

UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL SCIENCES

Mathematics Science

**Nonlinear Spatio-temporal Modelling Theory with
Application**

by

Xiaohang Ren

ORCID ID 0000-0002-9097-580X

Doctor of Philosophy

September 2020

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

Mathematics Science

Doctor of Philosophy

by Xiaohang Ren

A large number of spatio-temporal data with sophisticated structures exist in all kinds of disciplines such as energy research, environmental sciences, urban economics, etc. Nonlinear modelling of such spatio-temporal data is often a challenge, with irregularly observed locations and location-wide non-stationarity (c.f., [Lu et al., 2009](#); [Al-Sulami et al., 2017a](#)). This thesis sheds new light on both theoretical investigation into the spatio-temporal model with varying coefficient structure and empirical application to the complex dynamic nature of spatio-temporal data in energy market.

The main contributions of the thesis are as following:

(1) Chapter 2 empirically investigates the dynamic integrations or linkages of the USA's natural gas markets at the state level in a holistic manner by proposing a spatio-temporal network quantile analysis. This chapter empirically finds that spatial neighbouring effects significantly exist among the natural gas markets, not only in the eastern and middle states but also in some western and southwest states, and so do the dynamic linkages to the national crude oil market for the natural gas markets in the southern and eastern states, which are heterogeneous at different quantile levels. The empirical findings can help investors to hedge against the energy price risks in maximizing profits and local industry and government to mitigate the negative impacts from the expected or unexpected fluctuations in the national oil and the neighbouring natural gas markets.

(2) In Chapter 3, we are mainly concerned with the nonlinear modelling of econometric time series data from the view point of quantile regression. Specifically, we consider quantile regression for a (strictly) stationary time series that is near epoch dependent (NED). It is an extension of the work of [Welsh \(1996\)](#) and [Yu and Jones \(1998\)](#) under *i.i.d.* samples to econometric time series. Our asymptotic normality result generalizes [Lu et al. \(1998\)](#), [Honda \(2000\)](#) and [Hallin et al. \(2009a\)](#) who obtained Bahadur representations with different conditions under α -mixing

dependence to a more widely applicable family of data generating processes of near epoch dependence. It also lays the foundation for establishing asymptotic properties of nonlinear quantile modelling for spatio-temporal data under NED.

(3) In Chapter 4, we propose a semiparametric family of *Dynamic Functional-coefficient Autoregressive Spatio-Temporal (DyFAST)* models to address the modelling and analysis of spatio-temporal data, $Y_t(s_i)$, $t = 1, \dots, T$ and $i = 1, 2, \dots, N$, featuring non-stationarity over irregular locations in space. It is a useful extension and combination of spatial autoregressive models (Ord, 1975; Gao et al., 2006; Kelejian and Prucha, 2010; Su and Jin, 2010) and varying coefficient models (Zhang and Fan, 1999; Wang and Xia, 2009; Zhang et al., 2009) as well as Al-Sulami et al. (2017a, 2019). To model the dynamic spatial neighbouring temporal-lagged effects with the irregular locations, we consider using spatial weight matrix pre-specified either by experts or by the prior information of spatial locations, which is popular in spatial econometrics. Moreover, both one-step and two-step estimation methods are proposed to estimate the functional coefficients in DyFAST models. Accordingly, different semiparametric smoothing estimation procedures are suggested. Both theoretical properties and Monte Carlo simulations are investigated. Our empirical applications to energy market data sets further illustrate the usefulness of the models.

(4) In Chapter 5, we propose a semiparametric spatio-temporal quantile model to further study the dynamic nature of non-normal spatio-temporal data. The quantile estimators are more robust than the mean estimators considered in Chapter 4 for non-normal spatio-temporal data. Moreover, this approach can provide more information about the dynamic spatial effects at different quantile levels. Both theoretical asymptotic properties and Monte Carlo simulations for our model are established.

Contents

Declaration of Authorship	xi
Acknowledgements	xiii
1 Introduction	1
1.1 Research Context and Background	1
1.2 Literature Review	3
1.2.1 Quantile regression	3
1.2.2 Spatio-temporal model	5
1.2.3 Spatial weight matrix	7
1.2.3.1 Binary-based weight matrix	7
1.2.3.2 Distance-based weight matrix	8
1.3 Structure of the Thesis	10
2 On Dynamic Linkages of the State Natural Gas Markets in the USA: Evidence from an Empirical Spatio-temporal Network Quantile Analysis	11
2.1 Background	11
2.2 Literature Review	12
2.3 Data and Methodology	15
2.3.1 Data	15
2.3.2 Methodology	18
2.4 Empirical Findings	20
2.4.1 Global spatial autocorrelation analysis	20
2.4.2 Empirical results of spatio-temporal network quantile econometric modelling	23
2.4.3 Discussions	28
2.5 Conclusions	32
2.6 Appendix	36
3 Local Linear Quantile Regression for Time Series under Near Epoch Dependence	43
3.1 Introduction	43
3.2 Local Linear Quantile Regression Estimators	45
3.2.1 Notation and main assumption	45

3.2.2	Local linear fitting of quantile regression function	48
3.3	Asymptotic Behaviours	49
3.3.1	Bahadur representation	49
3.3.2	Asymptotic normality	50
3.4	Numerical Results	53
3.4.1	Simulation	53
3.4.2	An empirical application	56
3.5	Conclusions	61
3.6	Appendix	62
4	On Semiparametric Smoothing of DyFAST for Location-wide Non-stationary Spatio-Temporal Data	83
4.1	Introduction	83
4.2	Dynamical Functional-coefficient Autoregressive Spatio-Temporal (DyFAST) Models	85
4.3	Estimation	87
4.3.1	One-step estimation method	88
4.3.2	Asymptotic property	88
4.3.3	Bandwidth and Order Selection	92
4.3.4	Discussion on comparison with a two-step procedure	93
4.4	Discussion	96
4.5	Simulation Study	98
4.6	Real Data Example	103
4.6.1	Background	103
4.6.2	Data	104
4.6.3	Result	105
4.7	Conclusions	112
4.8	Appendix	113
5	Semiparametric Dynamic Varying Coefficients Spatio-temporal Quantile Regression	135
5.1	Introduction	135
5.2	Modelling Procedure	136
5.3	Asymptotic Theory	138
5.4	Simulation Study	141
5.5	Real Data Example	145
5.6	Conclusions	150
5.7	Appendix	151
6	Conclusion and Outlook	161
	References	165

List of Figures

1.1	Contiguity criteria in a regular square lattice grid	8
2.1	The time-series plots of the monthly WTI price data	15
2.2	Spatial Distributions of Natural Gas Prices: Mean and Standard Deviation	21
2.3	Scatter plot of spatial neighbouring effects of 48 states in the U.S. .	23
2.4	Maps of influence coefficients of spatial neighbouring effects.	25
2.5	Maps of influence coefficients of autoregressive effects.	26
2.6	Maps of influence coefficients of oil price effects.	27
2.7	Maps of the effects of the long-run relationship of the WTI crude oil and Henry Hub natural gas prices on the state-level natural gas price returns.	28
2.8	Region to region capacity map of natural gas in the U.S. in 2016 . .	32
2.9	The effects of driving forces on natural gas price returns in six typical states.	33
2.10	The effects of driving forces on natural gas price returns in other typical states.	34
3.1	Comparison of mean of 100 median and mean estimates of $g(\cdot)$ for sample size 100 and 500	55
3.2	Boxplots of the local linear fitting for the median regression and mean regression	56
3.3	Boxplots of squared estimation error of mean regression and median regression estimation	57
3.4	WTI crude oil and carbon future daily prices	58
3.5	Nonparametric quantile regression for the relationship between daily return of WTI crude oil price and carbon future price, from 27th July 2015 to 15th May 2017 with 454 observations.	59
4.1	Boxplots of 100 times SEE for the estimation of $\beta_j(\cdot)$ I	99
4.2	Boxplots of 100 times SEE for the estimation of $\beta_j(\cdot)$ II	100
4.3	Boxplots of 100 times SEE for the estimation of combining weight .	100
4.4	Simulation results: the mean in 100 replications	101
4.5	Boxplots of 100 times SEE for the estimation of $\beta_j(\cdot)$ III	102
4.6	Boxplots of 100 times SEE for the estimation of combining weight II	102
4.7	The impacts of Brent crude oil price return on diesel price returns in 23 EU countries by DyFAST with W_1 and W_2	107

4.8	The impacts of Brent crude oil price return on diesel price returns in 23 EU countries by DyFAST with combined weight matrix. . . .	107
4.9	The impacts of Brent crude oil price return on diesel price returns in 23 EU countries by DyFAST with W_2	109
4.10	Three EU country groups based on the results of DyFAST with W_2	109
5.1	Boxplots of 100 times mean squared estimation error for the esti- mation of $\beta_j(\cdot)$	143
5.2	Boxplots of 100 times estimation of conditional median quantile estimation	143
5.3	Simulation results: the mean of 100 replications at location $i = 1$. .	144
5.4	The results of semiparametric varying coefficients spatio-temporal quantile regression	148
5.5	The results of DyFAST	149

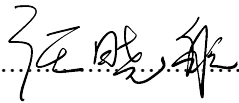
List of Tables

2.1	The test statistics for monthly natural gas prices of 48 states and Henry hub	16
2.2	The Moran's I test result	22
2.3	The AIC values for different pairs of orders for the 48 states	24
2.4	The estimation result for cointegration test	25
2.5	The estimation results for the 0.1-level quantile	37
2.6	The estimation results for the 0.25-level quantile	38
2.7	The estimation results for the 0.5-level quantile	39
2.8	The estimation results for the 0.75-level quantile	40
2.9	The estimation results for the 0.9-level quantile	41
2.10	The Wald tests for the equality of slopes	42
3.1	Estimation results of ARMA-GARCH	58
3.2	Results of quantile prediction error for three methods	59
4.1	Descriptive statistics for data.	104
4.2	Akaike Information Criterion with correction (AICc)	105
4.3	Country Groups	110
4.4	Comparison of forecasting results I	111
4.5	Comparison of forecasting results II	111
5.1	Normality Tests for data.	146
5.2	Comparison of prediction results	150

Declaration of Authorship

I, Xiaohang Ren , declare that the thesis entitled *Nonlinear Spatio-temporal Modelling Theory with Application* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: Xiaohang Ren, Zudi Lu, Cheng Cheng, Yukun Shi, Jian Shen. On dynamic linkages of the state natural gas markets in the USA: Evidence from an empirical spatio-temporal network quantile analysis. *Energy Economics*, 2019, 80: 234-252.[Chapter 2]
- One academic paper is available on-line in the SSRN eLibrary: Ren, Xiaohang and Lu, Zudi, Local Linear Quantile Regression for Time Series under near epoch dependence. Available at SSRN: <https://ssrn.com/abstract=3555740>. [Chapter 3]

Signed:.....

Date:.....29/9/2020

Acknowledgements

Throughout my 4 year PhD study at Southampton, I have to say, without the help, guidance, support and encouragement of my supervisors, Prof. Zudi Lu and Prof. Wei Liu, it would have been impossible for me to fulfil this task. I would like to express my heartfelt thanks to Zudi, who guides me to the spatio-temporal and semi- and non-parametric world. We always have long meetings which inspire me with new research ideas. I am also grateful to Prof. Wei for providing comments on my researches, and for the help and guidance during my time as a PhD student.

Doing a PhD is a lengthy and lonely journey. I would like to thank all my friends in the UK and China, who have accompanied and supported me during the past four years. My life would be significantly different without the following professors, colleagues and friends: Dr.Cheng Cheng, Dr.Kun Duan, Dr.MingMing Liu, Dr.Yukun Shi, Dr.Cheng Yan, Dr.Dan Yan, Prof.Tapas Mishra, Prof.Zhen Wang, Prof. Rongmao Zhang, Lizhu Tao, Jian Shen and many significant others. In addition, I would like to thank all the staff of the School of Mathematics Science for their help and support.

Finally, I owe my deepest gratitude to my beloved parents, Zhenpeng Ren and Lanting Wu, and my little brother, Hengyi Ren. Their love and encouragement have provided the strong support that has enabled me to finish this thesis.

Chapter 1

Introduction

1.1 Research Context and Background

A large number of spatio-temporal data with sophisticated structures exist in all kinds of disciplines such as energy research, environmental sciences, urban economics, etc. To measure spatial neighbouring effects and horizontal dimension of transmission effect, spatio-temporal models have been widely used in a variety of fields ([Fingleton, 2008](#); [Holly et al., 2010](#); [Auffhammer and Carson, 2008](#); [Yu, 2012](#)). In this project, the main objective is to develop statistical methods for temporal and spatial regression modellings with application to regional energy data analysis.

In the large majority of the literature on energy issues, the analyses typically do not take into account cross-market impacts and neglect spatial neighbouring effects. However, the dis-equilibria of prices in neighbouring regions or countries can affect the local energy price. For example, it is well-known that cross-country purchases of petroleum products is a common phenomenon in EU because of the differences in energy taxation ([Wlazlowski et al., 2009](#)). Due to this phenomenon, the local price is strongly related to the petroleum prices in nearby locations.

From a methodological perspective, the existing research has three main drawbacks. First, most literature considers the parametric models, such as time series models, cointegration models and vector autoregressive (VAR) models, etc. A drawback with parametric models is the possible bias of the specified model from the underlying relationship ([Fan and Yao, 2003](#)). As the relationship between energy markets has become increasingly complex, the problem that has arisen is that

parametric models may be biased and cannot uncover the underlying relationship well.

Second, most studies typically do not take into account price shock propagation and cross-market impacts. They neglected the dis-equilibria in neighbouring regions or countries that can affect the home region or country's energy markets. In other words, these studies do not consider the horizontal (multinational) dimension of transmission effect. For example, technological breakthroughs and policy changes (e.g., deregulation) have changed integrated monopoly American natural gas market into separate regional markets in recent years ([Park et al., 2008](#)). However, due to the significant differences in resources volumes and energy consumptions among different states in the USA, a well-developed pipeline system has been built to connect the natural gas markets of each state ([Cuddington and Wang, 2006](#)). In addition, the law of one price, holding in regional natural gas markets, contributes to the price interactions among natural gas spot markets ([Park et al., 2008](#)). Therefore, the natural gas market in a state may be strongly related to the natural gas markets in nearby states. These changes invite the research question addressed here: what is the relationship among regional natural gas markets in the current era, dependent or segmented? How does the national crude oil market impact the regional natural gas markets? Although there is a body of empirical literature analysing the gasoline price dispersion using gas station level data ([Clemen and Gugler, 2006](#); [Yilmazkuday and Yilmazkuday, 2016](#)), the study of spatial linkages in regional natural gas markets has essentially been overlooked.

Third, most of spatial models with energy data assume normality for variables. However, some energy price data may be non-Gaussian and have obvious peak and fat tails ([Hung et al., 2008](#)). Correctly estimating the tail behaviour of distributions is of crucial importance in the energy risk management and has policy implications. Quantile approach, therefore, can offer more information of oil or natural gas prices in specific market circumstances, such as the lower, upper and intermediate quantile markets. These are interesting for policymakers to know what will happen at the extreme energy prices.

In this research, we will, therefore, be concerned with statistical inference of spatio-temporal data in energy markets, developing novel semiparametric spatio-temporal mean and quantile models, and applying the theoretical results for empirically modelling real spatio-temporal data.

1.2 Literature Review

1.2.1 Quantile regression

The classical least squares regression has been one of the most widely used statistical methods in data analysis for many years. The reason is that the interpretation of least squares method is consistent with the intuitive imagination of people, and it is easy to calculate. In particular, if the error is normal distribution, the results of least squares method are unbiased and efficient. However, in the risk management, people often focus on the tail behaviour of distributions, and the financial or energy prices data do not comply with the normal distribution and have obvious peaks and fat tails, as well as the abnormal points. This will lead to the biased estimation results by the classic regression. Therefore, quantile regression method (Koenker and Bassett Jr, 1978) has been employed to estimate these data in many literature.

The τ -th quantile of a random variable Y is given by

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf \{y : F_Y(y) \geq \tau\}, \quad (1.2.1)$$

where $F_Y(y) = P(Y \leq y)$ and $\tau \in (0, 1)$. The quantile regression method provides a lot of advantages compared to the ordinary least squares estimation method. Firstly, a well known special case with 50% quantiles, i.e., median regression, is much explored and more robust than the mean regression when the distribution of variable is typically skewed or possesses a few abnormal observations (Bera et al., 2016). Secondly, a heteroscedastic model can be easily detected if the results of quantile regression at different quantile levels are not parallel (Efron, 1991). Thirdly, there do not need to be distributional assumptions (Sherwood and Wang, 2016). Fourthly, a collection of conditional quantiles can describe the whole distribution of the independent variable (Yu and Jones, 1998). Finally, pairs of extreme conditional quantiles can be used to depict the conditional prediction interval, which is important in econometric forecasting (see, Granger et al., 1989; Koenker and Zhao, 1996; Kuester et al., 2006, for example). A detailed review about quantile regression can be found in Yu et al. (2003).

There has been a considerable interest in modelling and estimation of time series data in energy markets. For instance, Ding et al. (2016) found that crude oil and stock index has a significant causal relationship in tail quantile intervals by using

the quantile causality test. The Granger quantile causality (Chuang et al., 2009) represents as:

$$Q_{Y_t}(\tau | (\mathcal{Y}, \mathcal{X})_{t-1}) = Q_{Y_t}(\tau | \mathcal{Y}_{t-1}), \quad \forall \tau \in (0, 1), \quad (1.2.2)$$

where $Q_{Y_t}(\tau | \mathcal{F})$ denotes the τ -th quantile of $F_{Y_t}(\tau | \mathcal{F})$, and set $(\mathcal{Y}, \mathcal{X})_{t-1}$ contains the information of Y_i and X_i up to time $t - 1$. If above equation holds, it means that X_t is not the Granger-cause of Y_t over the quantile interval $(0, 1)$.

Most of the literature in energy market has largely focused on conditional mean regression models. However, as the relationships between energy markets and financial markets have become somewhat increasingly complex at different market situations, the conditional mean regression may not capture this complex relationship well. These rise people's interests to quantile regression estimates, which are more robust and could discover more useful relationships between variables with weak or no relationship between the means of these variates (Sim and Zhou, 2015; Reboredo and Ugolini, 2016). The coefficients of τ -th quantile of the conditional distribution are estimated as:

$$\hat{\beta}(\tau) = \arg \min_{\beta} \sum_{i=1}^T \rho_{\tau}(Y_i - x_i' \beta(\tau) - \alpha(\tau)), \quad (1.2.3)$$

where $\rho_{\tau}(y) = y(\tau - I_{\{y < 0\}})$ with $y \in R^1$ is the traditional check function, and I_A is the indicator function of set A .

Meanwhile, due to the inevitable occasions when parametric approaches fail to estimate the unknown functional form, an extensive literature estimates conditional quantile functions by nonparametric approaches, such as spline approach (Koenker et al., 1994; Bosch et al., 1995) and kernel viewpoint (Koenker et al., 1994; Yu and Lu, 2004; Zhou et al., 2009; Hallin et al., 2009b). For example, Koenker et al. (1994) investigated a class of quantile smoothing splines through L_p penalties. Yu and Jones (1998) used local linear approach to investigate the nonparametric regression quantile regression. And Hallin et al. (2009b) considered quantile regression in spatial dependence structure and obtain a Bahadur representation for a local linear estimator $q_{\tau}(\cdot)$, which is the τ -th quantile. In this thesis, we will also focus on local linear fitting approach. The idea of local linear fitting (Fan and Gijbels, 1996) is to approximate the unknown function $q_{\tau}(\cdot)$ by a linear function

$$q_{\tau}(z) \approx q_{\tau}(x) + (\dot{q}_{\tau}(x))(z - x) \equiv a_0 + a_1'(z - x), \quad (1.2.4)$$

for z in a neighbourhood of x and $q_\tau(z)$ is the τ -th quantile of z . Then

$$(\hat{a}_0, \hat{a}_1) = \arg \min_{(a_0, a_1)} \sum_{i=1}^n \rho_\tau(Y_i - a_0 - a'_1(X_i - x))K_h(X_i - x), \quad (1.2.5)$$

where $K_h(x) = h_n^{-p}K(x/h_n)$, K is a kernel function on R^p , and $h_n > 0$ is the bandwidth.

1.2.2 Spatio-temporal model

There are a large number of spatio-temporal data with complex structures in all kinds of disciplines such as energy research, environmental sciences, urban economics, etc. To measure spatial neighbouring effects and horizontal (multinational) dimension of transmission effect, spatio-temporal models have been widely used in these fields, such as urban economics field (Fingleton, 2008; Holly et al., 2010) and environmental sciences (Auffhammer and Carson, 2008; Yu, 2012).

There are two basic models to estimate the spatial dependence. The first one is spatial autoregressive (SAR) model.

$$\mathbf{y} = \rho W \mathbf{y} + \varepsilon, \quad (1.2.6)$$

where W is a redefined $n \times n$ spatial weighting matrix and ρ is a spatial autoregressive parameter. This model shows that there is an interaction effect between the value of a variable at a given location and the values of the same variable at neighbour locations. W is almost always row-standardized, which could take into account a weighted average of neighbouring values.

Another one is conditional autoregressive (CAR) model proposed by Besag (1974).

$$\mathbf{E}(y_i | \text{all } y_{j \neq i}) = \mu_i + \rho \sum_{j \neq i} w_{ij}(y_j - \mu_j), \quad (1.2.7)$$

where μ_i is the expected value at i , and the spatial autocorrelation parameter ρ represents the size of spatial neighbouring effect. w_{ij} is a weight that describe the impact of mean adjusted values at locations j on i .

These models consider the linear spatial relationship, which may be violated in real spatial data analysis (Hallin et al., 2004; Lu et al., 2009; Cressie and Wikle, 2015). Thus, to better describe the nonlinear interactions in time and nonstationarity

across space, efficient nonlinear spatio-temporal models have been constructed in some literature. A general form of a nonlinear spatio-temporal model considering first-order lag can be written as

$$\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}, \boldsymbol{\epsilon}_t; \boldsymbol{\theta}); \quad t = 1, 2, \dots, T, \quad (1.2.8)$$

where for the spatial locations $\{s_1, s_2, \dots, s_N\}$, $\mathbf{Y}_t = (Y_t(s_1), Y_t(s_2), \dots, Y_t(s_N))'$ is a vector of spatio-temporal variables, $\mathcal{M}(\cdot, \cdot; \cdot)$ is a nonlinear function, such as local linear approximation and state-dependent models, $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \Sigma_t)$ and $\boldsymbol{\theta}$ is a vector of process parameters (Cressie and Wikle, 2015).

Some researchers use science-based models, such as partial differential equation (PDE) and integro-difference equation (IDE), to parameterize the nonlinearity in spatio-temporal models (Hooten and Wikle, 2008; Wikle and Holan, 2011). However, these models may not be appropriate for social and economic research, due to the lack of prior information and the empirical relationship among variables.

In contrast, nonparametric approaches can more effectively help to investigate these complex nonlinear relationships. However, due to the common existence of “curse of dimensionality” in spatial and spatio-temporal modelling, using pure nonparametric approach may get a poor estimation. Therefore, various semiparametric approaches have been suggested to address this problem for spatial data (Gao et al., 2006; Lu et al., 2007; Lu and Tjøstheim, 2014).

