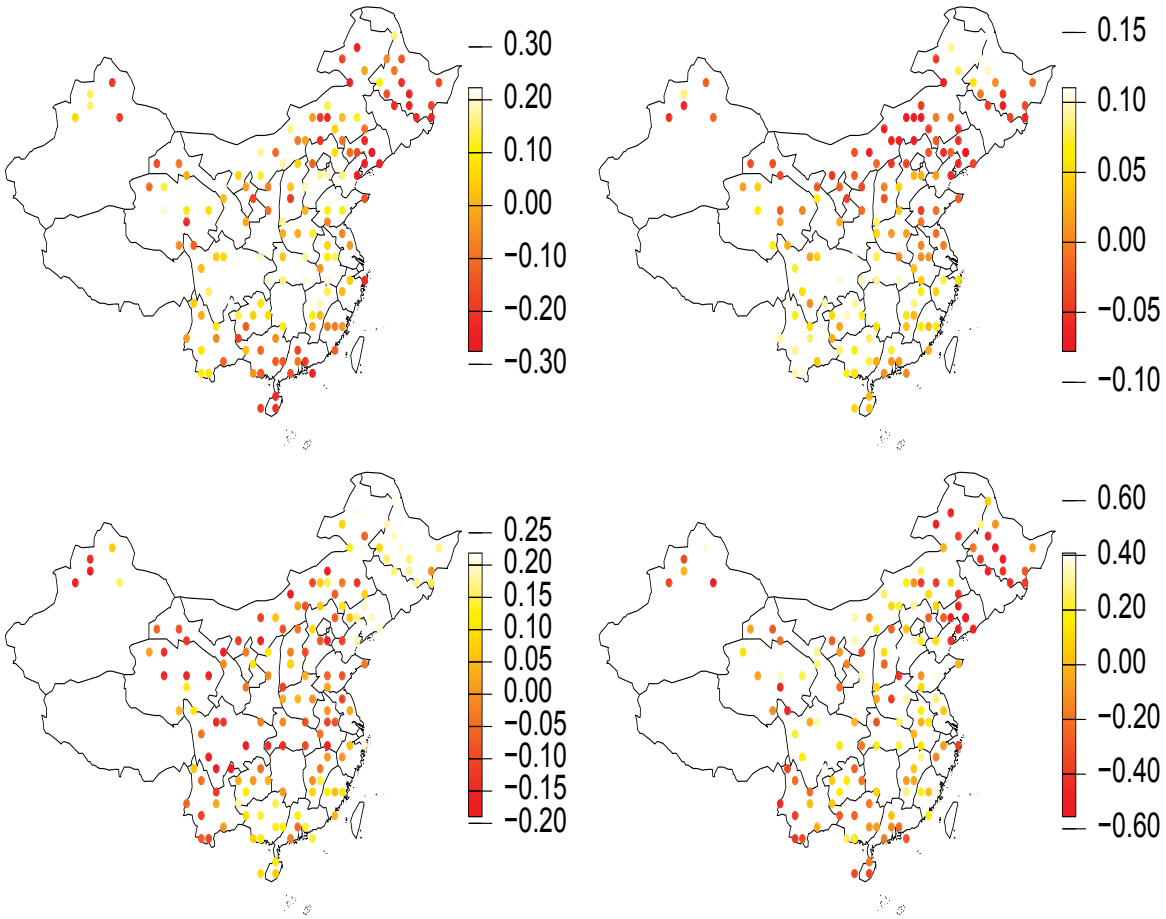

Figure 4.9: The estimated fixed effect α_2 for the 25%, 50%, 75% and 95% expectile curves of the temperature variation.

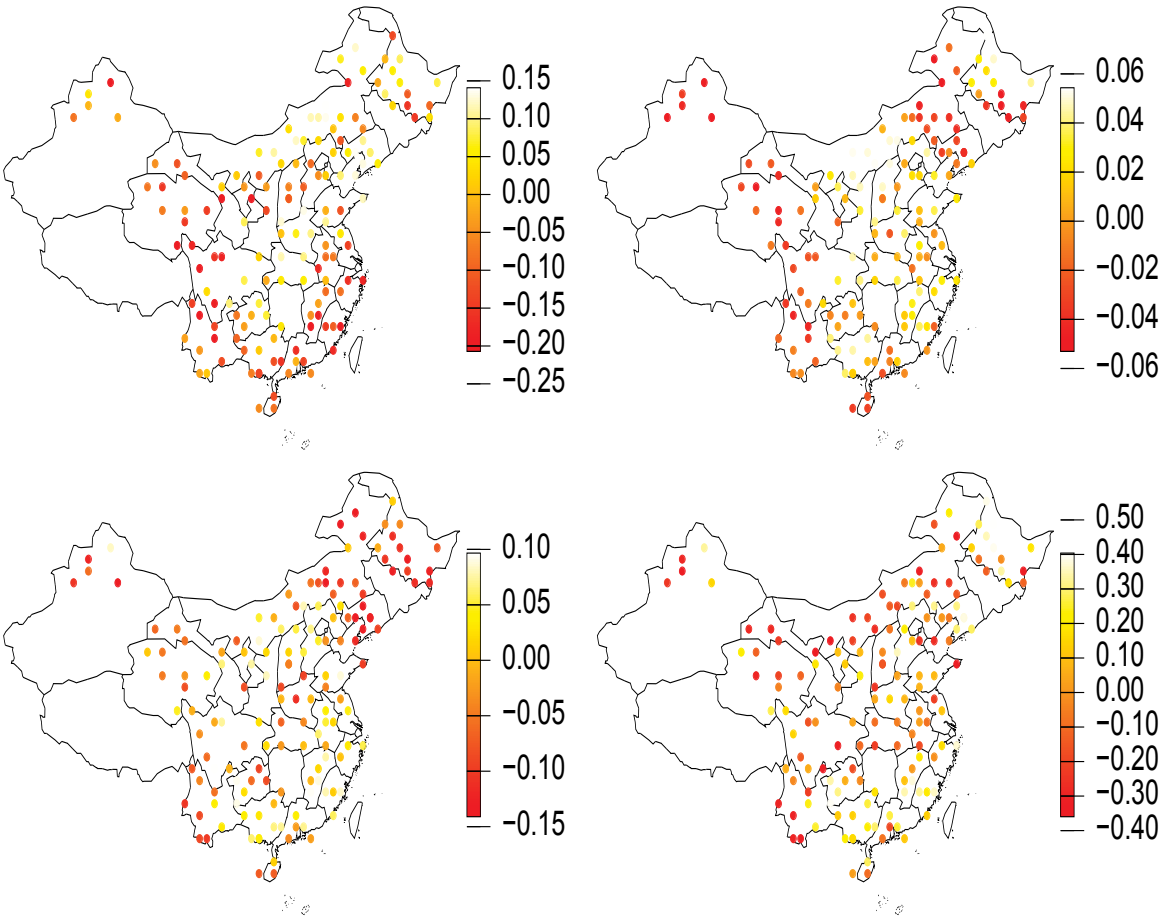

Figure 4.10: The estimated fixed effect α_3 for the 25%, 50%, 75% and 95% expectile curves of the temperature variation.

Bibliography

Y. A¨it-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9:385–426, 1996.

Z. Anastasiadou and B. López-Cabrera. Statistical modelling of temperature risk. *SFB discussing paper 2012-029*, 2012.

A. Ang and G. Bekaert. Regime switches in interest rates. *Journal of Business and Economic Statistics*, 20:163–182, 2002.

M. Arapis and J. Gao. Empirical comparisons in short-term interest rate models using nonparametric methods. *Journal of Financial Econometrics*, 4:310–345, 2006.

V. Bansal and H. Zhou. Term structure of interest rates with regime shifts. *The Journal of Finance*, 57(5):1997–2043, 2002.

F.S. Benth, J.S. Benth, and S. Koekebakker. Putting a price on temperature. *Scandinavian Journal of Statistics*, 34(4):746–767, 2007.

P. Bickel and M. Rosenblatt. On some global measures of the deviation of density function estimatiors. *Annals of Statistics*, 1:1071–1095, 1973.

F. Black and P. Karasinski. Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47:52–59, 1991.

J. Breckling and R. Chambers. M-quantiles. *Biometrika*, 74(4):761–772, 1988.

Z. Cai. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188, 2007.

S. Campbell and F. Diebold. Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 469:6–16, 2005.

H. Cardot, C. Crambes, and P. Sarda. Quantile regression when the covariates are funtions. *Nonparametric Statistics*, 17(7):841–856, 2005.

P. C´ižek, W. Härdle, and V. Spokoiny. Adaptive pointwise estimation in time-inhomogeneous condtional heteroscedasticity models. *Econometric Journal*, 12: 1–25, 2009.

G. Claeskens and I. Van Keilegom. Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics*, 31(6):1852–1884, 2003.

S. Csörgö and P. Hall. Upper and lower classes for triangular arrays. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61:207–222, 1982.

S. R. Das. The surprise element: jumps in interest rates. *Journal of Econometrics*, 106:27–65, 2002.

P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29(2):1–48, 2001.

F.X. Diebold and A. Inoue. Long memory and regime switching. *Journal of Econometrics*, 105:131–159, 2001.

B. Efron. Regression percentiles using asymmetric squared loss. *Statistica Sinica*, 1:93–125, 1991.

P.H.C. Eilers and B.D. Marx. Flexible smoothing with B-splines and penalties. *Journal of the American Statistical Association*, 11:89–121, 1996.

J. Fan and Y. Wang. Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, 102:1349–1362, 2007.

J. Fan, J. Jiang, C. Zhang, and Z. Zhou. Time-dependent diffusion models for term structure dynamics. *Statistica Sinica*, 13:965–992, 2003.

J. Franke and P. Mwita. Nonparametric estimates for conditional quantiles of time series. *Report in Wirtschaftsmathematik 87, University of Kaiserslautern*, 2003.

E. Giacomini, W. Härdle, and V. Spokoiny. Inhomogeneous dependence modeling with time-varying copulae. *Journal of Business and Economic Statistics*, 27(2): 224–234, 2009.