To better capture the nonlinearity and non-stationarity of spatio-temporal data, Lu et al. (2009) constructed an adaptive varying-coefficient spatio-temporal model.

$$Y_t(\mathbf{s}) = a[\mathbf{s}, \boldsymbol{\alpha}(\mathbf{s})^T \mathbf{X}_t(\mathbf{s})] + \mathbf{b}_1[\mathbf{s}, \boldsymbol{\alpha}(\mathbf{s})^T \mathbf{X}_t(\mathbf{s})]^T \mathbf{X}_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \quad (1.2.9)$$

where $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^2$ represents the spatial location. In addition, $a(\mathbf{s}, \cdot)$ is an unknown scalar, $\mathbf{b}_1(\mathbf{s}, \cdot)$ is $d \times 1$ functions and $\boldsymbol{\alpha}(\mathbf{s})$ is an unknown index vector. $\mathbf{X}_t(\mathbf{s})$ is a vector of time-lagged values of $Y_t(\cdot)$ in a neighbourhood of \mathbf{s} and possibly some exogenous variables. The noise process $\varepsilon_t(\mathbf{s})$ is a sequence of *i.i.d.* random variables over time. This model is widely applicable in nonlinear spatio-temporal dependence structure, but the computation is involved.

Based on Lu et al. (2009) and Gao et al. (2006), Al-Sulami et al. (2017a) proposed a spatio-temporal autoregressive partially non-linear regression model in the form

of

$$Y_t(s_j) = g(X_t(s_j), s_j) + \sum_{i=1}^p \lambda_i(s_j) Y_{t-i}^{sl}(s_j) + \sum_{l=1}^q \alpha_l(s_j) Y_{t-l}(s_j) + \varepsilon_t(s_j) \quad (1.2.10)$$

$$Y_t^{sl}(s_j) = \sum_{k=1}^N W_{jk} Y_t(s_k), \quad j = 1, \dots, N, t = 1, \dots, T$$

where a nonparametric function $g(X_t(s_j), s_j)$ varies by location. The temporally lagged variables, $Y_{t-i}^{sl}(s_j)$ denoting spatial neighbouring effects of region s_j . However, this model appears too restrictive to estimate the non-normal data and ignores the possible effects of exogenous variables on coefficients λ_i and α_l .

1.2.3 Spatial weight matrix

The spatial weight matrix reflects the spatial relations between different spatial locations. Mostly, the spatial weight matrix is represented by a $N \times N$ matrix W , and the element w_{ij} , for $1 \leq i, j \leq N$, measures the degree of spatial neighbouring effect of location s_i on s_j . There are different types of spatial weight matrix in the literature ([LeSage and Pace, 2009](#); [Arbia, 2014](#); [Cressie and Wikle, 2015](#)) and we will review some common types in this section.

1.2.3.1 Binary-based weight matrix

The data are supposed to be observed on a regular square lattice grid. In this case, two spatial locations/units share one boundary, which reflects spatial neighbouring influence.

Rooks contiguity

Let $\mathcal{A}(s_i)$ be a set of units that share a same side with unit s_i , then the element w_{ij} of spatial weight matrix W can be defined as

$$w_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{A}(s_i) \\ 0 & \text{otherwise} \end{cases}$$

Queens contiguity

Let $\mathcal{A}(s_i)$ be a set of units that share a side or an edge with unit s_i , then the element w_{ij} of spatial weight matrix W can be defined as

$$w_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{A}(s_i) \\ 0 & \text{otherwise} \end{cases}$$

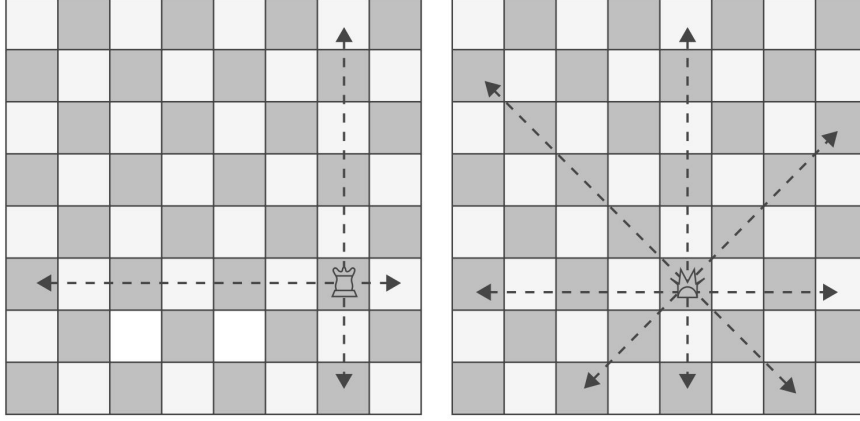


Figure 1.1: Contiguity criteria in a regular square lattice grid. Rook's move (left) and Queen's move (right) ¹

1.2.3.2 Distance-based weight matrix

Another important type of weight matrix depends on the centroid distances, d_{ij} ($1 \leq i, j \leq N$), between locations. The advantage of distance-based weight matrix is that distances are easily computed and the existence of descending effects with distance.

K-nearest neighbour weights

Let $\mathcal{A}(s_i)$ be a set of units that contains the k closest units to s_i . For each given k , the element w_{ij} of k-nearest neighbour weight matrix W can be defined as

$$w_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{A}(s_i) \\ 0 & \text{otherwise} \end{cases}$$

Radial distance weights

¹Source from: [Arbia \(2014\)](#)

Let d denote a threshold distance, beyond which there is no direct spatial dependence between units s_i , then the element w_{ij} of radial weight matrix W can be defined as

$$w_{ij} = \begin{cases} 1 & \text{if } 0 < d_{ij} < d \\ 0 & \text{otherwise} \end{cases}$$

Power distance weights

If there are supposed to be diminishing effects, a negative power function of distance can be defined as

$$w_{ij} = (d_{ij})^{-\alpha}$$

where α is any positive exponent, typically take as $\alpha = 1$ or $\alpha = 2$ and d_{ij} represents the distance between unit i and unit j .

Exponential distance weights

In this type, negative exponential functions of distance is a substitute to negative power functions, which can be defined as

$$w_{ij} = e^{-\alpha d_{ij}}$$

where α is a positive constant and d_{ij} represents the distance between unit i and unit j .

Double-power distance weights

Let d denote a threshold distance beyond which there is no direct spatial dependence between units s_i , then the element w_{ij} of radial weight matrix W can be defined as

$$w_{ij} = \begin{cases} [1 - (d_{ij}/d)^k]^k & \text{if } 0 < d_{ij} < d \\ 0 & d_{ij} > d \end{cases}$$

where k is any positive constant, typically take as $k = 2, 3, 4$.

1.3 Structure of the Thesis

This Ph.D. thesis sheds new light on both theoretical investigation into the spatio-temporal model with varying coefficient structure and empirical application to the complex dynamic nature of spatio-temporal data in energy market.

Specifically, in Chapter 2, rather than using conventional non-spatial model or spatial methodologies based on conditional mean, we will empirically investigate the dynamic integrations or linkages of the USA's natural gas markets at the state level holistically by proposing a spatio-temporal network quantile analysis.

In Chapter 3, we are mainly concerned with the nonlinear modelling of econometric time series data from view point of quantile regression. Specifically, we consider a (strictly) stationary time series that are near-epoch dependent (NED), which is an extension of the work of [Welsh \(1996\)](#) and [Yu and Jones \(1998\)](#) under *i.i.d.* samples to time series. Our asymptotic normality result also generalizes [Lu et al. \(1998\)](#), [Honda \(2000\)](#) and [Hallin et al. \(2009a\)](#) who obtained a Bahadur representation in sense of strong convergence with rather strong conditions.

In Chapter 4, we propose a semiparametric family of *Dynamic Functional-coefficient Autoregressive Spatio-Temporal (DyFAST)* models to address the difficulty in modelling and analysis of our spatio-temporal data, $Y_t(s_i)$, $t = 1, \dots, T$ and $i = 1, 2, \dots, N$, which are stationary along time but non-stationary over the irregular locations in space. Moreover, both one-step and two-step estimation methods will be proposed to model the dynamic spatial neighbouring temporal-lagged effects with the irregular locations by spatial weight matrix. Both theoretical asymptotic properties of our model and Monte Carlo simulations will be established.

As a continuation, in Chapter 5, we release the hypothetical constraint on normality of $Y_t(s_i)$ and propose a semiparametric spatio-temporal quantile approach to further study the dynamic nature of non-normal spatio-temporal data. Both theoretical properties and Monte Carlo simulations will be investigated. The proposed method will be applied to investigate the energy market data sets, which can further illustrate the usefulness of our models.

Chapter 2

On Dynamic Linkages of the State Natural Gas Markets in the USA: Evidence from an Empirical Spatio-temporal Network Quantile Analysis

2.1 Background

The extent of integration or segregation among different local natural gas markets has received increasing attention in the energy economics literature (c.f., [Renou-Maissant, 2012](#)). On one hand, the policy of deregulation has made natural gas market into separate regional markets. Due to the significant differences in resource volumes and energy consumptions among different states, the regional natural gas markets became relatively independent of each other. On the other hand, a well-developed pipeline system for regional natural gas supplies and the law of one price among markets contribute to the price interactions across different natural gas spot markets (c.f., [Park et al., 2008](#)). Therefore, a regional natural gas market could be strongly or weakly related to the nearby natural gas markets, and accurate identification about the linkages of regional natural gas markets is of crucial importance to both enterprise investment decisions and government energy policy.

In this chapter, our main objective is to empirically investigate the dynamic linkages of the state-level natural gas markets in the USA by introducing a novel spatio-temporal network quantile econometric model. Note that some determinants of regional natural gas prices (e.g., geographic location, weather or network characteristics) are unobserved and likely to be spatially correlated, which makes it difficult to accurately analyse the regional natural gas prices. However, the spatial econometric method can consider regional natural gas prices in neighbouring states as proxies of the unobserved determinants which are likely to be spatially correlated. Therefore, our model can accurately estimate the dynamic impacts of the neighbouring state markets on, or their dynamic integration with, the natural gas market in a concerned state at different quantile levels. [Yu et al. \(2003\)](#) suggested that, differently from the mean (regression) analysis, the quantile analysis is more popular in the energy risk management analysis; and correctly estimating the tail behaviour of the price change distribution is crucially important for energy portfolio risk management and has significant policy implications (c.f., [Aloui et al., 2014](#); [Fan et al., 2008](#); [Hung et al., 2008](#); [Lahiani et al., 2017](#); [Shahbaz et al., 2018](#)).

The empirical findings suggest that positive spatial neighbouring effects exist significantly in the eastern and middle states' natural gas markets, and the effects of the national crude oil market on the natural gas markets exist significantly in southern and eastern states, which are clearly heterogeneous at different quantile levels. These results can help us to discover the heterogeneous dynamic linkages or market integrations across different states with a broader perspective on decision-making for management of energy risks by considering the risk transmissions of local neighbour state natural gas markets and national oil market in this model. The recognised tail behaviour in price shocks from the quantile analysis can help investors to hedge against energy price risks maximizing their profits. Furthermore, the findings from this study can also help local industry or government to mitigate the negative impacts from the expected or unexpected fluctuations in the oil and the neighbouring natural gas markets, which will enact appropriate state-level price discovery, investment decisions and energy policies.

2.2 Literature Review

In this section, we review the studies that analyse the energy market integration, in particular those focusing on the relationship between the crude oil and the natural

gas markets, and summarize the theoretical and empirical literature which consider the spatial econometrics models and quantile methods.

Some previous literature has analysed the energy markets, either at a theoretical or at an empirical level. For example, focusing on the integration of regional retail fuel markets, [Holmes et al. \(2013\)](#) considered the pair-wise approach ([Pesaran, 2007](#)) to test the regional integration in the US gasoline market, which verifies the law of one price in the regional gasoline markets in the US. This approach was also applied by [Cárdenas et al. \(2017\)](#) on the diesel market integration in France. [Blair et al. \(2017\)](#) used an error-correction model to investigate the regional differences in the price pass-through in the US gasoline markets. However, on the natural gas markets in the USA concerned about in this Chapter, the literature appears not that extensive with regard to the market integration. See, for example, [De Vany and Walls \(1993\)](#), [Cuddington and Wang \(2006\)](#), [Park et al. \(2008\)](#), and [Ghoddusi \(2016\)](#) who considered the integration and liberalization of the North American natural gas market, while [Neumann et al. \(2006\)](#), [Renou-Maissant \(2012\)](#) and [Kuper and Mulder \(2016\)](#) examined the natural gas market integration in the European countries, by using time-series analysis. Another strand of the literature has examined the equilibrium relationship between the natural gas and the crude oil markets. Most of these studies adopted time series modelling methods; see, e.g., [Yücel and Guo \(1994\)](#), [Serletis and Rangel-Ruiz \(2004\)](#), [Bachmeier and Griffin \(2006\)](#), [Panagiotidis and Rutledge \(2007\)](#), [Brown and Yücel \(2008\)](#), [Wolfe and Rosenman \(2014\)](#). The existing literature has rarely investigated the relationship between the state-level natural gas markets and the national crude oil market in the U.S., and also typically neglected the spatial neighbouring linkages or dynamic market integrations and the tail behaviour of the extreme price changes in the state-level natural gas markets. Differently, we will propose a spatio-temporal network quantile econometric model, which combines the spatial econometrics and quantile regression, to explore the issues above in a more holistic manner.

The spatial neighbouring effects are mostly neglected in the previous studies on the natural gas markets. These studies, either based on a national perspective or ignoring the geographical heterogeneity in different states, may get a biased estimate of the market relationship. A spatial econometric model can help to measure the horizontal (multi-state) transmission effect or dynamic market integration. The methodology has been applied in other research fields. For example, in urban economics field, [Fingleton \(2008\)](#) developed a spatial GMM estimator for a house price model with moving average errors. [Holly et al. \(2010\)](#) constructed a spatio-temporal model to investigate the US house prices at the state level and

identify a significant spatial dependence. In environmental sciences, [Auffhammer and Carson \(2008\)](#) used province-level data to construct a spatial econometrics model to forecast China's emissions. [Yu \(2012\)](#) analysed the influential factors of China's regional energy intensity considering spatial dependence. Some empirical literature that use station-level data to analyse the gasoline price dispersion can be found in e.g., [Clemen and Gugler \(2006\)](#) and [Yilmazkuday and Yilmazkuday \(2016\)](#).

In addition, the tail behaviour in extreme price changes of the natural gas was also often neglected. Most of the existing literature examine the linkage in conditional mean between energy markets. See, e.g., [Hamilton \(2009\)](#), [Lee et al. \(2012\)](#), [Li et al. \(2012b\)](#), [Alquist and Gervais \(2013\)](#). For example, numerous studies use the time series regression or vector autoregressive (VAR) model in conditional mean to test the causal relationship between the oil price and other variables (c.f., [Hamilton, 1983](#); [Jones and Kaul, 1996](#); [Papapetrou, 2001](#); [Cong et al., 2008](#)). The classical conditional mean regression is a widely used statistical method in data analysis. In particular, if the error is normally distributed, the results of mean regression method are unbiased and efficient. However, from a risk management perspective, risk managers often focus on the tail behaviour of distributions, and the energy price returns do not comply with the normal distribution, having an obvious peak and fat tails. In this case, the classic conditional mean regression will yield less informed results. Moreover, the outliers in energy prices may also affect the estimation accuracy of the classical mean regression. Therefore, the quantile regression method proposed by [Koenker and Bassett Jr \(1978\)](#) has many advantages in comparison with conditional mean regression (c.f., [Bera et al., 2016](#); [Sherwood and Wang, 2016](#); [Xu and Lin, 2016](#), for a review). An increasing number of researchers have used quantile regression to analyse energy data. For example, [Sim and Zhou \(2015\)](#) and [Reboredo and Ugolini \(2016\)](#) applied quantile approach to examine the effect of oil price shocks on stock returns at different quantile levels, [Fan et al. \(2008\)](#) and [Hung et al. \(2008\)](#) estimated the value at risk for energy prices by using GARCH models, and [Aloui et al. \(2014\)](#) examined the dependence relationship between the WTI crude oil and the Henry Hub natural gas markets in extreme levels of quantiles.

Summarizing above, combining spatio-temporal and quantile regression methods can make it more comprehensive to investigate the complex relationships between regional natural gas markets and the national oil market. We therefore apply a spatio-temporal network quantile econometric model to investigate the dynamic relationships or market integration among the these markets.

2.3 Data and Methodology

2.3.1 Data

Our data consist of monthly natural gas commercial prices of 48 states (excluding Alaska, Hawaii and the District of Columbia) in the US and monthly West Texas Intermediate WTI crude oil, and Henry hub natural gas, spot prices¹. As there are missing data after Dec 2016 in some states, we choose the time interval from Jan 1997 to Dec 2016 as our sample period. Commodity prices are expressed in US dollars.

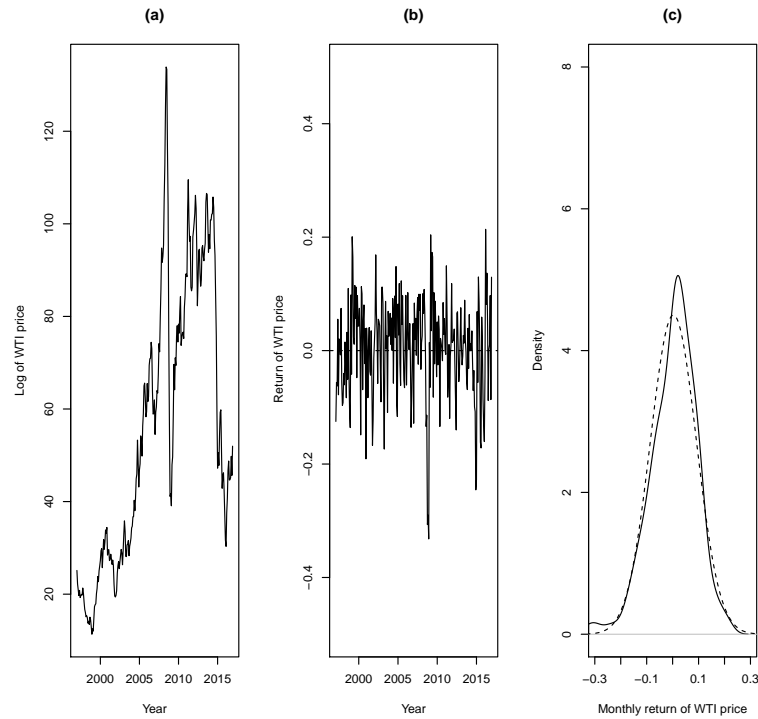


Figure 2.1: The time-series plots of the monthly WTI price data from Jan 1997 to Dec 2016 (left) and its monthly return data (middle) and the density plot (the kernel density estimate in solid line and the normal density estimate in dashed line) of the monthly return data (right).

¹Refer to the EIA Energy Glossary: <https://www.eia.gov/tools/glossary/>

Table 2.1: The test statistics for monthly natural gas prices of 48 states and Henry hub

States	Mean	SD	Skewness	Kurtosis	ADF	JB	KS
Alabama	11.607	2.975	-0.079	-0.717	0.913	0.077	0.000
Arizona	9.319	2.282	0.046	-0.904	0.957	0.019	0.000
Arkansas	8.366	2.135	0.073	-0.715	0.580	0.080	0.000
California	8.256	1.952	1.040	1.534	0.013	0.000	0.000
Colorado	7.274	1.904	-0.057	-0.628	0.094	0.147	0.000
Connecticut	9.575	2.730	0.432	-0.081	0.555	0.022	0.000
Delaware	11.846	3.296	0.051	-0.975	0.538	0.010	0.000
Florida	10.468	2.504	0.322	0.514	0.376	0.028	0.000
Georgia	10.427	3.294	0.339	0.468	0.463	0.028	0.000
Idaho	7.877	2.000	0.067	-0.612	0.815	0.158	0.000
Illinois	9.117	2.753	0.521	0.246	0.053	0.003	0.000
Indiana	8.621	2.474	0.674	0.286	0.316	0.000	0.000
Iowa	7.819	2.104	0.459	0.077	0.252	0.013	0.000
Kansas	9.716	3.085	0.180	-0.551	0.495	0.127	0.000
Kentucky	9.493	2.831	0.518	0.321	0.660	0.002	0.000
Louisiana	8.913	2.333	0.604	0.379	0.424	0.000	0.000
Maine	11.388	3.134	-0.056	-0.455	0.350	0.367	0.000
Maryland	9.867	2.216	0.245	-0.337	0.367	0.181	0.000
Massachusetts	10.930	2.883	0.362	-0.142	0.432	0.066	0.000
Michigan	8.013	2.082	0.068	-0.767	0.623	0.055	0.000
Minnesota	7.414	2.151	0.633	0.521	0.253	0.000	0.000
Mississippi	8.158	2.453	0.879	0.967	0.516	0.000	0.000
Missouri	9.696	2.570	-0.207	-0.771	0.838	0.025	0.000
Montana	8.393	2.315	0.405	0.245	0.431	0.025	0.000
Nebraska	6.747	1.919	0.702	0.635	0.398	0.000	0.000
Nevada	8.411	2.125	0.364	-0.936	0.942	0.001	0.000
New.Hampshire	11.878	3.086	-0.305	-0.747	0.531	0.011	0.000
New.Jersey	8.821	2.993	0.079	0.158	0.512	0.744	0.000
New.Mexico	7.108	2.280	0.689	0.791	0.535	0.000	0.000
New.York	8.601	2.536	0.468	-0.270	0.585	0.009	0.000
North.Carolina	9.772	2.593	0.669	0.215	0.647	0.000	0.000
North.Dakota	7.103	2.117	0.724	0.718	0.148	0.000	0.000
Ohio	8.672	2.509	0.727	-0.057	0.623	0.000	0.000
Oklahoma	9.975	3.679	0.381	-0.865	0.340	0.002	0.000
Oregon	9.102	2.382	-0.218	-0.722	0.814	0.032	0.000
Pennsylvania	10.478	2.387	0.569	0.135	0.543	0.001	0.000
Rhode.Island	12.918	3.307	0.178	-0.957	0.541	0.006	0.000
South.Carolina	9.750	2.653	1.007	1.696	0.609	0.000	0.000
South.Dakota	7.176	1.922	0.642	0.405	0.295	0.000	0.000
Tennessee	9.234	2.428	0.781	1.001	0.543	0.000	0.000

Continued on next page

Table 2.1 – continued from previous page

States	Mean	SD	Skewness	Kurtosis	ADF	JB	KS
Texas	7.486	2.224	0.708	0.945	0.501	0.000	0.000
Utah	6.686	1.595	-0.266	-0.404	0.235	0.115	0.000
Vermont	9.215	2.931	0.307	-1.018	0.942	0.001	0.000
Virginia	9.256	2.178	0.444	0.081	0.562	0.018	0.000
Washington	9.116	2.483	-0.363	-0.737	0.801	0.005	0.000
West.Virginia	10.085	3.046	0.649	-0.320	0.909	0.000	0.000
Wisconsin	7.598	2.189	0.580	0.052	0.488	0.001	0.000
Wyoming	7.064	1.887	0.130	-0.224	0.069	0.582	0.000
Henry hub	4.481	2.259	1.334	2.041	0.218	0.000	0.000

Note: SD represents the standard deviation. ADF is for the p -value of augmented Dickey-Fuller test for unit root, with alternative hypothesis of being a stationary series. JB is for the p -value of empirical statistic of the Jarque-Bera test for normality, with null hypothesis of being normal. KS is for the p -value of empirical statistic of the Kolmogorov-Smirnov test for normality, with null hypothesis of being normal.