A. Goyal and I. Welch. Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–654, 2003.

M. Guo and W. Härdle. Adaptive interest rate modeling. *SFB* 649 *Discussion Paper 2010-29*, 2010.

M. Guo and W. Härdle. Simultaneous confidence bands for expectile functions. *Advances in Statistical Analysis, DOI: 10.1007/s10182-011-0182-1*, 2011.

W. Härdle. Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis*, 29:163–179, 1989.

W. Härdle and B. López-Cabrera. Implied market price of weather risk. *Applied Mathematical Finance*, 5:1–37, 2011.

W. Härdle and S. Luckhaus. Uniform consistency of a class of regression function estimators. *Annals of Statistics*, 12:612–623, 1984.

W. Härdle and S. Song. Confidence bands in quantile regression. *Econometric Theory*, 3:1–21, 2009.

W. Härdle, P. Janssen, and R. Serfling. Strong uniform consistency rates for estimators of conditional functionals. *Annals of Statistics*, 16:1428–1429, 1988.

W. Härdle, Y. Ritov, and S. Song. Partial linear regression and bootstrap confidence bands. *SFB* 649 *Discussion Paper 2010-002, submitted to Journal of Multivariate Analysis*, 2010.

W. Härdle, O. Okhrin, and Y. Okhrin. Time varying hierarchical archimedean copulae. *SFB discussion paper, submitted to Computational Statistics and Data Analysis*, 2011.

P. Huber. *Robust Statistics*. Wiley, New York., 1981.

J. Hull and A. White. Pricing interest rate derivative securities. *Review of Financial Studies*, 3(4):573–592, 1990.

D.R. Hunter and K. Lange. Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics*, 9(1):60–77, 2000.

G. James, T. Hastie, and C. Sugar. Principle component models for sparse functinal data. *Biometrika*, 87:587–602, 2000.

M.S. Johannes. The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance*, 51(1):227–260, 2004.

G. Johnston. Probabilities of maximal deviations of nonparametric regression function estimates. *Journal of Multivariate Analysis*, 12:402–414, 1982.

M. Jones, O. Lamont, and R. Lumsdaine. Macroeconomic news and bond market volatility. *Journal of Financial Economics*, 47:315–337, 1998.

M.C. Jones. Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters*, 20:149–153, 1994.

R. Koenker. *Quantile Regression*. Cambridge University Press, Cambridge, 2005.

R. Koenker and G. W. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.

R. Koenker and J. A. F. Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94: 1296–1310, 1999.

C. M. Kuan, Y. H. Yeh, and Y. C. Hsu. Assesing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics*, 150:261–270, 2009.

M. Lettau and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849, 2001.

D. Mercurio and V. Spokoiny. Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics*, 32(2):577–602, 2004.

W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847, 1987.

M. Odening, E. Berg, and C. Turvey. Management of climate risk in agriculture. *Special Issue of the Agricultural Finance Review*, 68(1):83–97, 2008.

H.C. Ombao, J.A. Raz, R. von Sachs, and B.A. Malow. Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association*, 96:543–560, 2001.

M. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 32:1065–1076, 1962.

B.S. Paye and A. Timmermann. Instability of return prediction models. *Journal oF Empirical Finance*, 13:274–315, 2006.

J.O. Ramsay and B.W. Silverman. *Functional data analysis, 2nd ed.* New York: Springer, 2005.

M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23:470–472, 1952.

D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, 2003.

S.K. Schnabel and P.H.C. Eilers. An analysis of life expectancy and economic production using expectile frontier zones. *Demographic Research*, 21:109–134, 2009a.

S.K. Schnabel and P.H.C. Eilers. Optimal expectile smoothing. *Computational Statistics and Data Analysis*, 53:4168–4177, 2009b.

V. Spokoiny. Multiscale local change point detection with applications to value-at-risk. *The Annals of Statistics*, 37(3):1405–1436, 2009.

R. Stanton. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52:1973–2002, 1997.

J.W. Taylor. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6:231–252, 2008.

G. Tusnady. A remark on the approximation of the sample distribution function in the multidimensional case. *Periodica Mathematica Hungarica*, 8:53–55, 1977.

O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

Q. Yao and H. Tong. Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, 6(2-3):273–292, 1996.

B. Zhang. Nonparametric regression expectiles. *Nonparametric Statistics*, 3:255–275, 1994.

L. Zhou, J. Huang, and R. Carroll. Joint modelling of paired sparse functional data using principle components. *Biometrika*, 95(3):601–619, 2008.