Table 2.1 displays summary statistics for monthly natural gas prices, where ADF is for the p -value of augmented Dickey-Fuller test for unit root, with alternative hypothesis of being a stationary series. The result of unit root test shows that the prices series are not stationary (except for California). Jarque-Bera test indicates the non-normality of most natural gas price series. Moreover, the results of Kolmogorov-Smirnov test also be confirmed that the natural gas price series distribution differs from a normal distribution. For non-normal data, the quantile regression with $\tau = 0.5$ is more robust than the mean regression. In addition, quantile estimation could better describe the tail behaviour and offer more tail information of natural gas prices in specific market circumstances, which is of crucial importance in the energy risk management. The time series plots of the monthly WTI crude oil price and its return series as well as the density plot are displayed in Figure 1, which shows that the WTI price series is nonstationary while its return series appears stationary and non-normal. Therefore, we will consider the returns of the natural gas commercial prices and the WTI crude oil price by taking the difference in the logarithm of two consecutive monthly prices.

In addition, we consider the effects of the long-run price equilibrium of the WTI crude oil and the Henry Hub natural gas markets on the state-level natural gas markets in our model. Although the WTI crude oil and the Henry hub natural

gas prices are non-stationary, they share a common long-run trend (Hartley et al., 2008; Villar and Joutz, 2006). Therefore, a cointegrating relationship between them may exist and have significant impact on the regional natural gas markets in the U.S.. Ramberg and Parsons (2012) found a structural break in February 2009 by using the cointegration test based on the augmented Dickey-Fuller test statistic (c.f., Gregory and Hansen, 1996). Therefore, we estimate the following equation:

$$\ln Z_t = \phi_0 + \phi_1 \ln X_t + \phi_2 D_t + \epsilon_t, \quad (2.3.1)$$

and do the cointegration test by the approach of Gregory and Hansen (1996), where Z_t denotes the Henry hub natural gas price at time t , X_t is the WTI crude oil price at time t , and D_t is a dummy variable used to account for the structure change ($D_t = 0$ if time is between Jan 1997 and Feb 2009, and $= 1$ otherwise). Here ϵ_t is the stochastic error term, used to describe the long-run price disequilibrium between the WTI crude oil and Henry Hub natural gas markets, which will be shown to have important effect on the state-level natural gas markets.

2.3.2 Methodology

In order to comprehensively identify the market linkages among different state natural gas markets and uncover the underlying impacts of the national crude oil market on the state-level natural gas markets, we incorporate quantile regression method into spatio-temporal autoregressive model. We, therefore, construct a spatio-temporal network τ -th quantile econometric model as follows:

$$\begin{aligned} & Q_\tau(\Delta \ln Y_t(s_i) | \Delta \ln X_{t-1}, Y_{t-1}^{\text{sl}}(s_i), \dots, Y_{t-p}^{\text{sl}}(s_i), \Delta \ln Y_{t-1}(s_i), \dots, \Delta \ln Y_{t-q}(s_i)) \\ &= \alpha_{0,\tau}(s_i) + \sum_{j=1}^p \lambda_{\tau,j}(s_i) Y_{t-j}^{\text{sl}}(s_i) + \sum_{l=1}^q \beta_{\tau,l}(s_i) \Delta \ln Y_{t-l}(s_i) + \alpha_{1,\tau}(s_i) \Delta \ln X_{t-1} \\ & \quad + \gamma_\tau(s_i) \widehat{\epsilon}_{t-1}, \end{aligned} \quad (2.3.2)$$

$$Y_t^{sl}(s_i) = \sum_{k=1}^N W_{ik} \Delta \ln Y_t(s_k), \quad i = 1, \dots, N, t = 1, \dots, T.$$

Here $\Delta \ln Y_t(s_i) = \ln Y_t(s_i) - \ln Y_{t-1}(s_i)$ denotes the return of natural gas price in state i at time t and $\Delta \ln X_t$ is the return of oil price at time t . The parameter $\lambda_{\tau,j}(s_i)$ accounts for the spatial autoregressive effect of j -order lag in state i at τ quantile. The time autoregressive coefficient $\beta_{\tau,l}(s_i)$ denotes the time l -order lag coefficient of $\Delta \ln Y_t(s_i)$ in state i at τ quantile. Note $\hat{\epsilon}_{t-1} = \ln Z_{t-1} - \hat{\phi}_0 - \hat{\phi}_1 \ln X_{t-1} - \hat{\phi}_2 D_{t-1}$ represents the estimated residuals for the shocks in the long-run equilibrium relationship between the WTI crude oil and Henry Hub natural gas prices as given in (2.3.1) and $\gamma_\tau(s_i)$ represents the effect of this disequilibrium on the state i 's natural gas price return. Here p and q stand for the orders of the spatial neighbouring and the concerned state autoregressive temporal lags, respectively.

The spatially lagged response variable $Y_t^{sl}(s_i)$ denotes spatial neighbouring effects of the dependent variable, where w_{ik} is a spatial weight for the (i, k) -th element ($1 \leq i, k \leq N$) of a pre-specified nonnegative spatial weighting matrix $W = [w_{jk}]_{j,k=1}^N$. This matrix reflects the geographic relationship between natural gas markets among different states. There are different kinds of weighting matrices used in the literature, among which distance function matrix and binary contiguity matrix are the most commonly used specifications (LeSage and Pace, 2009). In this study, considering the fact that natural gas is mainly transported by pipeline together with the common practice in econometrics, we choose the binary contiguity matrix for the spatial weight matrix. Let $\mathcal{A}(s_i)$ be a set of the states that connect with state i by natural gas pipelines. Then the element w_{ij} between states i and j in the spatial weight matrix W is defined by

$$w_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{A}(s_i), \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, as usual, $w_{ii} = 0$, and we normalize the spatial weights by row such that the summation of the row elements is equal to 1. Besides, the orders p and q for spatial neighbouring autoregressive lags and the temporal lags, respectively, will be decided by Akaike information criterion (AIC).

Then the unknown parameters in the model can be estimated in two steps. First, we use OLS regression to estimate the model $\ln Z_t = \phi_0 + \phi_1 \ln X_t + \phi_2 D_t + \epsilon_t$ and

the predicted residuals $\hat{\epsilon}_t = \ln Z_t - \hat{\phi}_0 - \hat{\phi}_1 \ln X_t - \hat{\phi}_2 D$ from this regression are saved. The second step is to estimate the model (2.3.2) by:

$$\begin{aligned}
 & (\hat{\alpha}(s_i), \hat{\lambda}(s_i), \hat{\beta}(s_i), \hat{\gamma}(s_i)) \\
 &= \arg \min_{(\alpha(s_i), \lambda(s_i), \beta(s_i), \gamma(s_i))} \sum_{t=1}^T \rho_{\tau} \left\{ \Delta \ln Y_t(s_i) - \alpha_0(s_i) - \sum_{j=1}^p \lambda_j(s_i) Y_{t-j}^{sl}(s_i) \right. \\
 & \quad \left. - \sum_{l=1}^q \beta_l(s_i) \Delta \ln Y_{t-l}(s_i) - \alpha_1(s_i) \Delta \ln X_{t-1} - \gamma(s_i) \hat{\epsilon}_{t-1} \right\}, \quad (2.3.3) \\
 & Y_t^{sl}(s_i) = \sum_{k=1}^N W_{ik} \Delta \ln Y_t(s_k), \quad i = 1, \dots, N, \quad t = 1, \dots, T,
 \end{aligned}$$

where $\rho_{\tau}(y) = y(\tau - \mathbf{1}_{y < 0})$ is the quantile check function, and $\mathbf{1}_A$ is the indicator function of set A which equals 1 if A is true and 0 otherwise.

This equation can be solved by linear programming. In order to have comprehensive information on the dynamic neighbouring linkages among the state-level natural gas markets and the linkages of the natural gas markets in 48 states to the crude oil market, we set five representative quantiles (i.e., 10th, 25th, 50th, 75th and 90th). In addition, we also use nine quantiles (i.e., 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th) to display the varying characteristics of different coefficients of interest. A bootstrap method (He and Hu, 2002) is used to obtain the confidence intervals of the estimate parameters in above model. Specific discussion about quantile regression method is referred to Koenker (2005).

2.4 Empirical Findings

2.4.1 Global spatial autocorrelation analysis

First, we want to recognize whether the average and the standard deviation of natural gas prices have spatial affinities with corresponding state dispersions (Figure 2.2). From Figure 2.2(a), we can clearly observe that the eastern states in the U.S. (such as Alabama, Delaware and Rhode.Island) have relatively higher natural gas prices than the states in western and central USA (such as Nebraska, Idaho and Wyoming) over our sample period. It also shows that the mean natural gas price in one state tends to be close to those of neighbouring states in the

same region, indicating that there can be dynamic linkages between natural gas markets across states. Moreover, the dispersion of state-level natural gas prices (Figure 2.2(b)) have a similar pattern. For example, states in the eastern America (such as Delaware, Georgia, Maine and West Virginia) have higher dispersions and fluctuations in natural gas prices, whereas the western states, such as California, Nebraska and Utah present a low volatility in natural gas prices. Taking together the results of Figure 2.2, we can find the 'spatial herding behaviour' in the state-level natural gas markets. This spatial clustering phenomenon is also supported by Moran's I test.

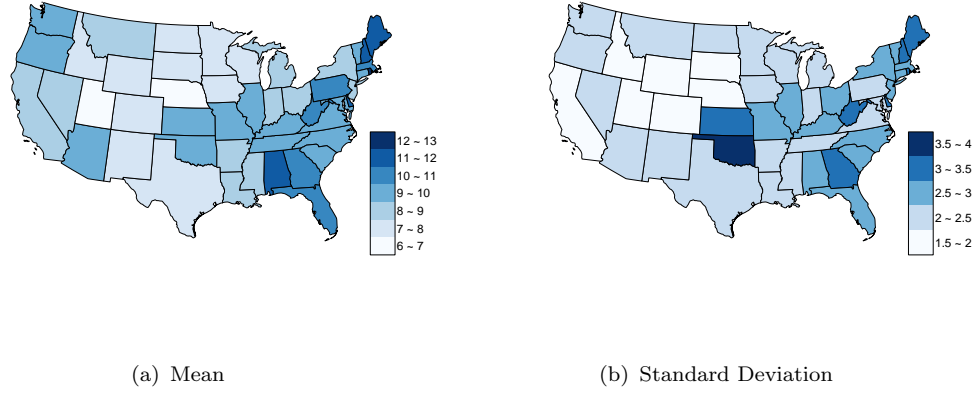


Figure 2.2: Spatial Distributions of Natural Gas Prices: Mean and Standard Deviation

The Moran's I test (Moran, 1950) defined as follows:

$$I_t = \frac{\sum_i^N \sum_j^N w_{ij}(x_{it} - \bar{x}_t)(x_{jt} - \bar{x}_t)}{\sum_i^N (x_{it} - \bar{x}_t)^2}, \quad (2.4.1)$$

where x_{it} is the natural gas price return for state i at time t , \bar{x}_t is the mean of x_{it} over the states i 's, w_{ij} is the row-standardised spatial weights matrix element defined in Section 3.2. The results of Moran's I over twenty years, given in Table 2.2, confirm that positive spatial autocorrelation is significant at the 5% significance level overall in most of the considered time periods.

The Moran's I test only assesses the overall pattern and the time trend of the spatial autocorrelation (Anselin, 1993). To further examine the clustering effects of spatial neighbours of the natural gas price returns in different states, we plot a scatter diagram of spatial neighbouring effects for year 2016, displayed in Figure

Table 2.2: The Moran's I test result

Date	Moran's I statistic	<i>p</i> -value
1997	0.0651	0.3726
1998	-0.0757	0.5559
1999	0.0395	0.5410
2000	0.1094	0.1643
2001	0.2738	0.0039
2002	0.2165	0.0214
2003	0.0691	0.3921
2004	0.2318	0.0165
2005	-0.0561	0.7460
2006	0.1593	0.0797
2007	0.2578	0.0087
2008	0.4587	0.0000
2009	0.2762	0.0054
2010	0.3010	0.0023
2011	0.2120	0.0298
2012	0.3551	0.0003
2013	0.3587	0.0004
2014	0.1776	0.0291
2015	0.3353	0.0008
2016	0.2823	0.0043

Note: This table displays Moran's I test results ([Moran, 1950](#)) for the spatial autocorrelation of regional natural gas price return of 48 states in the U.S. from 1997 to 2016.

2.3. The vertical axis represents the natural gas price returns x_{it} at time t for state i for 48 states, and the horizontal axis shows the corresponding states' temporal lag 1 spatial neighbouring natural gas price returns, $\sum_{j=1}^{48} w_{ij}x_{j,t-1}$, in these states. It clearly indicates that there appears a significant linear relationship between the two variables. Therefore, spatial neighbouring effect or local market integration seems to be an important factor that influences the state-level natural gas markets, and dynamic spatial models incorporating both spatial neighbouring effect and dynamic effect are then incorporated in the analysis of the state-level returns of the natural gas prices in the U.S. states in this Chapter.

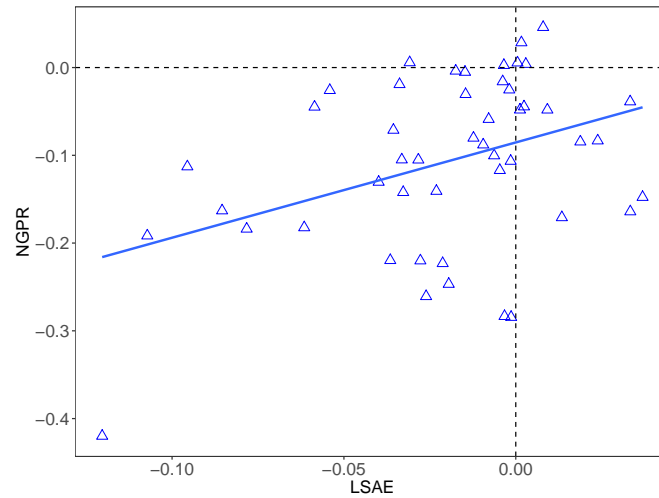


Figure 2.3: Scatter plot of spatial neighbouring effects of 48 states in the U.S. (2016). **Note:** The vertical axis represents the natural gas price returns in 48 states, and the horizontal axis shows the time lag 1 spatial neighbouring natural gas price returns in these states

2.4.2 Empirical results of spatio-temporal network quantile econometric modelling

To comprehensively investigate the effects of the crude oil market on state-level natural gas markets and the neighbour state effects on the monthly natural gas price returns, we choose five representative quantile levels (i.e., 10th, 25th, 50th, 75th and 90th) for implementing the quantile regression. Before proceeding to the rigorous empirical investigation, we first decide the orders of spatial neighbouring autoregressive lags and the temporal lags, p and q , in our model (2.3.2), by quantile regression Akaike information criterion (AIC), calculated by the function 'AIC.rq' of the R package "quantreg" (Koenker et al., 2018). The details of the AIC values for different orders of (p, q) for the 48 states are provided in Table 2.3. The result given in Table 2.3 shows that the chosen orders p and q by the AIC are only slightly different for each state. To ensure consistency, in our models, similarly to that in Blair et al. (2017), we use the same lag lengths for all states with $(p, q) = (1, 1)$ selected based on the AIC values for most of the states.

The estimation results for testing the cointegration between the WTI crude oil and the Henry hub natural gas market prices by using the approach of Gregory and Hansen (1996) under model (2.3.1) are provided in Table 2.4, and the results of the spatio-temporal network quantile econometric regression under model (2.3.2) in the Tables 2.5-2.9 (in the Appendix). In Figures 2.4-2.7, we depict, over states

Table 2.3: The AIC values for different pairs of orders for the 48 states

States	$p = q = 1$	$p = 1, q = 2$	$p = 2, q = 1$	$p = q = 2$
Alabama	-816.01	-812.88	-812.64	-811.09
Arizona	-895.42	-891.33	-893.46	-892.46
Arkansas	-524.64	-520.24	-526.02	-524.13
California	-472.52	-473.34	-469.58	-472.94
Colorado	-490.51	-488.34	-485.88	-487.22
Connecticut	-381.52	-382.56	-377.49	-380.80
Delaware	-488.04	-484.72	-490.36	-488.82
Florida	-801.50	-794.61	-794.57	-792.65
Georgia	-339.60	-344.10	-337.27	-342.21
Idaho	-958.52	-952.14	-952.89	-951.15
Illinois	-368.80	-366.12	-366.21	-364.22
Indiana	-307.95	-303.72	-312.52	-313.24
Iowa	-319.87	-316.17	-316.36	-320.40
Kansas	-440.02	-440.50	-441.05	-441.38
Kentucky	-543.71	-542.23	-561.66	-560.14
Louisiana	-573.12	-575.06	-571.32	-573.23
Maine	-366.91	-362.39	-363.10	-362.41
Maryland	-509.32	-509.04	-508.98	-508.88
Massachusetts	-396.49	-391.85	-395.65	-393.69
Michigan	-688.56	-681.71	-681.72	-679.74
Minnesota	-407.24	-415.90	-409.97	-414.07
Mississippi	-448.37	-451.81	-446.25	-450.19
Missouri	-683.83	-688.17	-690.54	-694.28
Montana	-528.90	-529.32	-529.27	-529.40
Nebraska	-462.39	-474.38	-462.43	-473.10
Nevada	-889.10	-886.07	-882.76	-884.19
New.Hampshire	-470.02	-472.30	-466.72	-471.46
New.Jersey	-310.45	-306.10	-305.42	-304.36
New.Mexico	-387.23	-399.91	-387.09	-402.98
New.York	-467.26	-466.70	-466.13	-469.95
North.Carolina	-577.54	-572.88	-574.06	-572.41
North.Dakota	-380.49	-377.06	-377.12	-375.35
Ohio	-572.81	-570.39	-570.42	-568.51
Oklahoma	-375.93	-374.50	-372.29	-374.80
Oregon	-682.83	-678.04	-678.01	-678.53
Pennsylvania	-672.05	-669.37	-669.55	-669.78
Rhode.Island	-545.84	-540.61	-539.77	-538.61
South.Carolina	-540.97	-536.36	-536.42	-534.53
South.Dakota	-382.68	-379.56	-379.36	-379.89
Tennessee	-572.07	-568.77	-570.95	-576.09
Texas	-468.81	-468.32	-468.70	-468.35
Utah	-636.50	-638.50	-631.13	-638.36
Vermont	-764.46	-757.86	-757.80	-755.98
Virginia	-635.56	-633.18	-633.52	-631.56
Washington	-764.25	-757.29	-759.43	-757.72
West.Virginia	-578.57	-575.79	-572.34	-575.92
Wisconsin	-348.33	-349.77	-348.38	-349.78
Wyoming	-521.61	-516.29	-524.60	-525.46

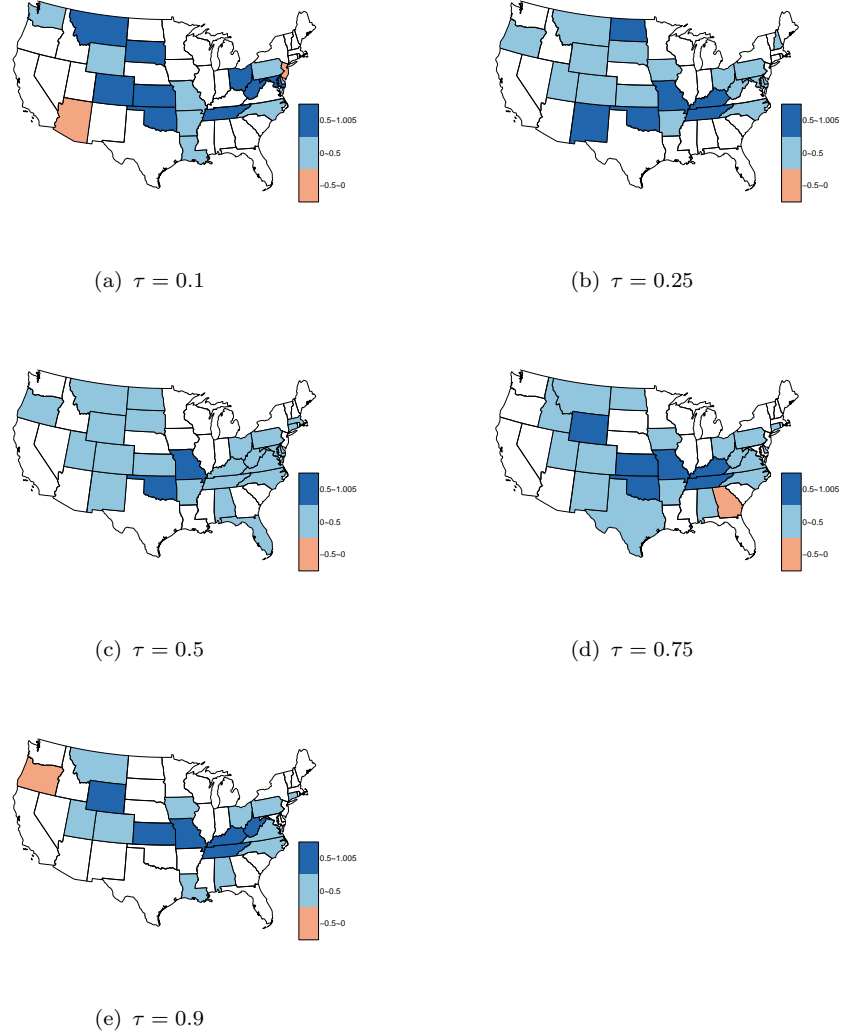


Figure 2.4: Maps of influence coefficients of spatial neighbouring effects.
Notes: The insignificant areas (the p -value of the coefficient is larger than 0.05) have been represented in white.

Table 2.4: The estimation result for cointegration test

	Estimate	Std. Error	t statistic	p -value
ϕ_0	-1.1322	0.1183	-9.5740	0.000
ϕ_1	0.7375	0.0327	22.5606	0.000
ϕ_2	-0.7987	0.0411	-19.4164	0.000

Note: This table displays the results of the estimation for the cointegration test by the method of [Gregory and Hansen \(1996\)](#), for the WTI crude oil and the Henry hub natural gas prices, in model (2.3.1).

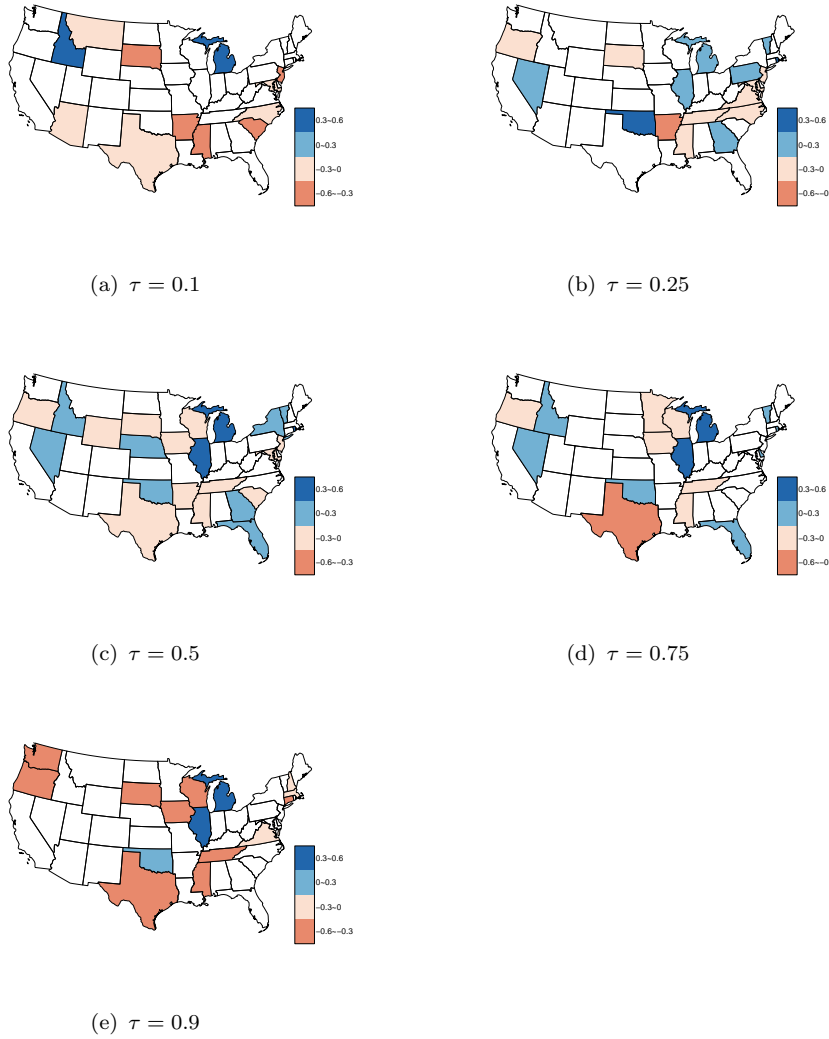


Figure 2.5: Maps of influence coefficients of autoregressive effects. **Notes:** The insignificant areas (the p -value of the coefficient is larger than 0.05) have been represented in white.

in maps, the results of estimating the coefficients $\alpha_1(s_i)$, $\lambda(s_i)$, $\beta(s_i)$ and $\gamma(s_i)$ at different quantile levels, respectively, where the null hypothesis of the coefficients being zero is rejected at 5% significance level. In Figure 2.4, it clearly shows that the large and significantly positive coefficients of the spatial neighbouring effects usually appear at the states in the middle of the USA. Further, Figure 2.5 shows the autoregressive effects of the natural gas price returns that are heterogeneous and more significant at the 50th quantile. Figure 2.6 illustrates the effects of the oil market on the state-level natural gas markets that are more significant in the south and southeast of the United States at different quantiles. Figure 2.7 displays the effects of the adjustment toward the long-run equilibrium relationship between

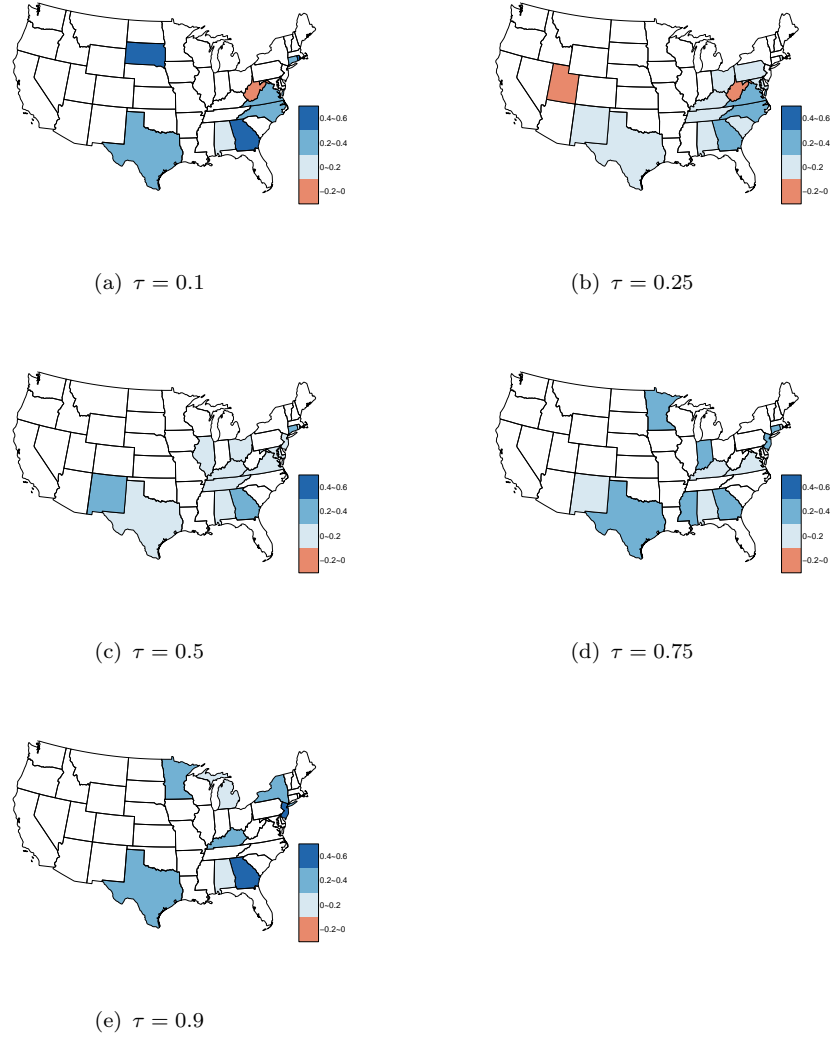


Figure 2.6: Maps of influence coefficients of oil price effects. **Notes:** The insignificant areas (the p -value of the coefficient is larger than 0.05) have been represented in white.

the WTI crude oil and the Henry Hub natural gas markets. These results clearly show that the tails of the price return distribution contain important information that the conditional mean regression cannot fully reveal. This suggests that the method of using spatial-temporal network quantile econometric regression model is appropriate and reasonable for this empirical analysis.

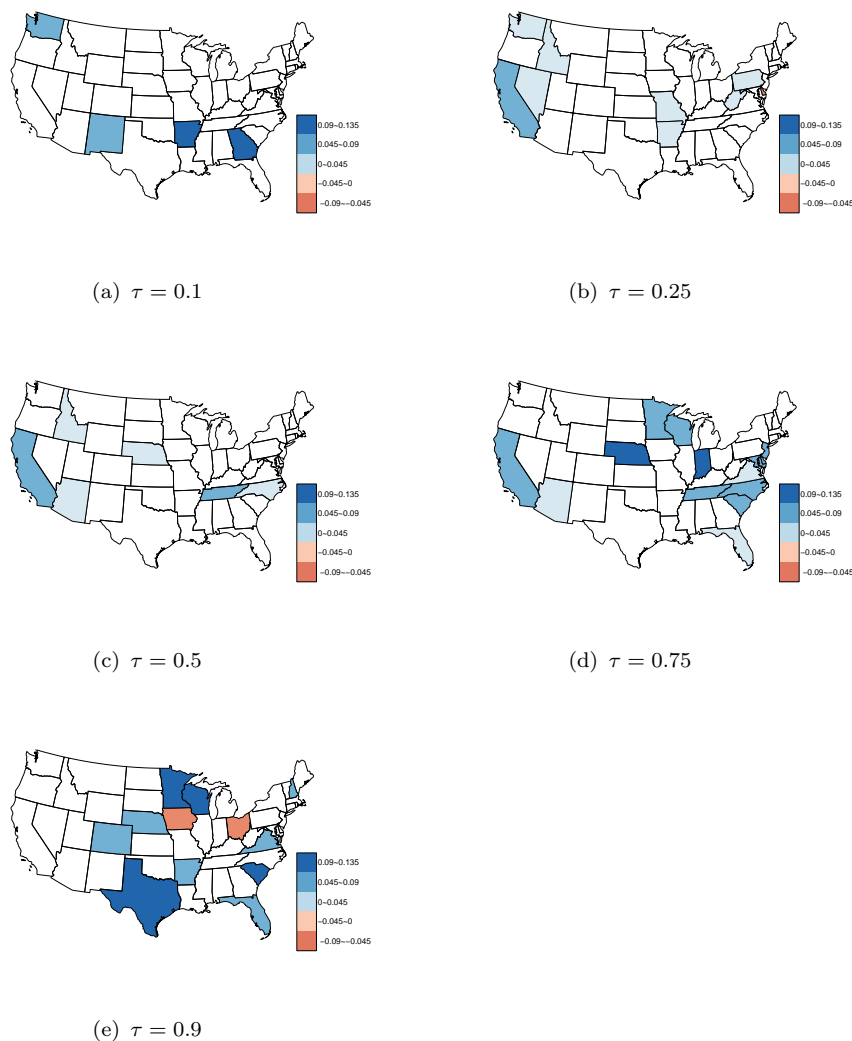


Figure 2.7: Maps of the effects of the long-run relationship of the WTI crude oil and Henry Hub natural gas prices on the state-level natural gas price returns. **Notes:** The insignificant areas (the p -value of the coefficient is larger than 0.05) have been represented in white.

2.4.3 Discussions

The spatial-temporal network quantile regression can visually reveal the spatial neighbouring effects of natural gas markets in the states of the USA and the effects of the oil market on state-level natural gas markets at different quantiles. The results for some typical states² are presented in Figures 2.9–2.10 on the spatial

²Here we choose Connecticut, Pennsylvania and West Virginia from the northeast region, Alabama and Tennessee from the southeast region, Ohio from the mid-east region, Colorado and Wyoming from the central region, New Mexico, Oklahoma and Texas from the southwest region, and Oregon from the western region (c.f., Figure 2.8). Figure 2.8 shows the information on the

neighbouring effects, autoregressive effects, the spillover effects of crude oil market and the effects of the shocks in the long-run price equilibrium between the WTI crude oil and the Henry Hub natural gas markets. From these figures, we can clearly see the dynamic linkages with significant heterogeneity on these effects³.

First, we analyze the impacts of the neighbouring states' natural gas markets or market integrations. From Figure 2.4, we can clearly observe that the linkages of local natural gas market to the neighbour natural gas markets are heterogeneous. In particular, there exist significant positive spatial neighbouring effects, indicating the strong linkage with the neighbour states, in the natural gas markets in most of the border states at the 5% significance level, as identified by comparing Figure 2.4 and Figure 2.8. For example, some eastern states of the USA, such as Connecticut, Ohio, Pennsylvania, Tennessee and West Virginia, have a significantly positive basin-shaped or U-shaped curve spatial neighbouring effect, respectively (Figure 2.9(c), Figure 2.9(e), Figure 2.10(b), Figure 2.10(c), and Figure 2.10(e)), which indicates the stronger integration of natural gas market at higher and lower quantiles. In some middle states, such as Colorado and Wyoming in the central region, the curve of the coefficients for the spatial neighbouring effects seems like a trigonometric function, with the values between 0.2 and 0.6 (Figure 2.9(b) and Figure 2.10(f)). And for New Mexico, we can observe an inverted U-shaped spatial neighbouring effect in Figure 2.9(d). In Figure 2.9(f), a decreasing trend of the spatial neighbouring effect implies that the lower the price changes in the natural gas markets of the neighbour states, the higher the influence on natural gas market in Oklahoma. Similar to Oklahoma, Oregon in the western region also has a decreasing trend of the spatial neighbouring effect (Figure 2.10(a)).

These significantly high degrees of linkages between state-level natural gas markets are also evident, implying that spatial arbitrage could enforce the law of one price and promote integration of the natural gas markets in the central and eastern states of the U.S. at different quantile levels, which is similar to the results of Cuddington and Wang (2006). Moreover, unlike Cuddington and Wang (2006), we find that the natural gas markets in southern states are under integration at

transportation capacity (in Million cubic feet per day) between the six geographical regions, international borders, and offshore Gulf of Mexico in 2016.

³To verify the asymmetry of the coefficients, we considered the Wald test for quantile regression coefficients. Due to the space limit, we only show the results of the Wald tests for the equality of slope coefficients (at the quantile levels of 0.1 against 0.5 and 0.9, and 0.5 against 0.9) for some typical states, provided in Table 2.10 in the Appendix. Complete results are available upon request. We found that there are significant differences between the coefficients under different quantile levels for some states, which largely support the heterogeneity of these effects.

upper quantiles while some western state-level natural gas markets are under integration at lower quantiles. There are two main reasons for the integration of the natural gas markets. The first one is owing to geographical locations. The transported volumes of the cross-state natural gas are large through the well-developed pipeline system in these border states (see Figure 2.8), and hence the transactions are more frequent in the border states. Therefore, the natural gas markets are more likely to be influenced by the nearby states in these areas. The second reason is the productions and consumptions of the natural gas, which also contribute to the spatial neighbouring effects. Specifically, for some states with production below consumption, such as Alabama, Connecticut, Missouri, Montana, Oregon and Wyoming, there are a high natural gas import from the nearby states to meet its own consumption requirement, which makes their natural gas markets significantly affected by the natural gas markets in the neighbour states. In contrast, in Colorado, Kansas, New Mexico, Ohio, Oklahoma, Pennsylvania and West Virginia, there exist vast resources of natural gas, where the status of the natural gas resources also makes their natural gas markets significantly influenced by the neighbour states.

Second, we examine the dynamic linkage between the future and the past of one state natural gas market. From Figure 2.5, there appear more states with significant dynamic linkage of the price changes (at the 5% significance level) in the natural gas markets around at 50th quantile than at other quantiles. This indicates that at the median level, the price changes in state-level natural gas markets in the future may be more easily influenced by its past. The results may attribute to the characteristics of the natural gas prices such as time-dependent and mean-reverting dynamics (Ramberg and Parsons, 2012; Thompson et al., 2009). In addition, there exist some states, e.g. Florida, Nevada, Vermont and Wisconsin, where significantly autoregressive effects are present in their natural gas markets, while spatial neighbouring or oil spillover effects on the natural gas markets are insignificant. Also, this finding may imply that the natural gas markets of these states are more segregated.

Third, the impacts of the national crude oil market are investigated. We can observe in Figure 2.6 that the effects of the national crude oil market on the natural gas markets exist significantly in the southern and eastern states, which are clearly heterogeneous. In particular, these effects in the northeast region, such as Maryland, Pennsylvania, Ohio, Rhode Island and Virginia are significant and positive at the lower quantile levels. This indicates that the natural gas markets in these states are vulnerable to the impact of price falling in crude oil market. However,

the effect of price changes in the crude oil market on the natural gas market in Connecticut is statistically significant at almost all quantiles. The curve of the coefficient seems like a trigonometric function, with the values between 0.1 and 0.4 (Figure 2.9(c)). Moreover, for the southern states, such as Alabama, Connecticut, Georgia, Kentucky, New Mexico, Tennessee and Texas, the effects from the crude oil market are statistically significant and positive at most quantile levels. More specifically, Alabama and Texas have a U-shaped curve for the impacts of the price changes in crude oil market on the natural gas markets at different quantile levels. As shown in Figure 2.9(a) and Figure 2.10(d), a U-shaped curve illustrates the higher (either positive or negative) price fluctuations (corresponding to higher or lower quantile levels) in the natural gas markets in Alabama and Texas are more susceptible to the impact of the crude oil price shocks. For New Mexico and Tennessee, the effects of price changes in the crude oil market are significantly positive and have an inverted U-shaped curve from the 30th to the 80th quantiles (Figure 2.9(d) and Figure 2.10(c)).

The main reason for the significant impacts of the national crude oil market on the state-level natural gas markets is owing to the relatively large-scale inflow-outflow capacity of the natural gas and the high proportion of the natural gas and oil consumptions in these states. This situation causes a strong substitutional relationship between the natural gas and the crude oil, implying that the price changes in the crude oil market have a significantly positive impact on the natural gas markets. In addition, in some border states (i.e. the states located on the boards of the six regions in Figure 2.8, such as Arkansas, Kansas, Missouri, etc.), although there exist no significant oil price effects, the natural gas markets in these states are still indirectly influenced by the price changes in the crude oil market. In other words, the natural gas markets in these states are affected, through the significantly spatial neighbouring effects, by the natural gas markets in the nearby states, which are influenced by the price changes in the crude oil market. Clearly, the impacts of the price changes in the crude oil market are transmitted through the spatial effects from the neighbour states' natural gas markets.

Finally, this chapter studies the effects of the long-run equilibrium shocks from the national crude oil and natural gas markets. As shown in Figure 2.7, the effects of the long-run relationship of the WTI crude oil and Henry Hub natural gas markets on the state-level natural gas markets are statistically significant at the 5% significance level, but they are heterogeneous in the western and eastern states of the USA. Averaging across all states, the adjustment process tends to be faster at higher or lower quantiles for deviations from the long run equilibrium level. In

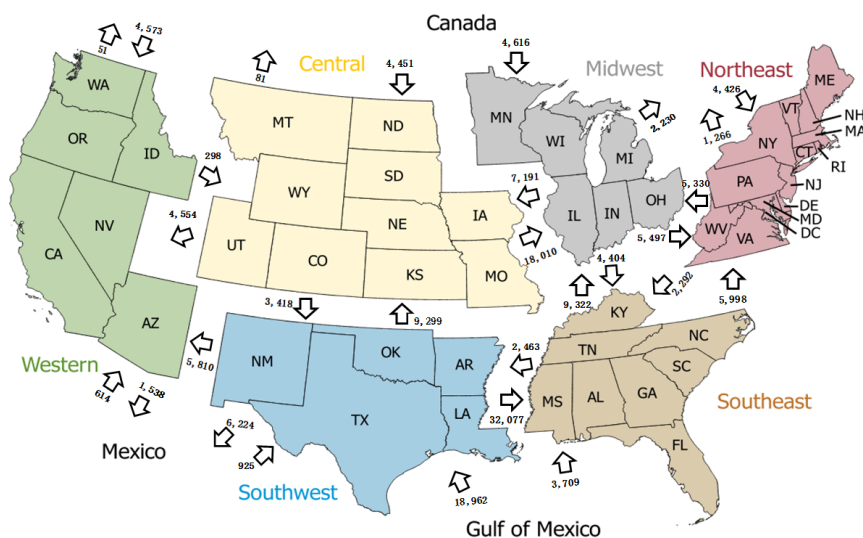


Figure 2.8: Region to region capacity map of natural gas in the U.S. in 2016. Source from: U.S. Energy Information Administration (<https://www.eia.gov/naturalgas>)

addition, for California, it only has significant long-run equilibrium shock effect. Specifically, there is statistically significant and positive effect of the long-run relationship of the WTI crude oil and Henry Hub natural gas markets around the middle quantile levels, while the adjustment effect is not statistically significant when the price has larger fluctuation in the natural gas market of California.

In summary, compared with the papers using mean regression analysis, our spatio-temporal network quantile model provides a more comprehensive picture of the dynamic linkages of the state-level natural gas markets, as indicated above on the spatial neighbouring effects, the autoregressive effects, the oil spillover effects and the effects of the long-run equilibrium shocks from the WTI crude oil and Henry Hub natural gas markets, for the natural gas markets of the 48 states in the USA. These findings can help us to better understand the heterogeneous effects of those dynamic linkages among different markets.

2.5 Conclusions

This chapter aims at exploring the dynamic linkages or market integrations of the state-level natural gas markets with the neighbouring states and the national crude oil market in terms of the spatial effects from the neighbour state natural

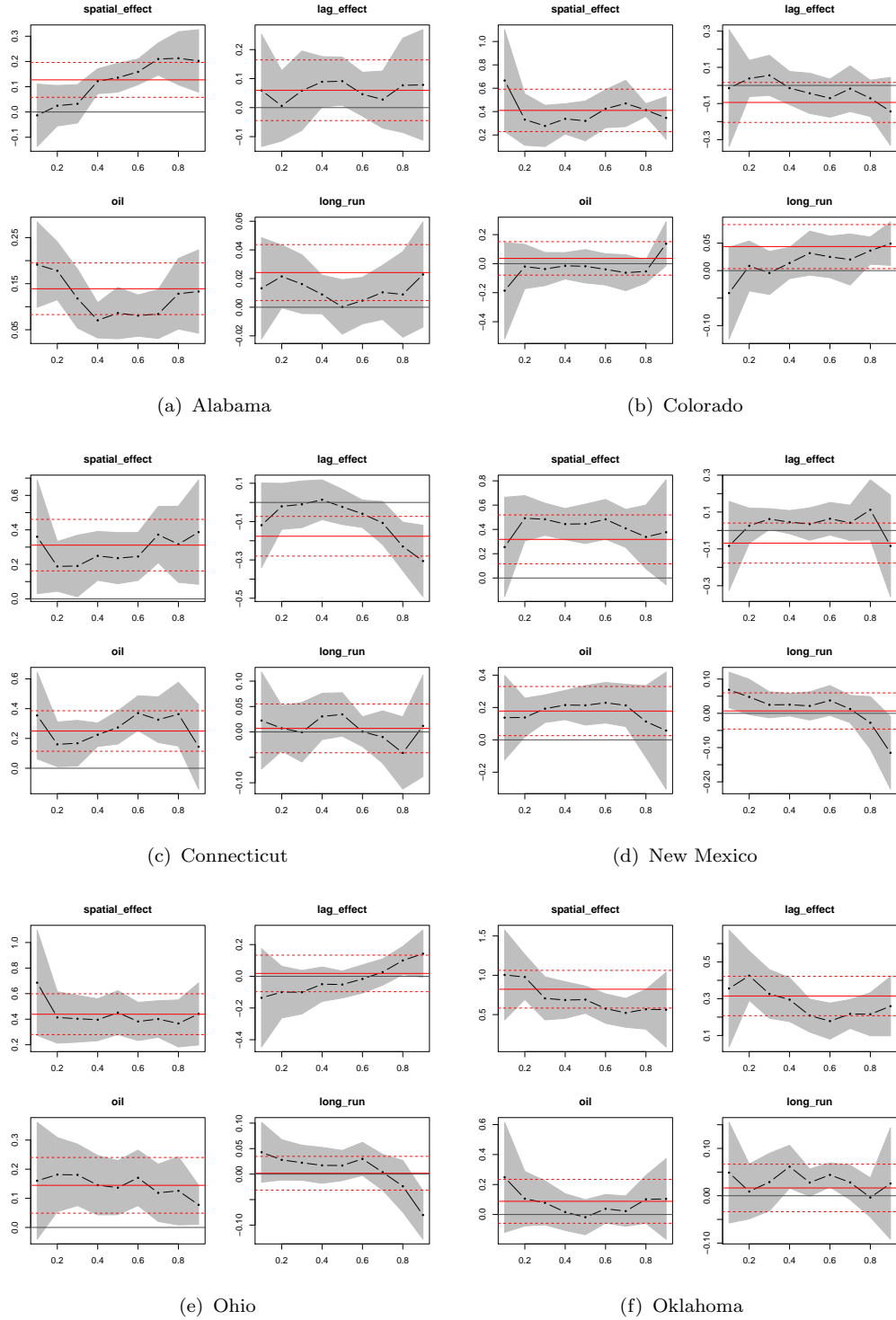


Figure 2.9: Quantile estimate: The effects of driving forces on natural gas price returns in six typical states. **Notes:** Shaded areas correspond to 95% confidence intervals of quantile estimation. The vertical axis indicates the elasticities of the explanatory variables. The red horizontal lines represent confidence intervals of OLS estimation.