Appendix A

A.1 Proofs of Chapter 3

Proof of Theorem 3.2.1. By the definition of $v_n(x)$ as a zero of (3.5), we have, for $\varepsilon > 0$,

$$\text{if } v_n(x) > v(x) + \varepsilon, \text{and then } H_n\{v(x) + \varepsilon, x\} > 0. \tag{A.1.1}$$

Now

$$H_n\{v(x) + \varepsilon, x\} \leqslant H\{v(x) + \varepsilon, x\} + \sup_{\theta \in I} |H_n(\theta, x) - H(\theta, x)|. \tag{A.1.2}$$

Also, by the identity $H\{v(x), x\} = 0$, the function $H\{v(x) + \varepsilon, x\}$ is not positive and has a magnitude $\geqslant m_1 \tilde{q} \varepsilon$ by assumption $(A6)$ and (3.12), for $0 < \varepsilon < \delta_1$. That is, for $0 < \varepsilon < \delta_1$,

$$H\{v(x) + \varepsilon, x\} \leqslant -m_1 \tilde{q} \varepsilon. \tag{A.1.3}$$

Combining (A.1.1), (A.1.2) and (A.1.3), we have, for $0 < \varepsilon < \delta_1$:

$$\text{if } v_n(x) > v(x) + \varepsilon, \text{and then } \sup_{\theta \in I} \sup_{x \in J} |H_n(\theta, x) - H(\theta, x)| > m_1 \tilde{q} \varepsilon.$$

With a similar inequality proved for the case $v_n(x) < v(x) + \varepsilon$, we obtain, for $0 < \varepsilon < \delta_1$:

$$\text{if } \sup_{x \in J} |v_n(x) - v(x)| > \varepsilon, \text{and then } \sup_{\theta \in I} \sup_{x \in J} |H_n(\theta, x) - H(\theta, x)| > m_1 \tilde{q} \varepsilon. \tag{A.1.4}$$

It readily follows that (A.1.4), and (3.11) imply (3.13). ■

Below we first show that $\|R_n\|_\infty = \sup_{x \in J} |R_n(x)|$ vanishes asymptotically faster than the rate $(nh \log n)^{-1/2}$; for simplicity we will just use $\|\cdot\|$ to indicate the sup-norm.

LEMMA A.1.1 *For the remainder term $R_n(t)$ defined in (3.16) we have*

$$\|R_n\| = o_p\{(nh \log n)^{-1/2}\}. \tag{A.1.5}$$

Proof. First we have by the positivity of the kernel K,

$$\begin{aligned}\|R_n\| &\leqslant \Big[\inf_{0\leqslant x\leqslant 1}\{|D_n(x)|\cdot p(x)\}\Big]^{-1}\{\|H_n\|\cdot\|p-D_n\|+\|D_n\|\cdot\|\,\mathsf{E}\,H_n\|\}\\ &\quad +C_1\cdot\|v_n-l\|^2\cdot\Big\{\inf_{0\leqslant t\leqslant 1}|D_n(x)|\Big\}^{-1}\cdot\|f_n\|,\end{aligned}$$

where $f_n(x)=(nh)^{-1}\sum_{i=1}^n K\{(x-X_i)/h\}$.

The desired result (A.1.1) will then follow if we prove

$$\|H_n\| = \wp_p\{(nh)^{-1/2}(\log n)^{1/2}\} \tag{A.1.6}$$
$$\|p-D_n\| = \wp_p\{(nh)^{-1/4}(\log n)^{-1/2}\} \tag{A.1.7}$$
$$\|\,\mathsf{E}\,H_n\| = \mathcal{O}(h^2) \tag{A.1.8}$$
$$\|v_n-v\|^2 = \wp_p\{(nh)^{-1/2}(\log n)^{-1/2}\} \tag{A.1.9}$$

Since (A.1.8) follows from the well-known bias calculation

$$\mathsf{E}\,H_n(x)=h^{-1}\int K\{(x-u)/h\}\,\mathsf{E}[\psi\{y-v(x)\}|X=u]f_X(u)du=\mathcal{O}(h^2),$$

where $\mathcal{O}(h^2)$ is independent of x in Parzen (1962), we have from assumption $(A2)$ that $\|\,\mathsf{E}\,H_n\|=\wp_p\{(nh)^{-1/2}(\log n)^{-1/2}\}$.

According to Lemma $A.3$ in Franke and Mwita (2003),

$$\sup_{x\in J}|H_n(x)-\mathsf{E}\,H_n(x)|=\mathcal{O}\{(nh)^{-1/2}(\log n)^{1/2}\}.$$

and the following inequality

$$\begin{aligned}\|H_n\| &\leqslant \|H_n-\mathsf{E}\,H_n\|+\|\,\mathsf{E}\,H_n\|.\\ &= \mathcal{O}\{(nh)^{-1/2}(\log n)^{1/2}\}+\wp_p\{(nh)^{-1/2}(\log n)^{-1/2}\}\\ &= \mathcal{O}\{(nh)^{-1/2}(\log n)^{1/2}\}\end{aligned}$$

Statement (A.1.6) thus is obtained.

Statement (A.1.7) follows in the same way as (A.1.6) using assumption $(A2)$ and the Lipschitz continuity properties of K, ψ', l.

According to the uniform consistency of $v_n(x)-v(x)$ shown before, we have

$$\|v_n-v\|=\mathcal{O}_p\{(nh)^{-1/2}(\log n)^{1/2}\}$$

which implies (A.1.9).

Now the assertion of the lemma follows, since by tightness of $D_n(x)$, $\inf_{0\leqslant t\leqslant 1}|D_n(x)|\geqslant q_0$ *a.s.* and thus

$$\|R_n\| = o_p\{(nh\log n)^{-1/2}\}(1+\|f_n\|).$$

Finally, by Theorem 3.1 of Bickel and Rosenblatt (1973), $\|f_n\| = \mathcal{O}_p(1)$; thus the desired result $\|R_n\| = o_p\{(nh\log n)^{-1/2}\}$ follows. ■

We now begin with the subsequent approximations of the processes $Y_{0,n}$ to $Y_{5,n}$.

LEMMA A.1.2

$$\|Y_{0,n} - Y_{1,n}\| = \mathcal{O}\{(nh)^{-1/2}(\log n)^2\} \quad a.s.$$

Proof. Let x be fixed and put $L(y) = \psi\{y - v(x)\}$ still depending on x. Using integration by parts, we obtain

$$\begin{aligned}
&\iint_{\Gamma_n} L(y)K\{(x-t)/h\}dZ_n(t,y) \\
&= \int_{u=-A}^{A}\int_{y=-a_n}^{a_n} L(y)K(u)dZ_n(x-h\cdot u,y) \\
&= -\int_{-A}^{A}\int_{-a_n}^{a_n} Z_n(x-h\cdot u,y)d\{L(y)K(u)\} \\
&\quad + L(a_n)(a_n)\int_{-A}^{A} Z_n(x-h\cdot u,a_n)dK(u) \\
&\quad - L(-a_n)(-a_n)\int_{-A}^{A} Z_n(x-h\cdot u,-a_n)dK(u) \\
&\quad + K(A)\Big\{\int_{-a_n}^{a_n} Z_n(x-h\cdot A,y)dL(y) \\
&\quad + L(a_n)(a_n)Z_{n_a}(x-h\cdot A,a_n) - L(-a_n)(-a_n)Z_n(x-h\cdot A,-a_n)\Big\} \\
&\quad - K(-A)\Big\{\int_{-a_n}^{a_n} Z_n(x+h\cdot A,y)dL(y) + L(a_n)(a_n)Z_n(x+h\cdot A,a_n) \\
&\quad - L(-a_n)(-a_n)Z_n(x+h\cdot A,-a_n)\Big\}.
\end{aligned}$$

If we apply the same operation to $Y_{1,n}$ with $B_n\{T(x,y)\}$ instead of $Z_n(x,y)$ and use Lemma 3.2.2, we finally obtain

$$\sup_{0\leqslant x\leqslant 1} h^{1/2}g(x)^{1/2}|Y_{0,n}(x) - Y_{1,n}(x)| = \mathcal{O}\{n^{-1/2}(\log n)^2\} \quad a.s..$$

■

LEMMA A.1.3 $\|Y_{1,n} - Y_{2,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof. Note that the Jacobian of $T(x,y)$ is $f(x,y)$. Hence

$$\begin{aligned} & Y_{1,n}(x) - Y_{2,n}(x) \\ &= \Big|\{g(x)h\}^{-1/2} \iint_{\Gamma_n} \psi\{y - v(x)\} K\{(x-t)/h\} f(t,y) dt dy\Big| \cdot |W_n(1,1)|. \end{aligned}$$

It follows that

$$\begin{aligned} h^{-1/2}\|Y_{1,n} - Y_{2,n}\| &\leqslant |W_n(1,1)| \cdot \|g^{-1/2}\| \\ &\quad \cdot \sup_{0 \leqslant t \leqslant 1} h^{-1} \iint_{\Gamma_n} |\psi\{y - v(x)\} K\{(x-t)/h\}| f(t,y) dt dy. \end{aligned}$$

Since $\|g^{-1/2}\|$ is bounded by assumption, we have

$$h^{-1/2}\|Y_{1,n} - Y_{2,n}\| \leqslant |W_n(1,1)| \cdot C_4 \cdot h^{-1} \int K\{(x-t)/h\} dx = \mathcal{O}_p(1).$$

■

LEMMA A.1.4 $\|Y_{2,n} - Y_{3,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof. The difference $|Y_{2,n}(x) - Y_{3,n}(x)|$ may be written as

$$\Big|\{g(x)h\}^{-1/2} \iint_{\Gamma_n} [\psi\{y - v(x)\} - \psi\{y - v(t)\}] K\{(x-t)/h\} dW_n\{T(t,y)\}\Big|.$$

If we use the fact that l is uniformly continuous, this is smaller than

$$h^{-1/2}|g(x)|^{-1/2} \cdot \mathcal{O}_p(h)$$

and the lemma thus follows. ■

LEMMA A.1.5 $\|Y_{4,n} - Y_{5,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof.

$$\begin{aligned} |Y_{4,n}(x) - Y_{5,n}(x)| &= h^{-1/2}\Big| \int \Big[\Big\{\frac{g(t)}{g(x)}\Big\}^{1/2} - 1\Big] K\{(x-t)/h\} dW(x)\Big| \\ &\leqslant h^{-1/2}\Big| \int_{-A}^{A} W(x-hu) \frac{\partial}{\partial u}\Big[\Big\{\frac{g(x-hu)}{g(x)}\Big\}^{1/2} - 1\Big] K(u) du\Big| \\ &\quad + h^{-1/2}\Big| K(A) W(t-hA)\Big[\Big\{\frac{g(x-Ah)}{g(x)}\Big\}^{1/2} - 1\Big]\Big| \\ &\quad + h^{-1/2}\Big| K(-A) W(x+hA)\Big[\Big\{\frac{g(x+Ah)}{g(x)}\Big\}^{1/2} - 1\Big]\Big| \\ &\quad S_{1,n}(x) + S_{2,n}(x) + S_{3,n}(x), \quad \textit{say.} \end{aligned}$$