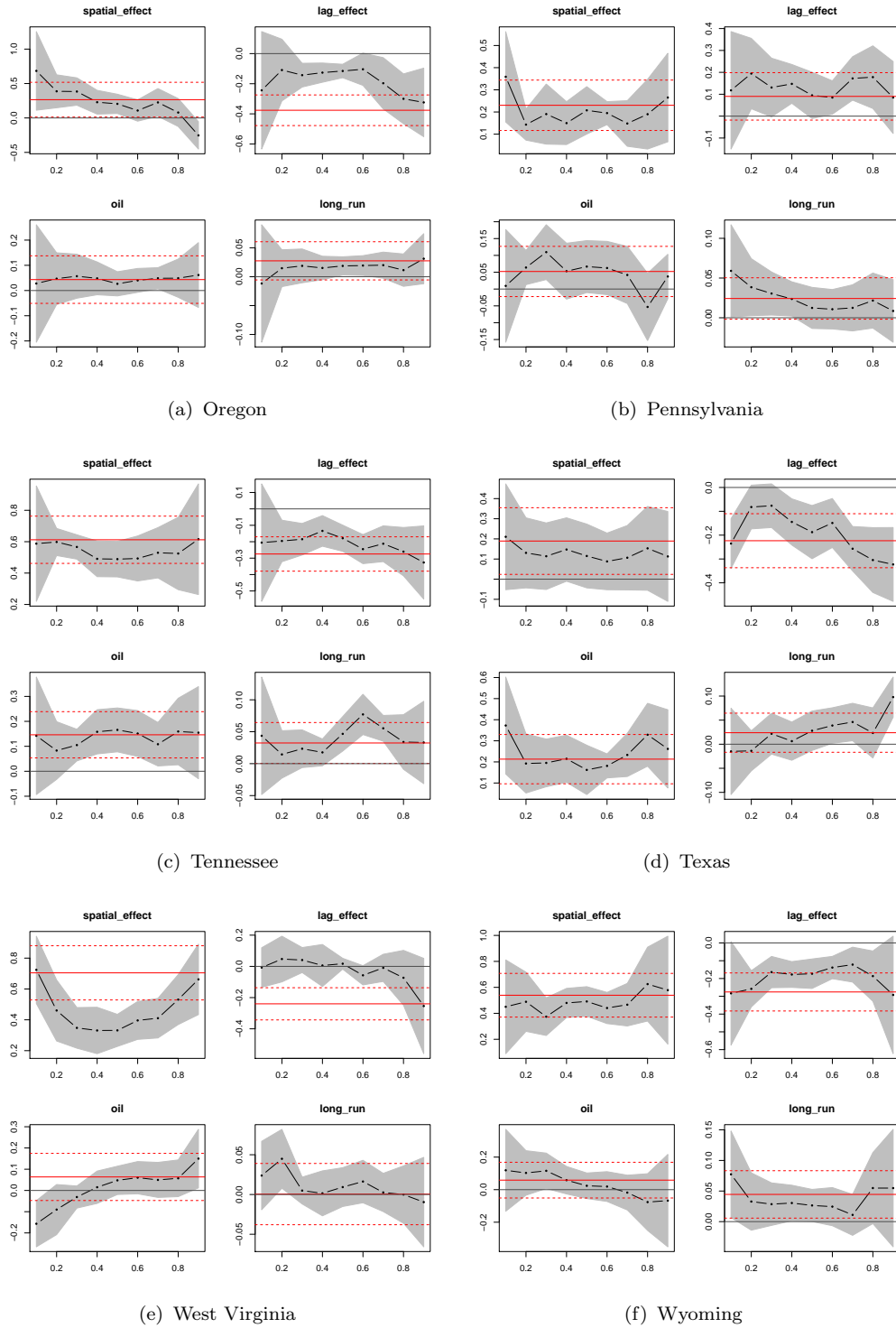


Figure 2.10: Quantile estimate: The effects of driving forces on natural gas price returns in other typical states. **Notes:** Shaded areas correspond to 95% confidence intervals of quantile estimation. The vertical axis indicates the elasticities of the explanatory variables. The red horizontal lines represent confidence intervals of OLS estimation.

gas markets as well as the oil price spillover effects from the WTI oil market. We introduce a spatio-temporal network quantile econometric model in this chapter. It can help to reveal the significant individual (state) and distributional (quantile) heterogeneities in the effects of the neighbouring state natural gas markets and the oil market price shocks among the 48 states in the USA. Compared with traditional mean regression, our model can provide us with comprehensive understanding of the spatial neighbouring or market integrations and the oil market price shock effects that impact the state-level natural gas prices.

This chapter uses the monthly data covering a period from Jan.1997 to Dec.2016 for the 48 states (excluding Alaska, Hawaii and the District of Columbia) in the USA. The empirical results indicate that the degree of the dynamic impacts owing to price shocks from both the local neighbouring natural gas markets and the national crude oil market, on the natural gas prices at a concerned state, varies by state. We find that significant local dynamic neighbouring market integrations exist in the natural gas markets not only in the eastern and central states as evidenced in the literature ([Cuddington and Wang, 2006](#); [Park et al., 2008](#)) but also in some western and southwest states at different quantile levels. Our results also show that there are significant linkages of the state-level natural gas markets to the national crude oil market through its lagged price shock and its long-run price equilibrium with the national gas market with varying price shock propagations at different quantile levels. The effects of the crude oil market price shocks on the state-level natural gas markets are heterogeneous and statistically significant in the southern states of the USA at different quantile levels. The impacts on the state-level natural gas markets may also come from the shocks or deviations from the long-run equilibrium of the WTI crude oil and Henry Hub natural gas markets, which are statistically significant in some western and eastern states and heterogeneous over states in the USA.

The results obtained can help local government and energy users to mitigate the negative impacts from the expected or unexpected fluctuations in the oil and the neighbouring natural gas markets. In particular, as a clean and cheap energy, natural gas can be used as a substitute to oil in many cases. An in-depth understanding of the relationship between the crude oil and the natural gas markets can help them better managing energy risk. Also, energy users in the border of the six geographical regions can pay more attention to the natural gas markets in the neighbour states to reduce their risk transmission from the neighbour states. Moreover, local state governments may formulate relevant energy policies on the basis of their geographical location and natural resource endowment. Quantile

regression methods can help to discover the full relationship between different variables, especially those which have fat tails in distributions and are important for energy risk management and policy decision-making. In short, the findings from this chapter can enact appropriate state-level price discovery, investment decisions and energy policies.

Combining the spatial econometrics and quantile regression method can help to better investigate the complex relationships between oil price and regional natural gas prices. The parametric models applied in this chapter may have some shortcomings in uncovering the underlying relationship for some states. Therefore, a semiparametric spatio-temporal quantile approach will be proposed in the Chapter 5.

2.6 Appendix

Table 2.5: The estimation results for the 0.1-level quantile

States	$\alpha_1(s_i)$		$\lambda(s_i)$		$\beta(s_i)$		$\gamma(s_i)$	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Alabama	0.1914	0.0008	-0.0133	0.8599	0.0593	0.6149	0.0132	0.5380
Arizona	-0.0329	0.3376	-0.0830	0.0008	-0.2458	0.0002	0.0177	0.3714
Arkansas	-0.1665	0.0951	0.4794	0.0142	-0.5564	0.0146	0.0949	0.0029
California	0.1413	0.3274	-0.1988	0.5083	0.1564	0.2517	0.0614	0.1971
Colorado	-0.1859	0.3577	0.6659	0.0126	-0.0148	0.9401	-0.0408	0.4211
Connecticut	0.3544	0.0465	0.3601	0.0741	-0.1187	0.3768	0.0224	0.6994
Delaware	-0.0665	0.4447	0.4796	0.0118	-0.0048	0.9786	-0.0330	0.4114
Florida	0.0991	0.0869	-0.0342	0.6430	-0.0121	0.8919	-0.0170	0.2575
Georgia	0.5325	0.0000	-0.2774	0.1200	0.0752	0.2845	0.0993	0.0143
Idaho	0.1145	0.0660	0.2008	0.1779	0.3982	0.0070	0.0166	0.2986
Illinois	0.1370	0.3476	0.3914	0.1080	0.1478	0.2631	0.0236	0.6623
Indiana	0.0721	0.7104	0.4953	0.0836	0.0315	0.8300	0.0469	0.4138
Iowa	0.1954	0.3817	0.0674	0.8232	-0.1023	0.6142	0.1137	0.0599
Kansas	0.0684	0.6699	0.5474	0.0254	0.3125	0.0590	-0.0226	0.7429
Kentucky	0.1123	0.2388	0.3324	0.1806	0.0839	0.6087	0.0544	0.0723
Louisiana	-0.0892	0.4094	0.4658	0.0355	-0.1621	0.3942	0.0283	0.5810
Maine	-0.1711	0.3877	0.1380	0.5984	0.0729	0.1327	0.0690	0.4768
Maryland	0.1533	0.1410	0.5032	0.0062	-0.2715	0.0307	0.0461	0.1010
Massachusetts	0.1191	0.4821	0.3839	0.1281	0.0757	0.5429	0.0483	0.3172
Michigan	0.0086	0.8989	0.1562	0.1946	0.3689	0.0037	-0.0090	0.7069
Minnesota	0.3219	0.0576	0.2803	0.1894	-0.2922	0.0640	-0.0302	0.5916
Mississippi	0.1193	0.4076	-0.1693	0.5260	-0.3215	0.0117	-0.0423	0.4090
Missouri	0.1766	0.1117	0.4208	0.0004	0.1781	0.2604	0.0333	0.2302
Montana	0.1047	0.2154	0.5221	0.0002	-0.2779	0.0026	0.0667	0.0524
Nebraska	0.2106	0.1766	-0.0646	0.7755	0.1558	0.3680	0.0502	0.3520
Nevada	0.0022	0.9502	0.0099	0.9445	0.1364	0.2809	0.0278	0.0890
New.Hampshire	-0.1064	0.2269	0.1205	0.3143	0.0299	0.8095	0.0792	0.0653
New.Jersey	0.2805	0.2323	-0.4261	0.0016	-0.3317	0.0000	-0.0620	0.3692
New.Mexico	0.1371	0.3886	0.2551	0.3067	-0.0840	0.5681	0.0685	0.0309
New.York	-0.0647	0.4655	0.1499	0.2910	0.0759	0.5197	0.0356	0.3146
North.Carolina	0.2135	0.0084	0.2924	0.0474	-0.2206	0.0343	-0.0059	0.8391
North.Dakota	0.1741	0.3531	0.6636	0.0521	-0.2428	0.3200	0.0317	0.5463
Ohio	0.1603	0.1905	0.6854	0.0067	-0.1356	0.4727	0.0429	0.2318
Oklahoma	0.2480	0.2680	1.0044	0.0042	0.3557	0.0671	0.0488	0.4510
Oregon	0.0279	0.8439	0.6836	0.0501	-0.2438	0.3050	-0.0117	0.8504
Pennsylvania	0.0094	0.9265	0.3589	0.0042	0.1179	0.4706	0.0590	0.0958
Rhode.Island	0.4488	0.0022	0.3276	0.0080	0.0470	0.7211	0.0068	0.8620
South.Carolina	0.0795	0.3834	-0.0040	0.9656	-0.3441	0.0004	-0.0245	0.4064
South.Dakota	0.5004	0.0015	0.6474	0.0055	-0.4369	0.0049	0.0373	0.4426
Tennessee	0.1421	0.3216	0.5875	0.0092	-0.2055	0.3458	0.0436	0.4362
Texas	0.3726	0.0082	0.2104	0.1891	-0.2353	0.0003	-0.0149	0.7840
Utah	-0.1413	0.1825	0.1954	0.3160	0.0664	0.6351	0.0207	0.6122
Vermont	-0.0315	0.7790	-0.0279	0.8280	0.2238	0.2688	0.0382	0.4272
Virginia	0.3624	0.0027	0.0933	0.6765	-0.1899	0.2643	0.0010	0.9812
Washington	0.1107	0.1084	0.3099	0.0336	-0.1949	0.0792	0.0798	0.0028
West.Virginia	-0.1568	0.0185	0.7239	0.0000	-0.0074	0.9237	0.0238	0.3633
Wisconsin	0.1098	0.6712	0.0232	0.9405	-0.1081	0.5886	0.0648	0.4867
Wyoming	0.1190	0.4375	0.4503	0.0413	-0.2835	0.1103	0.0774	0.0733

Table 2.6: The estimation results for the 0.25-level quantile

States	$\alpha_1(s_i)$		$\lambda(s_i)$		$\beta(s_i)$		$\gamma(s_i)$	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Alabama	0.1237	0.0017	0.0402	0.3906	0.0323	0.6716	0.0184	0.1702
Arizona	0.0238	0.3079	-0.0394	0.2078	-0.0454	0.5112	0.0190	0.0609
Arkansas	-0.0488	0.3366	0.3957	0.0000	-0.3393	0.0000	0.0393	0.0176
California	0.0374	0.6677	-0.2523	0.2107	0.0740	0.4232	0.0784	0.0103
Colorado	-0.0391	0.6245	0.3033	0.0193	0.0880	0.2780	-0.0135	0.6081
Connecticut	0.1145	0.3260	0.1837	0.1239	0.0677	0.4441	0.0049	0.9019
Delaware	0.0740	0.2321	0.4757	0.0000	0.0545	0.4715	-0.0568	0.0173
Florida	0.0524	0.1972	0.0252	0.7111	0.1484	0.1329	-0.0142	0.3756
Georgia	0.2826	0.0102	-0.1757	0.3153	0.1306	0.0409	0.0491	0.0783
Idaho	0.0332	0.2921	0.1015	0.1511	0.0616	0.3238	0.0225	0.0279
Illinois	0.1422	0.1168	0.1902	0.2019	0.2941	0.0004	0.0517	0.1057
Indiana	0.0483	0.6694	0.1919	0.2488	-0.0670	0.4896	0.0552	0.1053
Iowa	0.1592	0.1489	0.3632	0.0458	-0.1197	0.2656	0.0280	0.3765
Kansas	0.1044	0.2864	0.3860	0.0014	0.1219	0.1421	0.0486	0.0703
Kentucky	0.1858	0.0096	0.5575	0.0001	-0.1480	0.1382	0.0490	0.0644
Louisiana	0.0185	0.8041	0.1010	0.3547	0.0433	0.6804	0.0025	0.9211
Maine	0.0459	0.4500	-0.0332	0.7148	0.0487	0.5385	0.0370	0.2969
Maryland	0.1512	0.0168	0.3365	0.0010	-0.1859	0.0116	0.0393	0.0680
Massachusetts	0.0262	0.7396	0.2439	0.0762	-0.0047	0.9463	0.0267	0.3775
Michigan	0.0232	0.6970	0.1358	0.1679	0.2352	0.0093	-0.0018	0.9272
Minnesota	0.0319	0.7561	0.0619	0.6337	-0.0437	0.6232	-0.0360	0.2441
Mississippi	0.0441	0.5881	-0.1657	0.2695	-0.1694	0.0170	-0.0250	0.3938
Missouri	0.0319	0.4912	0.5128	0.0000	0.0869	0.2006	0.0312	0.0131
Montana	0.0268	0.7291	0.4615	0.0000	-0.1241	0.2862	0.0519	0.0636
Nebraska	0.1040	0.1579	-0.0257	0.7405	0.0849	0.1278	0.0106	0.5132
Nevada	0.0342	0.4300	0.0384	0.7333	0.2144	0.0110	0.0297	0.0432
New.Hampshire	0.0728	0.4577	0.0617	0.0416	0.0625	0.5292	0.0433	0.1635
New.Jersey	0.0778	0.4742	-0.0209	0.9002	-0.1505	0.0124	0.0254	0.5133
New.Mexico	0.1624	0.0043	0.5294	0.0000	0.0557	0.3521	0.0278	0.3118
New.York	-0.0101	0.8911	0.0135	0.8589	0.1373	0.0504	0.0270	0.2603
North.Carolina	0.2067	0.0011	0.2572	0.0118	-0.2030	0.0162	0.0203	0.2995
North.Dakota	0.1971	0.0531	0.6120	0.0013	-0.2408	0.0963	0.0492	0.1595
Ohio	0.1850	0.0037	0.4187	0.0001	-0.1333	0.1000	0.0154	0.4491
Oklahoma	0.0665	0.5026	0.7923	0.0000	0.4135	0.0000	0.0183	0.5996
Oregon	0.0810	0.2010	0.4017	0.0065	-0.1404	0.0000	0.0256	0.2186
Pennsylvania	0.0958	0.0456	0.1812	0.0200	0.2006	0.0044	0.0348	0.0327
Rhode.Island	0.1926	0.0273	0.0474	0.4900	0.3996	0.0001	0.0123	0.6188
South.Carolina	0.1585	0.0210	-0.0375	0.3069	-0.0956	0.3443	-0.0046	0.8298
South.Dakota	0.0637	0.5151	0.4412	0.0089	-0.2811	0.0092	0.0395	0.1755
Tennessee	0.1202	0.0153	0.5782	0.0000	-0.2231	0.0015	0.0108	0.5639
Texas	0.1912	0.0106	0.1311	0.1909	-0.0508	0.4124	0.0176	0.4881
Utah	-0.1262	0.0116	0.2945	0.0010	-0.0475	0.5178	0.0166	0.3166
Vermont	0.0108	0.8208	0.0265	0.6929	0.1773	0.0110	0.0055	0.7259
Virginia	0.2691	0.0000	0.2030	0.0607	-0.1741	0.0389	0.0214	0.2821
Washington	0.0133	0.7725	0.1237	0.1335	-0.0539	0.4179	0.0333	0.0013
West.Virginia	-0.0532	0.0175	0.4900	0.0000	0.0693	0.0768	0.0350	0.0486
Wisconsin	0.1577	0.1011	0.1770	0.1259	-0.1057	0.1187	0.0468	0.1455
Wyoming	0.0831	0.3025	0.3477	0.0130	-0.1350	0.1610	0.0190	0.5160

Table 2.7: The estimation results for the 0.5-level quantile

States	$\alpha_1(s_i)$		$\lambda(s_i)$		$\beta(s_i)$		$\gamma(s_i)$	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Alabama	0.0861	0.0118	0.1358	0.0001	0.0912	0.0718	0.0002	0.9856
Arizona	0.0085	0.7156	-0.0311	0.1919	-0.0237	0.5310	0.0204	0.0107
Arkansas	-0.0012	0.9824	0.2669	0.0026	-0.1268	0.0131	0.0283	0.1052
California	0.0948	0.2455	-0.3009	0.0897	-0.0503	0.5392	0.0761	0.0053
Colorado	-0.0179	0.7976	0.3204	0.0022	-0.0436	0.5203	0.0318	0.1931
Connecticut	0.2739	0.0001	0.2360	0.0094	-0.0232	0.6787	0.0344	0.1861
Delaware	0.0087	0.8884	0.3761	0.0000	0.1099	0.1450	0.0069	0.7067
Florida	0.0260	0.4388	0.0334	0.0175	0.2123	0.0016	0.0036	0.6683
Georgia	0.2406	0.0069	-0.0812	0.5097	0.0621	0.0020	0.0362	0.2023
Idaho	0.0089	0.5420	0.0491	0.0799	0.1377	0.0000	0.0087	0.0256
Illinois	0.1827	0.0337	0.1413	0.3306	0.3551	0.0000	0.0390	0.1783
Indiana	0.0225	0.8250	0.0117	0.9399	0.0309	0.7410	0.0678	0.0510
Iowa	0.1014	0.3454	0.3178	0.0692	-0.2146	0.0106	0.0480	0.1793
Kansas	0.0989	0.1918	0.2416	0.0479	0.1597	0.0760	0.0123	0.6150
Kentucky	0.1281	0.0357	0.4667	0.0000	0.0179	0.8022	0.0189	0.3713
Louisiana	0.0223	0.6666	0.0296	0.6881	0.1327	0.0523	-0.0057	0.7474
Maine	0.0442	0.3907	0.0100	0.8208	0.0313	0.5310	0.0055	0.7158
Maryland	0.1713	0.0187	0.4282	0.0007	-0.1852	0.0251	0.0249	0.3076
Massachusetts	0.0542	0.4830	0.2778	0.0241	-0.0817	0.1299	0.0263	0.3351
Michigan	0.0322	0.3634	0.0771	0.2619	0.3994	0.0000	0.0121	0.4269
Minnesota	0.0649	0.2863	0.0259	0.7957	-0.0175	0.7652	0.0145	0.4801
Mississippi	0.1080	0.0703	-0.1619	0.1989	-0.1881	0.0016	0.0190	0.4433
Missouri	0.0674	0.1374	0.6063	0.0000	-0.0194	0.7373	0.0078	0.5717
Montana	0.0558	0.2777	0.3216	0.0000	-0.0310	0.4706	0.0014	0.9396
Nebraska	0.0269	0.7205	-0.0024	0.9778	0.1229	0.0369	0.0403	0.0142
Nevada	-0.0053	0.7830	0.0662	0.2021	0.1353	0.0015	0.0074	0.2312
New.Hampshire	0.0718	0.1621	0.0257	0.7090	-0.0364	0.5567	0.0196	0.3839
New.Jersey	0.1533	0.0000	0.0248	0.7653	-0.1322	0.0014	0.0098	0.6718
New.Mexico	0.2136	0.0044	0.4460	0.0000	0.0346	0.5178	0.0217	0.3890
New.York	0.0420	0.5158	0.0247	0.7518	0.1883	0.0061	0.0141	0.5431
North.Carolina	0.0944	0.1185	0.3188	0.0003	-0.1323	0.0734	0.0398	0.0288
North.Dakota	0.0797	0.3421	0.3720	0.0048	-0.1434	0.1128	0.0492	0.0716
Ohio	0.1366	0.0165	0.4512	0.0000	-0.0519	0.3163	0.0168	0.3517
Oklahoma	-0.0180	0.8006	0.6902	0.0000	0.2081	0.0002	0.0278	0.1140
Oregon	0.0266	0.3633	0.2048	0.0168	-0.1154	0.0000	0.0188	0.0519
Pennsylvania	0.0667	0.1564	0.2074	0.0016	0.0949	0.1411	0.0124	0.4310
Rhode.Island	0.1108	0.0000	-0.0083	0.8401	0.3379	0.0000	-0.0001	0.9958
South.Carolina	0.0267	0.6078	-0.0275	0.2156	-0.1358	0.0067	0.0288	0.0906
South.Dakota	0.0746	0.3574	0.3300	0.0262	-0.2408	0.0165	-0.0067	0.8230
Tennessee	0.1658	0.0022	0.4888	0.0000	-0.1787	0.0003	0.0467	0.0048
Texas	0.1621	0.0233	0.1151	0.2331	-0.1874	0.0062	0.0279	0.2620
Utah	-0.0234	0.5618	0.2007	0.0015	-0.0795	0.1071	0.0245	0.0913
Vermont	-0.0014	0.9442	0.0216	0.3702	0.2085	0.0000	0.0042	0.5464
Virginia	0.1246	0.0204	0.2747	0.0026	-0.0825	0.2312	0.0148	0.4183
Washington	0.0064	0.7741	0.0520	0.1471	0.0061	0.8707	0.0070	0.3621
West.Virginia	0.0475	0.2413	0.3314	0.0000	0.0171	0.4178	0.0093	0.5314
Wisconsin	0.0422	0.6520	0.0313	0.7947	-0.2231	0.0002	0.0152	0.6023
Wyoming	0.0238	0.6181	0.4913	0.0000	-0.1731	0.0007	0.0264	0.0998