The second term can be estimated by

$$h^{-1/2}\|S_{2,n}\| \leqslant K(A) \cdot \sup_{0\leqslant x\leqslant 1} |W(x-Ah)| \cdot \sup_{0\leqslant x\leqslant 1} h^{-1}\Big|\Big[\Big\{\frac{g(x-Ah)}{g(x)}\Big\}^{1/2} - 1\Big]\Big|;$$

by the mean value theorem it follows that

$$h^{-1/2}\|S_{2,n}\| = \mathcal{O}_p(1).$$

The first term $S_{1,n}$ is estimated as

$$\begin{aligned} h^{-1/2}S_{1,n}(x) &= \Big|h^{-1}\int_{-A}^{A} W(x-uh)K'(u)\Big[\Big\{\frac{g(x-uh)}{g(x)}\Big\}^{1/2} - 1\Big]du \\ &\quad \frac{1}{2}\int_{-A}^{A} W(x-uh)K(u)\Big\{\frac{g(x-uh)}{g(x)}\Big\}^{1/2}\Big\{\frac{g'(x-uh)}{g(x)}\Big\}du\Big| \\ &= |T_{1,n}(x) - T_{2,n}(x)|, \quad say; \end{aligned}$$

$\|T_{2,n}\| \leqslant C_5 \cdot \int_{-A}^{A} |W(t-hu)|du = \mathcal{O}_p(1)$ by assumption on $g(x) = \sigma^2(x) \cdot f_X(x)$. To estimate $T_{1,n}$ we again use the mean value theorem to conclude that

$$\sup_{0\leqslant x\leqslant 1} h^{-1}\Big|\Big\{\frac{g(x-uh)}{g(x)}\Big\}^{1/2} - 1\Big| < C_6 \cdot |u|;$$

hence

$$\|T_{1,n}\| \leqslant C_6 \cdot \sup_{0\leqslant x\leqslant 1}\int_{-A}^{A} |W(x-hu)|K'(u)u/du = \mathcal{O}_p(1).$$

Since $S_{3,n}(x)$ is estimated as $S_{2,n}(x)$, we finally obtain the desired result. ■

The next lemma shows that the truncation introduced through $\{a_n\}$ does not affect the limiting distribution.

LEMMA A.1.6 $\|Y_n - Y_{0,n}\| = \mathcal{O}_p\{(\log n)^{-1/2}\}$.

Proof. We shall only show that $g'(x)^{-1/2}h^{-1/2}\iint_{\mathbb{R}-\Gamma_n}\psi\{y-v(x)\}K\{(x-t)/h\}dZ_n(t,y)$ fulfills the lemma. The replacement of $g'(x)$ by $g(x)$ may be proved as in Lemma A.4 of Johnston (1982). The quantity above is less than $h^{-1/2}\|g^{-1/2}\|\cdot\|\iint_{\{|y|>a_n\}}\psi\{y-v(x)\}K\{(x-t)/h\}dZ(t,y)\|$. It remains to be shown that the last factor tends to zero at a rate $\mathcal{O}_p\{(\log n)^{-1/2}\}$. We show first that

$$\begin{aligned} V_n(x) &= (\log n)^{1/2}h^{-1/2}\iint_{\{|y|>a_n\}}\psi\{y-v(x)\}K\{(x-t)/h\}dZ_n(t,y) \\ &\xrightarrow{p} 0 \quad \text{for all} \quad x \end{aligned}$$

and then we show tightness of $V_n(x)$, the result then follows:

$$\begin{aligned} V_n(x) &= (\log n)^{1/2}(nh)^{-1/2}\sum_{i=1}^{n}[\psi\{Y_i - v(x)\}\mathbf{I}(|Y_i| > a_n)K\{(x - X_i)/h\} \\ &\quad - \mathsf{E}\,\psi\{Y_i - v(x)\}\mathbf{I}(|Y_i| > a_n)K\{(x - X_i)/h\}] \\ &= \sum_{i=1}^{n} X_{n,x}(x), \end{aligned}$$

where $\{X_{n,x}(x)\}_{i=1}^{n}$ are i.i.d. for each n with $\mathsf{E}\,X_{n,x}(x) = 0$ for all $x \in [0,1]$. We then have

$$\begin{aligned} \mathsf{E}\,X_{n,x}^2(x) &\leqslant (\log n)(nh)^{-1}\,\mathsf{E}\,\psi^2\{Y_i - v(x)\}\mathbf{I}(|Y_i| > a_n)K^2\{(x - X_i)/h\} \\ &\leqslant \sup_{-A\leqslant u\leqslant A} K^2(u)\cdot(\log n)(nh)^{-1}\,\mathsf{E}\,\psi^2\{Y_i - v(x)\}\mathbf{I}(|Y_i| > a_n); \end{aligned}$$

hence

$$\begin{aligned} \mathrm{Var}\{V_n(x)\} &= \mathsf{E}\Big\{\sum_{i=1}^{n} X_{n,x}(x)\Big\}^2 = n\cdot\mathsf{E}\,X_{n,x}^2(x) \\ &\leqslant \sup_{-A\leqslant u\leqslant A} K^2(u)h^{-1}(\log n)\int_{\{|y|>a_n\}} f_y(y)dy\cdot M_\psi. \end{aligned}$$

where M_ψ denotes an upper bound for ψ^2. This term tends to zero by assumption $(A3)$. Thus by Markov's inequality we conclude that

$$V_n(x) \xrightarrow{p} 0 \quad \text{for all } x \in [0,1].$$

To prove tightness of $\{V_n(x)\}$ we refer again to the following moment condition as stated in Lemma A.1.1:

$$\mathsf{E}\{|V_n(x) - V_n(x_1)|\cdot|V_n(x_2) - V_n(x)|\} \leqslant C'\cdot(x_2 - x_1)^2$$

$$C' \text{ denoting a constant}, \quad x \in [x_1, x_2].$$

We again estimate the left-hand side by Schwarz's inequality and estimate each factor separately,

$$\begin{aligned} \mathsf{E}\{V_n(x) - V_n(x_1)\}^2 &= (\log n)(nh)^{-1}\,\mathsf{E}\Big[\sum_{i=1}^{n}\Psi_n(x, x_1, X_i, Y_i)\cdot\mathbf{I}(|Y_i| > a_n) \\ &\quad - \mathsf{E}\{\Psi_n(x, x_1, X_i, Y_i)\cdot\mathbf{1}(|Y_i| > a_n)\}\Big]^2, \end{aligned}$$

where $\Psi_n(x, x_1, X_i, Y_i) = \psi\{Y_i - v(x)\}K\{(x - X_i)/h\} - \psi\{Y_i - v(x_1)\}K\{(x_1 - X_1)/h\}$. Since ψ, K are Lipschitz continuous except at one point and the expectation is taken afterwards, it follows that

$$\begin{aligned}
&[\mathsf{E}\{V_n(x) - V_n(x_1)\}^2]^{1/2} \\
&\leqslant \; C_7 \cdot (\log n)^{1/2} h^{-3/2} |x - x_1| \cdot \Big\{ \int_{\{|y|>a_n\}} f_y(y)dy \Big\}^{1/2}.
\end{aligned}$$

If we apply the same estimation to $V_n(x_2) - V_n(x_1)$ we finally have

$$\begin{aligned}
&\mathsf{E}\{|V_n(x) - V_n(x_1)| \cdot |V_n(x_2) - V_n(x)|\} \\
&\leqslant \; C_7^2 (\log n) h^{-3} |x - x_1||x_2 - x| \times \int_{\{|y|>a_n\}} f_y(y)dy \\
&\leqslant \; C' \cdot |x_2 - x_1|^2 \text{ since } x \in [x_1, x_2] \quad \text{by (A3).}
\end{aligned}$$

■

LEMMA A.1.7 *Let $\lambda(K) = \int K^2(u)du$ and let $\{d_n\}$ be as in the theorem. Then*

$$(2\delta \log n)^{1/2}[\|Y_{3,n}\|/\{\lambda(K)\}^{1/2} - d_n]$$

has the same asymptotic distribution as

$$(2\delta \log n)^{1/2}[\|Y_{4,n}\|/\{\lambda(K)\}^{1/2} - d_n].$$

Proof. $Y_{3,n}(x)$ is a Gaussian process with

$$\mathsf{E}\, Y_{3,n}(x) = 0$$

and covariance function

$$\begin{aligned}
r_3(x_1, x_2) &= \mathsf{E}\, Y_{3,n}(x_1) Y_{3,n}(x_2) \\
&= \{g(x_1)g(x_2)\}^{-1/2} h^{-1} \iint_{\Gamma_n} \psi^2\{y - v(x)\}K\{(x_1 - x)/h\} \\
&\quad \times K\{(x_2 - x)/h\} f(t, y)dtdy \\
&= \{g(x_1)g(x_2)\}^{-1/2} h^{-1} \iint_{\Gamma_n} \psi^2\{y - v(x)\} f(y|x)dy K\{(x_1 - x)/h\} \\
&\quad \times K\{(x_2 - x)/h\} f_X(x)dx \\
&= \{g(x_1)g(x_2)\}^{-1/2} h^{-1} \int g(x) K\{(x_1 - x)/h\} K\{(x_2 - x)/h\} dx \\
&= r_4(x_1, x_2)
\end{aligned}$$

where $r_4(x_1, x_2)$ is the covariance function of the Gaussian process $Y_{4,n}(x)$, which proves the lemma. ■