Table 2.8: The estimation results for the 0.75-level quantile

States	$\alpha_1(s_i)$		$\lambda(s_i)$		$\beta(s_i)$		$\gamma(s_i)$	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Alabama	0.1063	0.0041	0.2007	0.0000	0.0504	0.3989	0.0059	0.6460
Arizona	-0.0169	0.5139	-0.0290	0.2327	-0.0127	0.8332	0.0259	0.0032
Arkansas	0.0330	0.5879	0.2323	0.0270	-0.0946	0.1448	0.0405	0.0746
California	0.0915	0.2334	-0.1539	0.3484	-0.0567	0.4586	0.0539	0.0292
Colorado	-0.0848	0.1471	0.4092	0.0000	-0.0426	0.5093	0.0216	0.2204
Connecticut	0.3543	0.0015	0.3865	0.0007	-0.1497	0.0671	-0.0222	0.5104
Delaware	0.0386	0.5299	0.1889	0.0000	0.1017	0.0148	0.0495	0.0007
Florida	0.0570	0.0611	0.0023	0.9115	0.2252	0.0012	0.0293	0.0321
Georgia	0.3489	0.0000	-0.3363	0.0322	0.0685	0.3717	0.0400	0.1891
Idaho	0.0242	0.3942	0.0738	0.0280	0.1767	0.0037	0.0069	0.5077
Illinois	0.0746	0.4769	0.1230	0.4851	0.4018	0.0000	0.0342	0.2864
Indiana	0.2114	0.0082	0.2057	0.2126	-0.0908	0.3060	0.0979	0.0002
Iowa	0.0553	0.5948	0.4489	0.0076	-0.2497	0.0021	-0.0306	0.3284
Kansas	0.1603	0.1194	0.5083	0.0006	-0.0841	0.4815	0.0364	0.2953
Kentucky	0.1526	0.0193	0.6530	0.0000	-0.0809	0.3256	0.0422	0.0666
Louisiana	0.0115	0.8547	0.1686	0.0827	0.0432	0.6342	0.0048	0.8350
Maine	-0.0589	0.4764	-0.0908	0.2208	-0.0013	0.9874	-0.0159	0.5789
Maryland	0.1006	0.2672	0.1396	0.3613	-0.1605	0.1068	0.0655	0.0295
Massachusetts	0.0402	0.7114	0.2737	0.0565	-0.0766	0.2195	0.0622	0.0710
Michigan	0.0586	0.2959	0.1563	0.1052	0.4035	0.0000	0.0335	0.0926
Minnesota	0.3112	0.0006	0.1559	0.1301	-0.1695	0.0205	0.0528	0.0018
Mississippi	0.2064	0.0028	0.0140	0.9324	-0.2165	0.0000	0.0029	0.8802
Missouri	0.0563	0.1922	0.6197	0.0000	-0.0130	0.7958	-0.0093	0.4964
Montana	0.0429	0.4328	0.3857	0.0000	-0.1406	0.1253	0.0389	0.0585
Nebraska	0.0279	0.7736	0.0119	0.9248	0.0120	0.8842	0.0966	0.0028
Nevada	-0.0233	0.4236	-0.0054	0.9425	0.1299	0.0452	0.0039	0.6962
New.Hampshire	0.1007	0.2590	-0.0255	0.5261	-0.0690	0.3277	0.0194	0.4793
New.Jersey	0.2910	0.0025	0.0158	0.9095	-0.1055	0.3249	0.0766	0.0490
New.Mexico	0.1791	0.0131	0.4190	0.0000	0.0768	0.1886	0.0168	0.3559
New.York	0.0108	0.8744	0.1429	0.1828	0.0556	0.4526	-0.0341	0.1721
North.Carolina	-0.0023	0.9724	0.2137	0.0284	-0.1291	0.1343	0.0580	0.0120
North.Dakota	0.1446	0.2026	0.3753	0.0408	-0.2215	0.0752	0.0294	0.4353
Ohio	0.0898	0.1851	0.4524	0.0000	0.0650	0.2426	-0.0013	0.9286
Oklahoma	0.0822	0.2953	0.6077	0.0000	0.2282	0.0006	0.0025	0.9258
Oregon	0.0353	0.4096	0.2684	0.0987	-0.2121	0.0422	0.0167	0.3483
Pennsylvania	0.0072	0.8820	0.1586	0.0348	0.1273	0.0840	0.0202	0.2583
Rhode.Island	0.1673	0.0050	-0.0508	0.4962	0.2137	0.0029	0.0376	0.0921
South.Carolina	0.0142	0.8331	-0.0232	0.5143	-0.0764	0.4855	0.0570	0.0309
South.Dakota	-0.0723	0.5165	0.3000	0.1113	-0.2082	0.0681	-0.0119	0.7405
Tennessee	0.1070	0.1455	0.5604	0.0000	-0.1977	0.0155	0.0571	0.0090
Texas	0.3316	0.0000	0.2153	0.0434	-0.3340	0.0001	0.0487	0.1016
Utah	-0.0106	0.8489	0.2698	0.0007	0.0552	0.4812	0.0321	0.1630
Vermont	-0.0140	0.8078	0.0163	0.8023	0.2820	0.0000	0.0143	0.3965
Virginia	0.1063	0.0256	0.3506	0.0009	-0.1003	0.0768	0.0383	0.0183
Washington	-0.0064	0.8865	0.1478	0.2473	-0.0346	0.6630	0.0128	0.3980
West.Virginia	0.0364	0.4311	0.4244	0.0000	-0.0089	0.9221	-0.0033	0.8454
Wisconsin	0.0938	0.3436	-0.0347	0.7386	-0.2889	0.0000	0.0495	0.0250
Wyoming	-0.0530	0.5286	0.5072	0.0001	-0.1243	0.1116	0.0403	0.1657

Table 2.9: The estimation results for the 0.9-level quantile

States	$\alpha_1(s_i)$		$\lambda(s_i)$		$\beta(s_i)$		$\gamma(s_i)$	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Alabama	0.1330	0.0162	0.2023	0.0077	0.0784	0.4995	0.0228	0.3069
Arizona	-0.1222	0.1628	0.0046	0.9598	-0.1131	0.6131	0.0522	0.1434
Arkansas	0.0385	0.7385	0.2089	0.0819	-0.1323	0.0953	0.0728	0.0054
California	-0.0268	0.8540	-0.0436	0.8974	-0.2262	0.1312	0.0895	0.0828
Colorado	0.1369	0.1371	0.3460	0.0020	-0.1436	0.2104	0.0491	0.0422
Connecticut	0.1445	0.4037	0.3867	0.0364	-0.3058	0.0075	0.0119	0.8448
Delaware	0.1662	0.0903	0.2877	0.1374	-0.2237	0.2560	0.0764	0.1048
Florida	0.0980	0.1239	-0.0195	0.7964	0.2950	0.1368	0.0554	0.0327
Georgia	0.5685	0.0005	-0.5558	0.1772	-0.0669	0.6576	0.0495	0.5025
Idaho	0.0912	0.1004	0.1914	0.1137	0.1870	0.2078	0.0047	0.7933
Illinois	-0.0854	0.7364	-0.0740	0.8655	0.5816	0.0224	0.0528	0.5615
Indiana	0.2482	0.1989	0.2118	0.4316	-0.1484	0.2938	0.0373	0.4344
Iowa	-0.1904	0.2013	0.4494	0.0129	-0.3016	0.0021	-0.0885	0.0175
Kansas	0.1240	0.1530	0.6569	0.0000	-0.1688	0.1309	0.0129	0.3678
Kentucky	0.2501	0.0122	0.8734	0.0000	-0.1907	0.1754	0.0337	0.3823
Louisiana	-0.0811	0.4858	0.3712	0.0065	0.0749	0.5212	0.0117	0.7215
Maine	-0.0963	0.7614	-0.2578	0.2433	-0.0376	0.6921	0.0556	0.4640
Maryland	0.0917	0.4723	0.1695	0.4102	-0.1702	0.2196	0.0440	0.3188
Massachusetts	0.1637	0.2512	-0.0609	0.8065	-0.2320	0.0320	0.0892	0.1019
Michigan	0.1058	0.0443	0.1248	0.1952	0.4270	0.0000	0.0199	0.3071
Minnesota	0.3022	0.0388	0.0978	0.6410	-0.2612	0.0864	0.1236	0.0165
Mississippi	0.2070	0.2204	0.4144	0.0861	-0.5015	0.0003	0.0979	0.0755
Missouri	0.0057	0.9621	0.5989	0.0004	-0.0896	0.6572	-0.0350	0.2927
Montana	0.1018	0.2886	0.4829	0.0000	-0.3116	0.0991	0.0316	0.1745
Nebraska	-0.0252	0.7484	-0.1518	0.2672	0.0533	0.4038	0.0789	0.0040
Nevada	-0.0164	0.8613	0.1852	0.4354	0.1201	0.4375	0.0008	0.9815
New.Hampshire	0.0600	0.4660	-0.1226	0.2916	-0.2844	0.0133	0.0638	0.0340
New.Jersey	0.4243	0.0000	0.1962	0.5790	-0.2113	0.4189	0.0411	0.6741
New.Mexico	0.0573	0.7944	0.3765	0.1529	-0.0840	0.6172	-0.1151	0.0749
New.York	0.2479	0.0076	0.0220	0.9048	0.1372	0.4533	-0.0436	0.1982
North.Carolina	-0.0438	0.6703	0.3619	0.0199	-0.1193	0.3907	0.0609	0.0963
North.Dakota	0.1380	0.5196	0.1611	0.6417	-0.2048	0.3537	-0.0062	0.9300
Ohio	0.0776	0.0590	0.4413	0.0031	0.1422	0.1195	-0.0802	0.0059
Oklahoma	0.1037	0.5261	0.5630	0.0522	0.2593	0.0087	0.0258	0.7184
Oregon	0.0615	0.4317	-0.2528	0.0376	-0.3240	0.0200	0.0312	0.2368
Pennsylvania	0.0372	0.3538	0.2652	0.0300	0.0851	0.3955	0.0086	0.7204
Rhode.Island	0.1337	0.3651	-0.0306	0.8288	0.1794	0.1239	0.0135	0.7326
South.Carolina	-0.1071	0.5250	-0.0613	0.2600	0.0180	0.9231	0.1283	0.0304
South.Dakota	-0.1342	0.4623	0.2513	0.2709	-0.3430	0.0044	-0.0164	0.7326
Tennessee	0.1547	0.1689	0.6158	0.0044	-0.3265	0.0167	0.0331	0.3992
Texas	0.2617	0.0211	0.1128	0.4063	-0.3229	0.0007	0.0978	0.0001
Utah	0.0268	0.7962	0.3861	0.0141	-0.0486	0.6833	0.0393	0.3298
Vermont	-0.1869	0.1080	-0.0233	0.7716	-0.0609	0.5245	0.0005	0.9885
Virginia	0.1377	0.0769	0.3175	0.0079	-0.2448	0.0397	0.0602	0.0164
Washington	-0.0566	0.6226	0.3516	0.0695	-0.3005	0.0277	0.0081	0.8024
West.Virginia	0.1499	0.0751	0.6623	0.0000	-0.2540	0.1725	-0.0096	0.7802
Wisconsin	0.0373	0.8258	0.0487	0.8145	-0.4502	0.0016	0.1215	0.0398
Wyoming	-0.0674	0.6965	0.5786	0.0228	-0.2924	0.1468	0.0548	0.3489

Table 2.10: The Wald tests for the equality of slopes

State	0.1 vs 0.5	$\lambda(s_t)$ 0.1 vs 0.9	0.5 vs 0.9	0.1 vs 0.5	$\beta(s_t)$ 0.1 vs 0.9	0.5 vs 0.9	0.1 vs 0.5	$\alpha_1(s_t)$ 0.1 vs 0.9	0.5 vs 0.9	0.1 vs 0.5	$\gamma(s_t)$ 0.1 vs 0.9	0.5 vs 0.9
Alabama	0.041	0.033	0.371	0.779	0.903	0.909	0.058	0.434	0.396	0.531	0.743	0.297
Colorado	0.173	0.246	0.840	0.876	0.552	0.378	0.381	0.132	0.107	0.135	0.094	0.541
Connecticut	0.515	0.918	0.393	0.452	0.261	0.010	0.632	0.369	0.434	0.827	0.894	0.695
Georgia	0.278	0.518	0.222	0.843	0.372	0.377	0.015	0.849	0.038	0.132	0.536	0.849
New Mexico	0.428	0.724	0.780	0.394	1.000	0.456	0.616	0.757	0.451	0.208	0.009	0.025
Ohio	0.325	0.378	0.948	0.645	0.167	0.031	0.839	0.510	0.337	0.451	0.005	0.001
Oklahoma	0.343	0.301	0.645	0.420	0.644	0.603	0.209	0.583	0.434	0.735	0.801	0.976
Oregon	0.147	0.009	0.000	0.576	0.759	0.115	0.992	0.828	0.636	0.609	0.507	0.616
Pennsylvania	0.210	0.569	0.623	0.882	0.857	0.922	0.555	0.794	0.595	0.167	0.213	0.873
Tennessee	0.643	0.923	0.533	0.898	0.620	0.251	0.861	0.942	0.917	0.954	0.871	0.719
Texas	0.546	0.623	0.987	0.585	0.429	0.161	0.119	0.515	0.373	0.411	0.053	0.019
West Virginia	0.003	0.739	0.013	0.742	0.203	0.134	0.002	0.003	0.203	0.573	0.413	0.563
Wyoming	0.844	0.686	0.715	0.513	0.972	0.533	0.512	0.393	0.577	0.213	0.743	0.609

Note: This table displays the results of the Wald tests (p -value) for the equality of slopes (0.1 against 0.5 and 0.9 quantiles) in some typical states.

Chapter 3

Local Linear Quantile Regression for Time Series under Near Epoch Dependence

3.1 Introduction

Nonlinear modelling of time series data has drawn a lot of attention from researchers in the past few decades (c.f., [Tong, 1990](#); [Tjøstheim and Auestad, 1994](#); [Härdle et al., 1997](#); [Fan and Yao, 2003](#), etc). In these literatures, most modellings are based on conditional mean regression perspective, while fewer work has focused on nonlinear modelling by quantile regression for time series. In this chapter, we are mainly concerned with the nonlinear modelling of time series data from view point of quantile regression. Specifically, we consider a (strictly) stationary time series that are near-epoch dependent (NED) introduced by [Ibragimov \(1962\)](#).

Assume that $\{(Y_t, X_t)\}$ is a stationary multivariate time series on a probability space (Ω, \mathcal{F}, P) in general context, where X_t and Y_t are random variables taking their values in \mathbb{R}^p and \mathbb{R}^1 respectively. In time series econometrics, X_t may involve both the lags of endogenous and/or exogenous variables. We are here interested in the τ -th ($0 < \tau < 1$) conditional quantile function of Y_t given $X_t = x$.

$$q_\tau(x) = \arg \min_{a \in \mathbb{R}^1} E \{ \rho_\tau(Y_i - a) | X_i = x \}, \quad (3.1.1)$$

where $\rho_\tau(y) = y(\tau - I_{\{y < 0\}})$ with $y \in \mathbb{R}^1$ and I_A is the indicator function of set A . This *conditional quantile regression* was initially developed under *i.i.d.* samples for

linear regression models in econometric literature (Koenker and Bassett Jr, 1978, 1982).

In comparison with the mean regression, the modelling based on quantile regression has some essential advantages. Firstly, a well known special case with $\tau = 1/2$, i.e., median regression, is much explored and more robust than the mean regression when the data distribution is typically skewed or possesses a few outliers. Secondly, a heteroscedastic model can be easily detected if the regression quantiles of the model are not parallel (c.f., Efron, 1991). Thirdly, a collection of conditional quantiles can describe the whole distribution of the independent variable (c.f., Yu and Jones, 1998). Fourthly, pairs of extreme conditional quantiles can be used to depict the conditional prediction interval, which is important in econometric forecasting (see, Granger et al., 1989; Koenker and Zhao, 1996; Kuester et al., 2006, for example). Finally, the value-at-risk (VaR) has become a popular tool to measure market risk, which is just the quantile of the potential loss to be expected over a given future period (for a given probability) (c.f., Jorion, 1997). Therefore, the regression quantile would be helpful to factor analysis of the risk modelling based on VaR.

The asymptotic properties of kernel estimators for quantile regression have been investigated under *i.i.d.* or α -mixing conditions. To be specific, Yu and Jones (1998) used local linear fitting to estimate the quantile regression, Cai (2002) introduced an estimate of conditional quantile based on the inverse of conditional distribution function (Hall et al., 1999), and Honda (2000) estimate the quantile regression by local polynomial fitting. Recently, for time-varying coefficient models, Honda (2004) and Kim (2007) considered the nonparametric quantile estimation based on the *i.i.d.* errors term. Cai and Xu (2008) assumed that error is α -mixing process. In Wu and Zhou (2017), the errors are locally stationary processes with cross-dependent. Moreover, Cai and Xiao (2012) investigated the partially varying coefficients quantile regression models based on β -mixing assumption, and Wang et al. (2009) without any specification of the error distribution. Oberhofer and Haupt (2016) established the asymptotic properties of the nonlinear quantile regression, with allowing the error process to be heterogeneous and mixing. Differently, we will develop a theory for nonlinear modelling of the quantile regression function in nonparametric way for time series under NED condition. The main motivation for the NED condition is that α -mixing is a stricter assumption, which is hard to cover compound processes (Davidson, 1994; Lu and Linton, 2007). For example, Andrews (1984) has shown that a very simple linear AR(1) model of the

form

$$X_t = \frac{1}{2}X_{t-1} + e_t,$$

with e_t 's being independent symmetric Bernoulli random variables taking values -1 and 1 , is not α -mixing. Therefore, α -mixing may not be satisfied in the case of discrete innovations. Moreover, [Gorodetskii \(1978\)](#) has demonstrated that when the coefficients of the linear process do not drop off sufficiently fast, even the continuously distributed innovations may not satisfy the α -mixing property. For these reasons, the NED processes, which can cover more kinds of processes, have been considered extensively in econometrics ([Andrews, 1995](#); [Lu, 2001](#); [Gallant, 2009](#); [Jenish, 2012](#); [Li et al., 2012a](#)).

We will introduce a local linear kernel method to estimate both the quantile regression and its derivatives under NED condition, which is an extension of the work of [Welsh \(1996\)](#) and [Yu and Jones \(1998\)](#) under *i.i.d.* samples. In addition, we establish the asymptotic distribution for quantile regression estimators, which is important both for assessing the estimates and constructing an asymptotic confidence interval as well as testing hypothesis on the quantile regression, where a powerful tool of Bahadur representation will be established in convergence in probability for our aim. Our asymptotic normality result generalizes [Lu et al. \(1998\)](#), [Honda \(2000\)](#) and [Hallin et al. \(2009a\)](#) who obtained Bahadur representations with different conditions under alpha-mixing dependence to a more widely applicable family of data generating processes of near epoch dependence.

3.2 Local Linear Quantile Regression Estimators

3.2.1 Notation and main assumption

Supposing that \mathbb{R}^1 -valued Y_t and \mathbb{R}^p -valued X_t are stationary processes, we can define

$$Y_t = \Psi_Y(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \tag{3.2.1}$$

$$X_t = (X_{t1}, \dots, X_{tp})' = \Psi_X(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \tag{3.2.2}$$

where $\{\varepsilon_t\}$ is a stationary process, $'$ denotes the transpose. $\Psi_Y : \mathbb{R}^\infty \longrightarrow \mathbb{R}^1$ and $\Psi_X : \mathbb{R}^\infty \longrightarrow \mathbb{R}^p$ are two Borel measurable functions, respectively. $(X_1, Y_1), \dots, (X_n, Y_n)$ are a realization of n observations from this stationary sequence $\{(X_t, Y_t)\}$.

We consider the case of general p ($p \geq 1$) for the dimension of X . In practice, to avoid the curse of dimensionality, p should be chosen to be as small as possible through some variable selection procedure (c.f., [Tjøstheim and Auestad, 1994](#), in the mean regression setting). In what follows, we denote by $F_{Y|X}(x) = P(Y_t < y|X_t = x)$ and $f_X(y|x)$ the conditional distribution and the conditional density function of Y_t given $X_t = x$, respectively, and $f_X(x)$ will denote the marginal density function of X_t .

Definition 0. For a positive real number $\nu > 0$, if

$$v_\nu(m) = E|Y_t - Y_t^{(m)}|^\nu + E\|X_t - X_t^{(m)}\|^\nu \rightarrow 0 \quad (3.2.3)$$

as $m \rightarrow \infty$, where $|\cdot|$ and $\|\cdot\|$ are the absolute value and the Euclidean norm of \mathbb{R}^p , respectively. $Y_t^{(m)} = \Psi_{Y,m}(\varepsilon_t, \dots, \varepsilon_{t-m+1})$, $X_t^{(m)} = (X_{t1}^{(m)}, \dots, X_{tp}^{(m)})^\tau = \Psi_{X,m}(\varepsilon_t, \dots, \varepsilon_{t-m+1})$, where $\Psi_{Y,m}$ and $\Psi_{X,m}$ are \mathbb{R}^1 - and \mathbb{R}^p -valued Borel measurable functions with m arguments, respectively. We will call $\{(Y_t, X_t)\}$ the *near epoch dependent in L_ν norm* w.r.t. a stationary α -mixing process $\{\varepsilon_t\}$, and $v_\nu(m)$ the *stability coefficients of order ν* of the process $\{(Y_t, X_t)\}$.

Definition 1. For a stationary sequence $\{\varepsilon_t, t = 0, \pm 1, \dots\}$, if

$$\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{n+k}^\infty} |P(AB) - P(A)P(B)| \rightarrow 0$$

as $k \rightarrow \infty$, where $\mathcal{F}_{-\infty}^n$ and \mathcal{F}_{n+k}^∞ are the σ -fields generated by $\{\varepsilon_t, t \leq n\}$ and $\{\varepsilon_t, t \geq n+k\}$, respectively. We call ε_t the α -mixing process.

Clearly, $\{(Y_t^{(m)}, X_t^{(m)})\}$ is an α -mixing process with mixing coefficient

$$\alpha_m(k) \leq \begin{cases} \alpha(k-m) & k \geq m+1, \\ 1 & k \leq m. \end{cases} \quad (3.2.4)$$

For the sake of convenience, we will provide the main assumptions on the data-generating process (**A1**), the kernel function (**A2-A3**) and the bandwidth (**B1-B4**) in the following, which will be used in later lemmata and theorems.

A1: (i) The data generating process (DGP) is a strictly stationary NED process, with order $\nu = 2 + \delta/2$, with respect to the α -mixing process $\{\varepsilon_t\}$, where the constant $\delta > 0$ is satisfied in $E[|Y_t|^{1+\delta}] < \infty$ $\lim_{k \rightarrow \infty} k^a \sum_{j=k}^\infty \alpha(j) = 0$, for some positive real number a .

(ii) $f_X(\cdot)$ denotes the continuous density function of X_i and satisfies $f_X(x) > 0$ at x .

(iii) $f_{Y|X}(y|\tilde{x})$ is a conditional density function of Y , which is continuous in a neighbourhood of $q_\tau(x)$ uniformly for \tilde{x} in a neighbourhood of x , and continuous as a function of \tilde{x} in a neighbourhood of x for all y in a neighbourhood of $q_\tau(x)$. Also, $f_{Y|X}(q_\tau(x)|x) > 0$.

(iv) The joint density (X_i, X_j) at (x, \tilde{x}) is satisfying

$$\sup_{i,j} \sup_{(x,\tilde{x})} f_{i,j}(x, \tilde{x}) \leq C, \quad (3.2.5)$$

where C is a generic positive constant, and i and j are in \mathbb{Z} .

A2. (i) For kernel function $K : \mathbb{R}^p \rightarrow \mathbb{R}$, we define

$$K_c(u) := (c_0 + c_1' u)K(u). \quad (3.2.6)$$

where $\mathbf{c} := (c_0, c_1')' \in \mathbb{R}^{p+1}$. $|K_c(u)|$ is uniformly bounded by some constant K_c^+ , and is integrable: $\int_{\mathbb{R}^p} |K_c(x)| dx < \infty$.

(ii) $Q_c^K(x) := \sup_{\|y\| \geq \|x\|} [\|y\|^2 K_c(y)]$ is integrable.

(iii) The kernel function $K(\cdot)$ is a bounded density function such that $\int uK(u)du = 0$ and $\mu_2 = \int uu'K(u)du > 0$ (positive definite).

(iv) For some constant $C > 0$, $|K_c(u) - K_c(v)| \leq C\|u - v\|$ for any $u, v \in \mathbb{R}^p$.

A3. The quantile function $q_\tau(\cdot)$ is twice continuously differentiable. The v -th order derivative $q_\tau^{(v)}(\cdot)$ of the quantile function $q_\tau(\cdot)$ satisfies the Lipschitz condition of degree δ ($0 < \delta \leq 1$), such that

$$\|q_\tau^{(v)}(x) - q_\tau^{(v)}(\tilde{x})\| \leq C\|x - \tilde{x}\|^\delta, \text{ for any } x, \tilde{x} \in \mathbb{R}^p, \quad (3.2.7)$$

where $q_\tau^{(0)}(x) = q_\tau(x)$, $q_\tau^{(1)}(x) = \dot{q}_\tau(x)$, and $\|\cdot\|$ is the Euclidean norm.