A.2 Algorithm in Chapter 4

A.2.1 Identification

Firstly, to make sure there exists global minima, that is to say, (4.18) is convex, one needs a constraint on α_i. $\sum_{i=1}^{N} \alpha_i = 0$ guarantee that the convexity of the penalized loss function S^*. To identify Θ_f and α's, note that $\Theta_f \alpha_i = \tilde{\Theta}_f \tilde{\alpha}_i$, where $\tilde{\Theta}_f = \Theta_f C$ and $\tilde{\alpha}_i = C^{-1}\alpha_i$ for any invertible $K \times K$ matrix C. Therefore, by requiring that D_α be diagonal and that the Θ_f have orthonormal columns, we prevent reparameterization by linear transformation, for more clarification of this issue we refer to Lemma 1 in Zhou et al. (2008). We also need to order α_i according to their second moments, since α_i's are already centered, that is $\mathsf{E}(\alpha_{i1}^2) > \cdots > \mathsf{E}(\alpha_{iK}^2)$, which makes sure all the equations are identifiable.

A.2.2 The complete PLAWS Algorithm

We give the complete algorithm in this appendix. The parameters that appear on the right hand side of the equations are all fixed at the values from the last iteration.

1. Initialization the algorithm using the procedure described in Appendix A.2.3.
2. Update $\widehat{\theta}_\mu$.

$$\widehat{\theta}_\mu = \left\{ \sum_{i=1}^{N} B_i^\top \widehat{W}_i B_i + \lambda_\mu \Omega \right\}^{-1} \left\{ \sum_{i=1}^{N} B_i^\top \widehat{W}_i (Y_i - B_i \widehat{\Theta}_f \widehat{\alpha}_i) \right\}$$

3. For $l = 1, \cdots, K$, update the l-th column of $\widehat{\Theta}_f$ using

$$\widehat{\theta}_{f,l} = \left\{ \sum_{i=1}^{N} \widehat{\alpha}_{il}^2 B_i^\top \widehat{W}_i B_i + \lambda_f \Omega \right\}^{-1} \left\{ \sum_{i=1}^{N} \widehat{\alpha}_{il} B_i^\top \widehat{W}_i (Y_i - B_i \widehat{\theta}_\mu - B_i Q_{il}) \right\}$$

   where

$$Q_{il} = \sum_{k \neq l} \widehat{\theta}_{f,k} \widehat{\alpha}_{ik}, \qquad i = 1, \cdots, N$$

   $\widehat{\theta}_{f,k}$ is the k-th column of $\widehat{\Theta}_f$.

4. Orthogonalize Θ_f.

5. Update $(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_N)$ using

$$\widehat{\alpha}_i = (\widehat{\Theta}_f^\top B_i^\top \widehat{W}_i B_i \widehat{\Theta}_f)^{-1} \left\{ \widehat{\Theta}_f^\top B_i^\top \widehat{W}_i (Y s_i - B_i \widehat{\theta}_\mu) \right.$$

6. Update the weight, defined in (4.21) for expectile and (4.22) for quantile.

7. Iterate Steps 2-6 until Convergence is reached.

A.2.3 Initial Values Selection

We propose the following procedure to initialize the PLAWS algorithm:

1. We estimate N single expectile/quantile curves $\widehat{l}_i(t)$ by applying the single curve estimation algorithm described in Section 4.2.

2. Set $\widehat{L}_i = \{\widehat{l}(t_{i1}), \ldots, \widehat{l}(t_{iT_i})\}^\top$. Run the linear regression

$$\widehat{L}_i = B_i \theta_\mu + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{A.2.1}$$

to get the initials of $\widehat{\theta}_\mu$ as follows

$$\widehat{\theta}_{\mu 0} = \left(\sum_{i=1}^{N} B_i^\top B_i \right)^{-1} \left(\sum_{i=1}^{N} B_i^\top \widehat{L}_i \right).$$

3. Calculate the residuals of the regression (A.2.1), denoted as $\widetilde{L}_i = \widehat{L}_i - B_i \widehat{\theta}_{\mu 0}$. For each i, run the following linear regression:

$$\widetilde{L}_i = B_i \Gamma_i + \varepsilon_i,$$

the solution, denoted as $\widehat{\Gamma}_{i0}$, is used in later steps for finding initials of Θ_f and α_i. Set $\widehat{\Gamma}_0 = (\widehat{\Gamma}_{10}, \cdots, \widehat{\Gamma}_{N0})$.

4. Calculate the singular value decomposition of $\widehat{\Gamma}_0^\top$:

$$\widehat{\Gamma}_0^\top = U D V^\top$$

The initial of Θ_f is chosen as $\widehat{\Theta}_f = V_k D_k$ where V_k consists of the first K columns of V and D_k is the corresponding $K \times K$ block of D.

5. Do regression on $\widehat{\Theta}_f$

$$\widehat{\Gamma}_{i0} = \widehat{\Theta}_f(\alpha_{i1}, \cdots, \alpha_{iK}) + \varepsilon_i \tag{A.2.2}$$

to get the initials of $\widehat{\alpha}_i$. We centered the initial values α_i's for further iterating procedures.

Selbständigkeitserklärung

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Berlin, 15 Mai 2012

Mengmeng Guo