B1.

$$h_n \rightarrow 0 \quad \text{and} \quad nh_n^{(1+2/a)p} \rightarrow \infty, \quad \text{as } n \rightarrow \infty, \quad (3.2.8)$$

where a is the positive constant defined in **A1**(i).

B2. There is two positive integers $m = m_n \rightarrow \infty$ and $L = L_n = \sqrt{v_1(m)} \rightarrow 0$, satisfying

$$h_n^{-1-p}v_1(m) \rightarrow 0, \quad n^6v_1(m) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (3.2.9)$$

B3. For two positive integer sequences, $p^* = p_n^* \in \mathbb{Z}$ and $q = q_n = 2m_n \in \mathbb{Z}$, with $m = m_n \rightarrow \infty$ such that $p^* = p_n^* = o((nh_n^p)^{1/2})$, $q/p^* \rightarrow 0$ and $n/p^* \rightarrow \infty$, and $\frac{n}{p^*}\alpha(m) \rightarrow 0$.

B4. h_n tends to zero in such a manner that $qh_n^p = O(1)$ such that

$$h_n^p \sum_{t=q}^{\infty} \alpha_m(t) \rightarrow 0 \quad \text{as } T \rightarrow \infty.$$

Remark. Assumption **B1** is the common used in *i.i.d.* case. Assumptions **B2-B4** restrict the decay rates of the stability and mixing coefficients for a given bandwidth.

3.2.2 Local linear fitting of quantile regression function

Denote by $\dot{q}_\tau(x) = (\partial q_\tau(x)/\partial x_1, \dots, \partial q_\tau(x)/\partial x_p)'$ the vector of the first order partial derivatives of $q_\tau(x)$ at $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$, where x' denotes the transpose of a vector x . The idea of the local linear fitting ([Fan and Gijbels, 1996](#); [Loader, 1999](#)) is to approximate the unknown function $q_\tau(\cdot)$ by a linear function

$$q_\tau(z) \approx q_\tau(x) + (\dot{q}_\tau(x))'(z - x) \equiv a_0 + a_1'(z - x), \quad (3.2.10)$$

for z in a neighbourhood of x . Locally, estimating $(q_\tau(x), \dot{q}_\tau(x))$ is equivalent to estimating (a_0, a_1) . This motivates us to define an estimator by setting $\hat{q}_\tau(x) \equiv \hat{a}_0$ and $\hat{\dot{q}}_\tau(x) \equiv \hat{a}_1$. Then

$$(\hat{a}_0, \hat{a}_1) = \arg \min_{(a_0, a_1)} \sum_{i=1}^n \rho_\tau(Y_i - a_0 - a_1'(X_i - x))K_h(X_i - x), \quad (3.2.11)$$

where $K_h(x) = h_n^{-p}K(x/h_n)$, K is a kernel function on \mathbb{R}^p , and $h_n > 0$ is the bandwidth. Note that if $q_\tau(x)$ has no derivative, $q_\tau(x)$ can still be estimated by (3.2.11), but the estimate of its derivative is vacuous. (See also [Lemma 3.2](#))

The regression coefficients in model (3.2.11) are often of interest in econometrics. For example, in the nonparametric option pricing model introduced by Aït-Sahalia and Lo (1998), if the modelling is carried out based on quantile regression method, then a_0 would estimate the quantile of the option price and a_1 estimate the related Greeks. Another example of potential interest in financial econometrics is the value-at-risk modelling of the risk factors, where a_0 would estimate the quantity of value-at-risk given the factor variables and a_1 estimate the influence of the changes of the factor variables on the risk. These interesting topics will be developed in other work. In this chapter, we are concerned with the theoretical property of the estimators.

3.3 Asymptotic Behaviours

The main interest in this section is to establish the asymptotic distribution of the quantile regression estimators under near epoch dependence. To this purpose, we need to develop a powerful tool of Bahadur representation in weak convergence sense for the quantile regression estimators, which are not of analytical expression as the mean regression.

3.3.1 Bahadur representation

In this subsection, we consider the weak conditions to ensure the Bahadur representation of the local linear estimators of $q_\tau(\cdot)$ and its derivatives. If $q_\tau(x)$ is first order differentiable, then its derivatives can be estimated reasonably well by the local linear fitting. We first introduce a notation. Denote by $Z_n = o_P(1)$ if the random sequence $Z_n \xrightarrow{P} 0$, where \xrightarrow{P} denotes convergence in probability.

Theorem 3.1. *Assume that Assumptions A1 –B4 are satisfied for some $a \geq 1$, and that the quantile function $q_\tau(x)$ is twice continuously differentiable at x . Then*

$$\sqrt{nh_n^p} \begin{pmatrix} \hat{q}_\tau(x) - q_\tau(x) \\ h_n (\hat{\dot{q}}_\tau(x) - \dot{q}_\tau(x)) \end{pmatrix} = \phi_\tau(x) \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^n \psi_\tau(Y_i^*) \begin{pmatrix} 1 \\ \frac{X_i - x}{h_n} \end{pmatrix} K\left(\frac{X_i - x}{h_n}\right) + o_P(1), \quad (3.3.1)$$

as $n \rightarrow \infty$, where $\psi_\tau(y) = \tau - I_{\{y < 0\}}$, $Y_i^* = Y_i^*(\tau) = Y_i - q_\tau(x) - (\dot{q}_\tau(x))'(X_i - x)$ and $\phi_\tau(x) = (f_{Y|X}(q_\tau(x)|x)f_X(x))^{-1}$.

The next theorem shows that even if $q_\tau(x)$ is not differentiable, it may still be estimated reasonably well by the local linear fitting of (3.2.11), but the estimate of $\dot{q}_\tau(x)$ is vacuous. This is similar to Remark 2(b) of Lu and Gijbels (2001) in mean regression setting.

Theorem 3.2. *Under the Assumptions in Theorem 3.1, and the quantile function $q_\tau(x)$ is continuous at x . Then*

$$\sqrt{nh_n^p}(\hat{q}_\tau(x) - q_\tau(x)) = \phi_\tau(x) \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^n \psi_\tau(Y_i^*) K\left(\frac{X_i - x}{h_n}\right) + o_P(1), \quad (3.3.2)$$

as $n \rightarrow \infty$, where $Y_i^* = Y_i^*(\tau) = Y_i - q_\tau(x)$.

The proofs of Theorem 3.1 and 3.2 are given in Section 3.6.

3.3.2 Asymptotic normality

Based on the powerful tool of the weak Bahadur representation, we can establish the convergence rates and the asymptotic distribution of the local linear quantile regression estimates under near epoch dependence. If $q_\tau(x)$ has the first order derivatives which are Lipschitz continuous, then $q_\tau(x)$ and its derivatives can be estimated with optimal convergence rates of Stone (1980) as in the *i.i.d.* setting.

Theorem 3.3. *Under the conditions of Theorem 3.1, if Assumption A3 with $v = 1$ is satisfied, then*

$$\hat{q}_\tau(x) - q_\tau(x) = O_P(h_n^{1+\delta}) + O_P((nh_n^p)^{-1/2}), \quad (3.3.3)$$

and

$$\hat{\dot{q}}_\tau(x) - \dot{q}_\tau(x) = O_P(h_n^\delta) + O_P((nh_n^{p+2})^{-1/2}). \quad (3.3.4)$$

Furthermore, if in A1(i) $a > p/(v + \delta)$ and $h_n = n^{-1/[p+2(1+\delta)]}$, then

$$\hat{q}_\tau(x) - q_\tau(x) = O_P(n^{-(1+\delta)/[p+2(1+\delta)]}), \quad (3.3.5)$$

$$\hat{\dot{q}}_\tau(x) - \dot{q}_\tau(x) = O_P(n^{-\delta/[p+2(1+\delta)]}) \quad (3.3.6)$$

as $n \rightarrow \infty$.

Remark. The condition $a > p/(v + \delta)$ in the theorem is to ensure that the requirement in Assumption **B1** can be satisfied with the optimal bandwidth, $h_n = n^{-1/[p+2(1+\delta)]}$.

If $q_\tau(x)$ is not differentiable but Lipschitz continuous, then the estimator $\hat{q}_\tau(x)$ can converge to $q_\tau(x)$ at the rate specified as follows.

Theorem 3.4. *If Assumption **A3** with $v = 0$ is satisfied in Theorem 3.2, then*

$$\hat{q}_\tau(x) - q_\tau(x) = O_P(h_n^\delta) + O_P((nh_n^p)^{-1/2}). \quad (3.3.7)$$

Furthermore, if in **A1(i)** $a > p/\delta$ and $h_n = n^{-1/[p+2\delta]}$, then

$$\hat{q}_\tau(x) - q_\tau(x) = O_P(n^{-\delta/[p+2\delta]}) \quad (3.3.8)$$

as $n \rightarrow \infty$.

Remark. According to [Stone \(1980\)](#), the convergence rates in equation 3.3.3-3.3.8 are also optimal for *i.i.d.* observations.

The asymptotic distribution of the estimators can be established under near epoch dependence. And the asymptotic normality of our estimators relies on the following lemmata. Suppose

$$W_n := \begin{pmatrix} w_{n0} \\ w_{n1} \end{pmatrix}, \quad (W_n)_j := (nh_n^p)^{-1} \sum_{i=1}^n \psi_\tau(Y_i^*) \left(\frac{X_i - x}{h_n} \right)_j K \left(\frac{X_i - x}{h_n} \right),$$

$$j = 0, \dots, p, \text{ with } \left(\frac{X_i - x}{h_n} \right)_0 = 1.$$

The usual Cramér-Wold device will be adopted. For all $c := (c_0, c_1)' \in \mathbb{R}^{1+p}$, let

$$A_n := (nh_n^p)^{1/2} c' W_n = \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^n \psi_\tau(Y_i^*) K_c \left(\frac{X_i - x}{h_n} \right),$$

with $K_c(u)$ defined in (3.2.6). The following lemma provides the expectation and asymptotic variance of A_T for all c , hence that of $(nh_n^p)^{1/2} W_n$.

Lemma 3.1. *Assume that Assumptions **A1** and **A2** hold, and that h_n satisfies Assumptions **B1** and **B2**. The expectation is as*

$$E[\phi_\tau(x)A_n] = \sqrt{nh_n^p} \left[(1 + o(1)) \begin{pmatrix} B_0(x) \\ B_1(x) \end{pmatrix} \right],$$

where $B_0(x) = \frac{1}{2}tr[\ddot{q}_\tau(x) \int uu'K(u)du]h_n^2$, and $B_1(x) = (B_{11}(x), \dots, B_{1p}(x))'$, $B_{1j}(x) = \frac{1}{2}tr[\ddot{q}_\tau(x) \int uu'u_jK(u)du]h_n$, $j = 1, \dots, p$. The asymptotic variance is as

$$\lim_{n \rightarrow \infty} \text{Var}[\phi_\tau(x)A_n] = c'\Sigma c,$$

where

$$\Sigma := \phi_\tau^2(x)\tau(1-\tau)f_X(x) \begin{pmatrix} \int K^2(u)du & \int u'K^2(u)du \\ \int uK^2(u)du & \int uu'K^2(u)du \end{pmatrix}$$

Lemma 3.2. *Suppose that Assumptions **A1** and **A2** hold, and that the bandwidth h_n satisfies Assumptions **B1**–**B4**. Denote by σ^2 the asymptotic variance of A_T . Then $(nh_n^p)^{1/2}(c'[W_n(x) - EW_n(x)]/\sigma)$ is asymptotically standard normal as $n \rightarrow \infty$.*

The proof of Lemma 3.1–3.2 will be given in Section 3.6. We now turn to the main consistency and asymptotic normality result.

Theorem 3.5. *If Assumption **A3** with $v = 2$ is satisfied, $a > p/(2 + \delta)$ in Assumption **A1**(i) and $nh_n^{p+2(2+\delta)} \rightarrow 0$, then for any $0 < \tau_1, \tau_2 < 1$,*

$$\sqrt{nh_n^p} \begin{pmatrix} \hat{q}_{\tau_1}(x) - q_{\tau_1}(x) - B_{01}(x) \\ \hat{q}_{\tau_2}(x) - q_{\tau_2}(x) - B_{02}(x) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \begin{pmatrix} \sigma_{000}^2(x) & \sigma_{001}^2(x) \\ \sigma_{010}^2(x) & \sigma_{011}^2(x) \end{pmatrix} \right), \quad (3.3.9)$$

$$\sqrt{nh_n^{p+2}} \begin{pmatrix} \hat{\dot{q}}_{\tau_1}(x) - \dot{q}_{\tau_1}(x) - B_{11}(x) \\ \hat{\dot{q}}_{\tau_2}(x) - \dot{q}_{\tau_2}(x) - B_{12}(x) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \begin{pmatrix} \sigma_{100}^2(x) & \sigma_{101}^2(x) \\ \sigma_{110}^2(x) & \sigma_{111}^2(x) \end{pmatrix} \right), \quad (3.3.10)$$

as $n \rightarrow \infty$. Here $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution,

$B_{0i}(x) = \frac{1}{2}tr[\ddot{q}_{\tau_i}(x) \int uu'K(u)du]h_n^2$, and $B_{1i}(x) = (B_{1i1}(x), \dots, B_{1ip}(x))'$, $B_{1ij}(x) = \frac{1}{2}tr[\ddot{q}_{\tau_i}(x) \int uu'u_jK(u)du]h_n$, $i = 1, 2, j = 1, \dots, p$.

$$\sigma_{000}^2(x) = \phi_1^*(x) \int K^2(u)du, \quad \sigma_{011}^2(x) = \phi_2^*(x) \int K^2(u)du,$$

$$\sigma_{001}^2(x) = \sigma_{010}^2(x) = \phi_3^*(x) \int K^2(u)du, \quad \sigma_{100}^2(x) = \phi_1^*(x) \int uu'K^2(u)du,$$

$$\sigma_{111}^2(x) = \phi_2^*(x) \int uu' K^2(u) du, \quad \sigma_{101}^2(x) = \sigma_{110}^2(x) = \phi_3^*(x) \int uu' K^2(u) du,$$

and where

$$\phi_i^*(x) = \phi_{\tau_i}^2(x) \tau_i (1 - \tau_i) f_X(x) = \frac{\tau_i (1 - \tau_i)}{f_X(x) f_{Y|X}^2(q_{\tau_i}(x)|x)}, \quad i = 1, 2$$

$$\begin{aligned} \phi_3^*(x) &= \phi_{\tau_1}(x) \phi_{\tau_2}(x) (\min(\tau_1, \tau_2) - \tau_1 \tau_2) f_X(x) \\ &= \frac{\min(\tau_1, \tau_2) - \tau_1 \tau_2}{f_X(x) f_{Y|X}(q_{\tau_1}(x)|x) f_{Y|X}(q_{\tau_2}(x)|x)} \end{aligned}$$

Remark. When the kernel function $K(\cdot)$ is symmetric, we can have $B_{1i}(x) = 0$, $i = 1, 2$. Then (3.3.10) can be rewritten as

$$\sqrt{nh_n^{p+2}} \begin{pmatrix} \hat{q}_{\tau_1}(x) - \dot{q}_{\tau_1}(x) \\ \hat{q}_{\tau_2}(x) - \dot{q}_{\tau_2}(x) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \begin{pmatrix} \sigma_{100}^2(x) & \sigma_{101}^2(x) \\ \sigma_{110}^2(x) & \sigma_{111}^2(x) \end{pmatrix} \right). \quad (3.3.11)$$

If $q_\tau(x)$ is not differentiable but Lipschitz continuous, one can obtain the asymptotic normality result for $q_\tau(x)$ as follows.

Theorem 3.6. *If Assumption A3 with $v = 0$ is satisfied, $a > p/\delta$ in Assumption A1(i) and $nh_n^{p+2\delta} \rightarrow 0$, then*

$$\sqrt{nh_n^p} \begin{pmatrix} \hat{q}_{\tau_1}(x) - q_{\tau_1}(x) \\ \hat{q}_{\tau_2}(x) - q_{\tau_2}(x) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \begin{pmatrix} \sigma_{000}^2(x) & \sigma_{001}^2(x) \\ \sigma_{010}^2(x) & \sigma_{011}^2(x) \end{pmatrix} \right), \quad (3.3.12)$$

as $n \rightarrow \infty$

Remark. The proofs of Theorems 3.3-3.6 are relegated to the Section 3.6.

3.4 Numerical Results

3.4.1 Simulation

In this section, we will consider the Monte Carlo simulation to study the finite sample performance of median regression and mean regression under near epoch dependence. The purpose is to illustrate that local linear median regression can

work more efficient and robust for the processes under near epoch dependence. We considered the following model(Reference to [Lee, 2003](#)):

$$Y_t = g(X_t) + a_3 \xi_t, \quad (3.4.1)$$

where ξ_t was drawn from the Student t -distribution with the 2 degrees of freedom (denoted by t_2) and $a_3 = 0.2$. The Student t -distribution is of great interest in financial modelling of market volatility(c.f., [Bollerslev et al., 1992](#)).In particular the variance of the time series may not exist if the ξ_t is heavily tailed enough (e.g., t_2 distribution).

$$g(x) = a_0 x + a_1 \sin(a_2 x), \quad (3.4.2)$$

where $a_0 = a_1 = 1, a_2 = 4$. And X_t was drawn from the ARMA(1,1)-GARCH(1,1) model, which is as follows:

$$X_t = \mu + \phi X_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t, \quad (3.4.3)$$

$$\varepsilon_t = e_t h_t^{1/2}, \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \quad (3.4.4)$$

where $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$, and e_t is *i.i.d.* $N(0, 1)$ random sequence Referring to the estimations of the model (3.4.3) with (3.4.4) in section 3.4.2, we take the parameters $\mu = 0.011223, \phi = 0.975815, \theta = -1, \alpha_0 = 0.234255, \alpha_1 = 0.106487$, and $\beta_1 = 0.861989$, which are obtained from the real oil price data by ARMA-GARCH module in R. [Lu and Linton \(2007\)](#) proofed that X_t is a stationary NED of order 1 w.r.t. a strongly (α -) mixing process.

We then generate data X_t from the model (3.4.3) and (3.4.4), and Y_t from model (3.4.1) with (3.4.2), denoted as $\{(X_t, Y_t)\}$ for $t = 1, \dots, n$. We consider two time series of sample size: $n = 100$ and $n = 500$. We are assessing the estimate of the median regression $q_{0.5}(x) = \operatorname{argmin}_{a \in R^1} E\{\rho_{0.5}(Y_i - a) | X_i = x\}$ and mean regression $m(x) = E(Y_t | X_t = x)$, which are both equal to $g(x)$, we partition 41 points of x between the -0.8 and 0.8, which are approximately 10th and 90th percentiles of X_t . In the simulations, the bandwidth b_T was chosen by the conventional cross-validation rule of [Stone \(1984\)](#).

Figure 3.1 shows that the biases of 100 median and mean estimates of $g(\cdot)$ for sample size 100 and 500 are acceptable. Figure 3.2 displays the results of local linear estimators of median regression $q_{0.5}(x) = g(x)$ and mean regression function $m(x) = g(x)$ in 41 points of x , based on 100 replications with each sample size

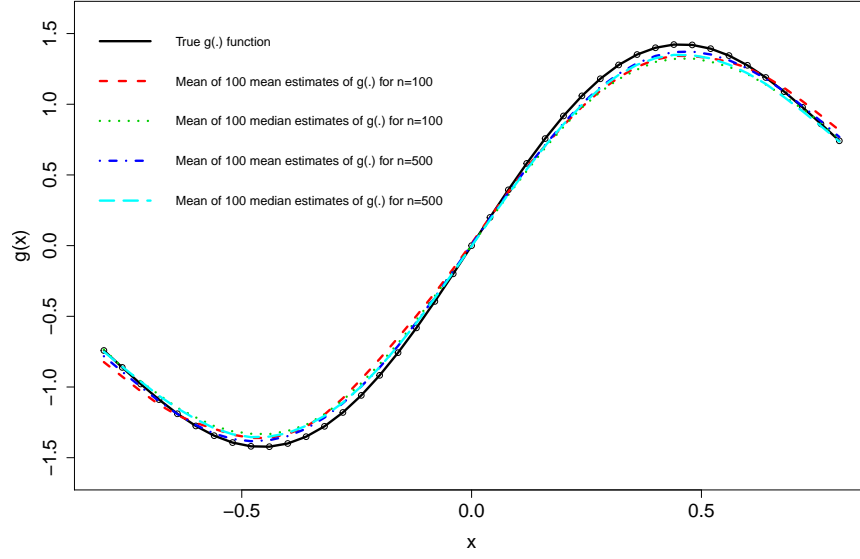


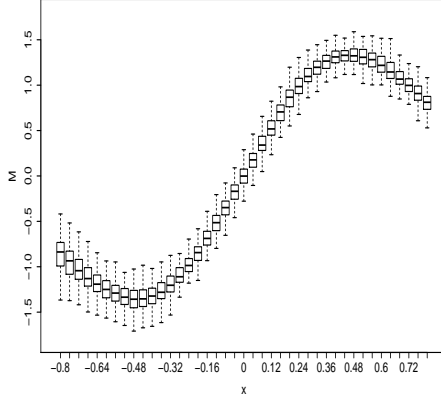
Figure 3.1: Comparison of mean of 100 median and mean estimates of $g(\cdot)$ for sample size 100 and 500

$n = 100$ and $n = 500$. Figure 3.3 assesses the accuracy of estimation by defining a squared estimation error (SEE) of conditional mean regression and median regression estimation of $g(\cdot)$

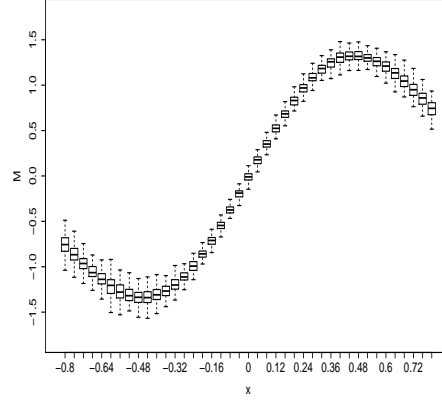
$$SEE(\hat{q}_{0.5}(\cdot)) = \frac{1}{41} \sum_{k=1}^{41} \left\{ \hat{q}_{0.5}^{(r)}(x_k) - g(x_k) \right\}^2$$

$$SEE(\hat{m}(\cdot)) = \frac{1}{41} \sum_{k=1}^{41} \left\{ \hat{m}^{(r)}(x_k) - g(x_k) \right\}^2$$

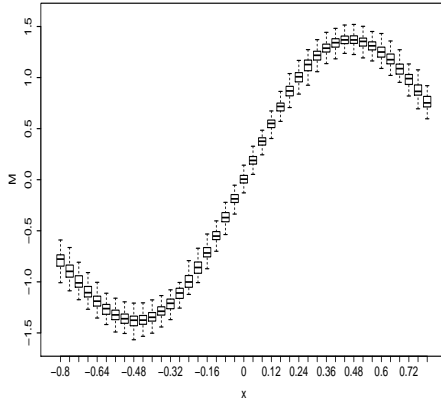
where $r = 1, \dots, 100$ is the simulation times, for the sample size $n = 100$ and $n = 500$. Overall, the simulation results of Model (3.4.1) adapt very well to asymptotic theory: with the sample size increasing, both the conditional mean and median regressions with a cross validation method for bandwidth selection become more stable and fit better to the true $g(\cdot)$ function. Even for the sample size of 100, the estimate procedure and bandwidth selection looks acceptable with the median regression. Clearly, the median regression has better performance in estimated results than the mean regression, and the median regression in sample size 500 works very well in all cases.



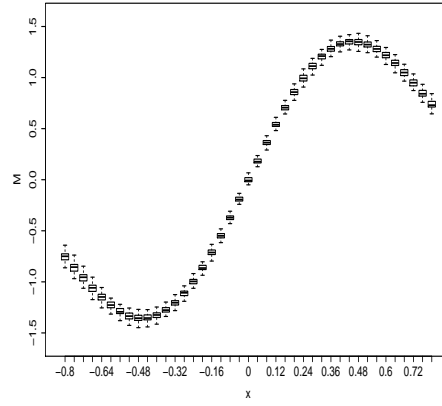
(a) mean regression, $n = 100$



(b) median regression, $n = 100$



(c) mean regression, $n = 500$



(d) median regression, $n = 500$

Figure 3.2: Simulation results-Boxplots of the local linear fitting for the median regression and mean regression, for $n = 100$ (top) and $n = 500$ (bottom)

3.4.2 An empirical application

Climate change has become one of the world's supreme policy challenges. In order to promote greener growth, the economists suggested providing an appropriate design of a price on carbon emissions. In 2005, the European Union Emissions Trading System (EU ETS) was established, which is the biggest greenhouse gas emissions trading scheme around world.

It is well known that crude oil is one of the biggest sources of world greenhouse gases. Low oil prices could restrain the development in cleaner energy technologies, which lead to higher carbon emissions ([Balaguer and Cantavella, 2016](#)). Through

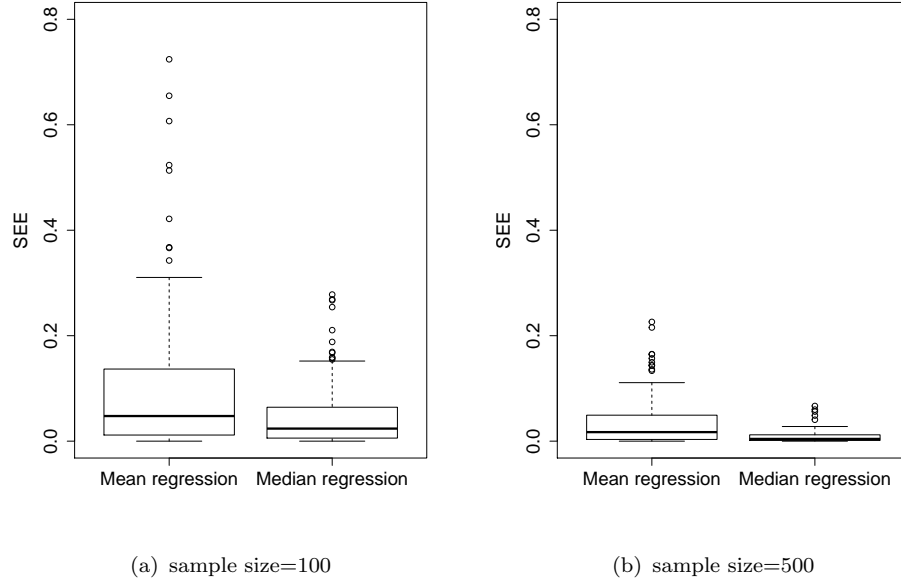


Figure 3.3: Simulation results-Boxplots of squared estimation error(SEE) of mean regression and median regression estimation of $g(\cdot)$, for $n = 100$ (left) and $n = 500$ (right)

corrective carbon pricing, governments could restore appropriate price incentives, and mitigate climate change. Correspondingly, the interaction between carbon price and crude oil price has been increasingly closer, due to the rapid development and steady expansion of carbon market. The relationship oil price and carbon futures price has drawn attention in many recent studies ([Mansanet-Bataller et al., 2007](#); [Chevallier, 2011](#); [Benz and Trück, 2009](#)). To capture well the underlying impact of crude oil futures price on the carbon futures price, we investigate the impact of daily return of WTI crude oil futures price on carbon futures price return¹, with sample size 454 from 27th July 2015 to 15th May 2017, for an illustration.

Figure 3.4(a) and 3.4(b) show WTI crude oil futures price (W_t) and carbon futures price (C_t), and the daily return of WTI crude oil futures price X_t and daily return of carbon futures price Y_t , defined by

$$X_t = \log(W_t/W_{t-1}) \times 100, \quad Y_t = \log(C_t/C_{t-1}) \times 100,$$

which are plotted in Figure 3.4(c) and 3.4(d), respectively.

¹Historical Futures Prices: ECX EUA Futures, Continuous Contract # 1. Non-adjusted price based on spot-month continuous contract calculations. Raw data from ICE.

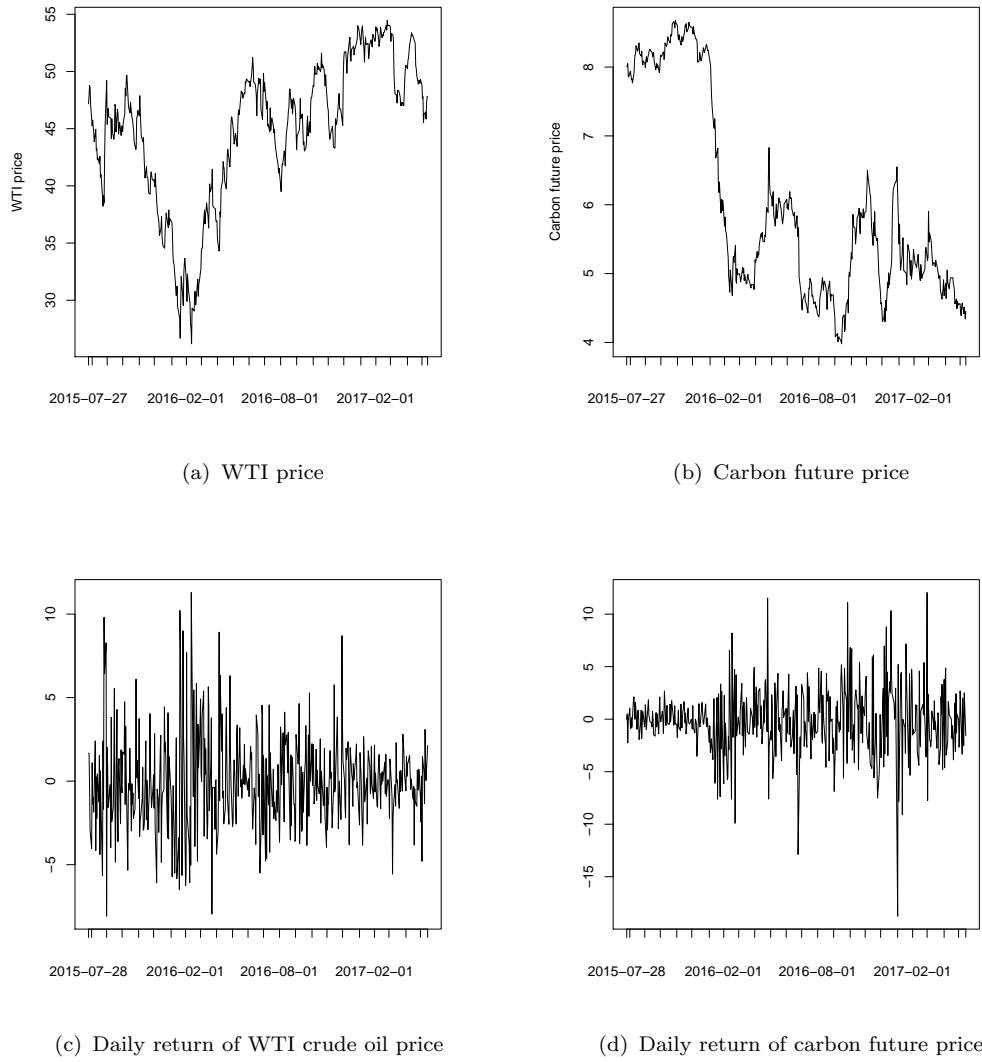


Figure 3.4: WTI crude oil and carbon future daily prices

Table 3.1: Estimation results of ARMA-GARCH

Coefficients	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
μ	0.011223	0.015269	0.73501	0.462332
ϕ	0.975815	0.004564	213.813	0.000000
θ	-1.000000	0.000457	-2188.80	0.000000
α_0	0.234255	0.107747	2.17411	0.029697
α_1	0.106487	0.035673	2.98508	0.002835
β_1	0.861989	0.035991	23.9501	0.000000

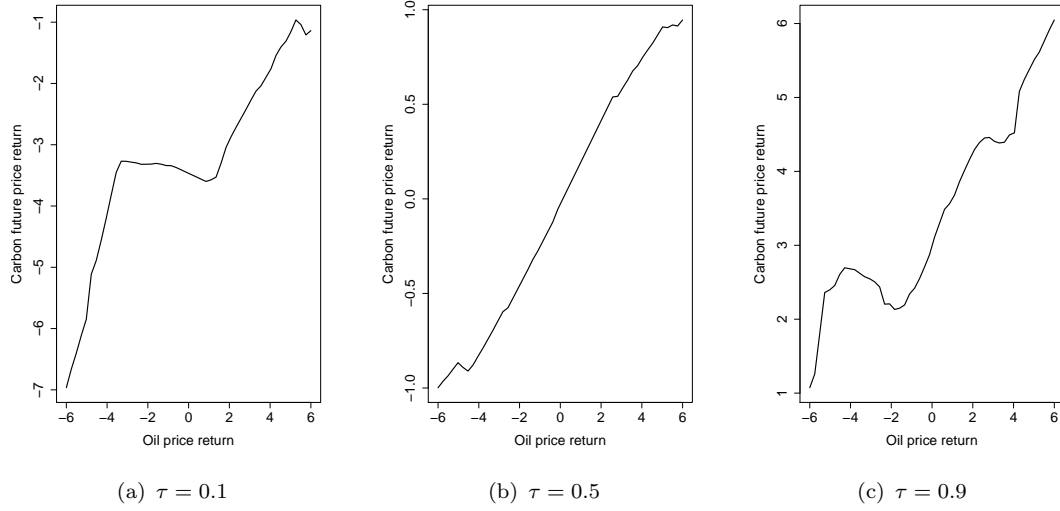


Figure 3.5: Nonparametric quantile regression for the relationship between daily return of WTI crude oil price and carbon future price, from 27th July 2015 to 15th May 2017 with 454 observations.

ARMA-GARCH models have been used to describe the oil price return in many studies (Chang et al., 2010; Lee and Chiou, 2011; Zhang and Chen, 2011, 2014). Therefore, we used the daily data of WTI crude oil price return to estimate the model (3.4.3) with (3.4.4) and check whether the conditional variance follows the GARCH process. The estimation results show in Table 3.1, which all the coefficients are statistically significant at 5% significance level except the intercept.

Table 3.2: Results of quantile prediction error for three methods

Methods	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
parametric quantile regression	0.3937184	0.97913	0.4405764
local linear quantile regression	0.3806771	0.98394	0.4240236
nonlinear threshold quantile regression	0.3511665	0.97913	0.4210472

The leave-one-out cross-validation (Li and Racine, 2004) is used to choose the optimal bandwidths, which are 1.30 ($\tau = 0.9$), 3.03 ($\tau = 0.5$), and 1.33 ($\tau = 0.1$) for the conditional quantiles, respectively. Figure 3.5 shows the local linear estimates of the conditional quantile regression at 0.1, 0.5 and 0.9 quantile levels, respectively. From Figure 3.5, we can find that both the upper and lower quantiles (e.g. $\tau = 0.1$ and $\tau = 0.9$) functions appear to be nonlinear.

To evaluate the local linear quantile method, we will compare the prediction of different parametric forms of quantile regression. The first is a linear quantile

function $Q_{Y_t}^0(\tau|X_t) = a_\tau^0 + \alpha_\tau^0 X_t$, where a_τ^0 and α_τ^0 are linear quantile coefficients (Koenker, 2005). In general, nonparametric specification can help to explore the possibly nonlinear relationship, but it may give poor prediction. Therefore, we consider a parametric nonlinear threshold quantile model with the thresholds given based on the nonparametric estimation results (Figure 3.5). Then, we construct the nonlinear threshold functions as follow.

- For $\tau = 0.1$

$$a(x) = \{a_{10} + a_{11}x\} \mathcal{I}(x < -3) + \{a_{20} + a_{21}x\} \mathcal{I}(-3 \leq x < 1.5) \\ + \{a_{30} + a_{31}x\} \mathcal{I}(1.5 \leq x)$$

- For $\tau = 0.5$

$$a(x) = a_{10} + a_{11}x$$

- For $\tau = 0.9$

$$a(x) = \{a_{10} + a_{11}x\} \mathcal{I}(x < -4) + \{a_{20} + a_{21}x\} \mathcal{I}(-4 \leq x < -2) \\ + \{a_{30} + a_{31}x\} \mathcal{I}(-2 \leq x < 2) + \{a_{40} + a_{41}x\} \mathcal{I}(2 \leq x < 4) \\ + \{a_{50} + a_{51}x\} \mathcal{I}(4 \leq x)$$

where $I(\cdot)$ is an indicator function, $a_{ij}(x)$, $i = 1, 2, \dots, 5$ and $j = 0, 1$, are the linear coefficients.

We use the first $T = 404$ as estimation sample to predict the last 50 quarters. A quantile prediction error (QPE) of prediction result is computed for the linear quantile regression and nonlinear threshold quantile regression under different quantiles. Here, we define QPE as:

$$QPE(\hat{q}_\tau(\cdot)) = \frac{1}{n} \sum_{i=1}^n \rho_\tau(\hat{q}_\tau(x_i) - q_\tau(x_i)), \quad (3.4.5)$$

where $\rho_\tau(y) = y(\tau - I_{\{y < 0\}})$ and I_A is the indicator function of set A . The results show in Table 3.2. The QPE values are 0.3937184 and 0.3511665 for linear and nonlinear threshold quantile regression in 10% quantile, respectively. Compared with the linear quantile function, the threshold quantile regression outperforms

in prediction, with a relative improvement of 11%. This result further illustrates that local linear quantile regression can help to uncover the impact of WTI crude oil price return on carbon futures price return, which is more complex than linear.

3.5 Conclusions

This chapter develops a theory for nonlinear modelling of the quantile regression function in nonparametric way for econometric time series under NED condition, which can extensively cover more kinds of processes in econometrics. We introduce a local linear kernel method to estimate both the quantile regression and its derivatives under NED condition, which extends the work of [Welsh \(1996\)](#) and [Yu and Jones \(1998\)](#) under *i.i.d.* samples to time series. In addition, we establish the asymptotic distribution for quantile regression estimators. A powerful tool of Bahadur representation will be established in weak convergence sense, which generalizes [Honda \(2000\)](#) who obtained a Bahadur representation with rather strong conditions. At last, we also apply the nonparametric quantile method to investigate the impact of crude oil futures price on the carbon futures price. The empirical findings shows that this impact is more complex than linear.

3.6 Appendix

Proofs of lemmata

Some basic lemmata are given in this section for later reference.

Lemma A.1 *If X and Y are two random variables which are measurable with respect to \mathcal{A} and \mathcal{B} , respectively, and there exist two constants C_1, C_2 , such that $|X| \leq C_1, |Y| \leq C_2$, a.s., then*

$$|EXY - EXEY| \leq 4C_1C_2\alpha(\mathcal{A}, \mathcal{B}), \quad (\text{A.1})$$

where \mathcal{A} and \mathcal{B} are two σ -algebras generated, respectively, by X and Y , and

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(AB) - P(A)P(B)|.$$

Proof. See the appendix of [Hall and Heyde \(2014\)](#).

Lemma A.2 *Let $m = m_n$ be a positive integer tending to ∞ as $n \rightarrow \infty$. Then under Assumptions A1(ii) and A2, if $E|b(X_i^{(m)})| < \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, for any $x \in R^p$ at which $b(\cdot)$ is continuous,*

$$h_n^{-p}Eb(X_i^{(m)})K((x - X_i^{(m)})/h_n) \rightarrow b(x)f_X(x) \int K(u)du + O(h_n^{-1-p}v_1(m)), \quad (\text{A.2})$$

$$h_n^{-p}Eb(X_i^{(m)})K^2((x - X_i^{(m)})/h_n) \rightarrow b(x)f_X(x) \int K^2(u)du + O(h_n^{-1-p}v_1(m)), \quad (\text{A.3})$$

as $n \rightarrow \infty$. for $i \neq j$,

$$\begin{aligned} &Eb(X_i^{(m)})b(X_j^{(m)})K((x - X_i^{(m)})/h_n)K((x - X_j^{(m)})/h_n) \\ &= Eb(X_i)b(X_j)K((x - X_i)/h_n)K((x - X_j)/h_n) + O(h_n^{-1-p}v_1(m)) \end{aligned} \quad (\text{A.4})$$

Furthermore, if Assumption A1(iv) also holds, then for $j > 0$,

$$h_n^{-p} EK((x - X_1^{(m)})/h_n)K((x - X_{j+1}^{(m)})/h_n) = O(h_n^{\min(p,j)}) + O(h_n^{-1-p}v_1(m)), \quad (\text{A.5})$$

where (A.4) holds uniformly for $j \geq p$.

Proof. The main idea of the proof is similar to that of Lemma A.2 of [Lu and Linton \(2007\)](#), though details are different. We only briefly sketch it here.

To prove (A.2), note first that

$$\begin{aligned} & Eb(X_i^{(m)})K((x - X_i^{(m)})/h_n) \\ &= Eb(X_i)K((x - X_i^{(m)})/h_n) + E(b(X_i^{(m)}) - b(X_i))K((x - X_i^{(m)})/h_n) \\ &=: Eb(X_i)K((x - X_i^{(m)})/h_n) + \delta_{1T}. \end{aligned} \quad (\text{A.6})$$

Here, using the bounded property of the kernel function $K(\cdot)$,

$$\begin{aligned} |\delta_{1T}| &\leq E|b(X_i^{(m)}) - b(X_i)| K((x - X_i^{(m)})/h_n) \\ &\leq CE|b(X_i^{(m)}) - b(X_i)| \leq CE\|X_i^{(m)} - X_i\| = O(v_1(m)). \end{aligned}$$

Next,

$$\begin{aligned} & Eb(X_i)K((x - X_i^{(m)})/h_n) \\ &= Eb(X_i)K((x - X_i)/h_n) + Eb(X_i)(K((x - X_i^{(m)})/h_n) - K((x - X_i)/h_n)) \\ &=: \delta_{2T} + \delta_{3T}, \end{aligned}$$

$$\delta_{2T} = Eb(X_i)K((x - X_i)/h_n) = h_n^p b(x) f_X(x) \int K(u) du; \quad (\text{A.7})$$

owing to the Lipschitz continuity of the kernel function $K(\cdot)$,

$$\begin{aligned} |\delta_{3T}| &\leq E|b(X_i)| |K((x - X_i^{(m)})/h_n) - K((x - X_i)/h_n)| \\ &\leq CE \left\| \frac{X_i^{(m)} - X_i}{h_n} \right\| = O(h_n^{-1}v_1(m)); \end{aligned} \quad (\text{A.8})$$

Then it follows from equations, we can get

$$Eb(X_i)K((x - X_i^{(m)})/h_n) = h_n^p b(x) f_X(x) \int K(u) du + O(h_n^{-1} v_1(m)). \quad (\text{A.9})$$

.

For (A.3), (A.4) and (A.5), it can be proved in an argument similar to that in the above.

Lemma A.3 (Cross-term lemma) *Let $\{(Y_j^{(m)}, X_j^{(m)}); 1 \leq j \leq q\}$ be a stationary sequence with mixing coefficient*

$$\alpha_m(j) := \sup \left\{ |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{B}(\{Y_i^{(m)}, X_i^{(m)}\}), B \in \mathcal{B}(\{Y_{i+j}^{(m)}, X_{i+j}^{(m)}\}) \right\}.$$

Let $(y, x) \mapsto \tilde{b}(y, x)$ be a bounded measurable function defined on $\mathbb{R} \times \mathbb{R}^p$. Set $\eta_j^{(m)}(x) = \tilde{b}(Y_j^{(m)}, X_j^{(m)}) \tilde{K}((x - X_j^{(m)})/h_n)$, where \tilde{K} is a kernel function satisfying Assumption (A2), and $\Delta_j^{(m)}(x) = \eta_j^{(m)}(x) - E\eta_j^{(m)}(x)$, $\tilde{R}(x) = \sum_{1 \leq i < j \leq n} E\Delta_i^{(m)}(x) \Delta_j^{(m)}(x)$. Then, under Assumptions A1, A2 and A3, there exists a constant $C > 0$ such that

$$|\tilde{R}(x)| \leq C n h_n^p \left[\tilde{J}_1(x) + \tilde{J}_2(x) \right]. \quad (\text{A.10})$$

where $\tilde{J}_1(x) := h_n^p N_n \max \{1, h_n^{-2-2p} v_1(m)\}$ and

$$\tilde{J}_2(x) := h_n^p \left(\sum_{j=N_n}^n \alpha_m(j) \right) \max \{1, h_n^{-2-2p} v_1(m)\}.$$

Proof. The main idea of the proof is similar to that of Lemma A.2 of [Lu and Linton \(2007\)](#), though details are different. We only briefly sketch it here. Writing $Z_j^{(m)}$ for $\tilde{b}(Y_j^{(m)}, X_j^{(m)})$, $|Z_j^{(m)}|$ is bounded by some $\tilde{L} > 0$.

$$E\Delta_j^{(m)}(x) \Delta_i^{(m)}(x) = \left\{ E \left(\eta_j^{(m)}(x) \eta_i^{(m)}(x) \right) - E \left(\eta_j^{(m)}(x) \right) E \left(\eta_i^{(m)}(x) \right) \right\}$$

Then applying Lemma A.2,

$$\begin{aligned}
 & \mathbb{E} \left[\Delta_j^{(m)}(x) \Delta_i^{(m)}(x) \right] \\
 &= [\mathbb{E} Z_i K((x - X_i)/h_n) Z_j K((x - X_j)/h_n) - \mathbb{E} Z_i K((x - X_i)/h_n) \mathbb{E} Z_j K((x - X_j)/h_n)] \\
 &\quad + [O(h_n^{-2} v_1^2(m)) + O(h_n^{-1} v_1(m))] + h_n^p [O(h_n^{-1} v_1(m))] \\
 &\leq C h_n^{2p} [1 + O(h_n^{-2p-2} v_1(m))], \tag{A.11}
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 & \sum_{i=1}^n \sum_{j=i+1}^{N_n} \mathbb{E} \left[\Delta_j^{(m)}(x) \Delta_i^{(m)}(x) \right] \leq C h_n^{2p} (n N_n) \max \{1, h_n^{-2p-2} v_1(m)\} \\
 &= n h_n^p \tilde{J}_1(x). \tag{A.12}
 \end{aligned}$$

where N_n is a positive integer depending on n to be specified later. On the other hand,

$$\begin{aligned}
 & \sum_{i=1}^n \sum_{j=i-N_n}^n \mathbb{E} \left[\Delta_j^{(m)}(x) \Delta_i^{(m)}(x) \right] \leq C h_n^{2p} \left(n \sum_{j=N_n}^n \alpha_m(j) \right) \max \{1, h_n^{-2p-2} v_1(m)\} \\
 &= n h_n^p \tilde{J}_2(x). \tag{A.13}
 \end{aligned}$$

Proofs for Section 3.3

The proofs of Lemmas 3.1 and 3.2 are similar. In this section, we will mainly focus on the proof of 3.1. C will denote a generic constant in the proof.

For convenience of the proof, we introduce some notations. Denote $X_{hi} = (X_i - x)/h_n$, $\mathcal{X}_{hi} = (1, X_{hi}')'$, $K_i = K(X_{hi})$, $X_{hi}^{(m)} = (X_i^{(m)} - x)/h_n$, $\mathcal{X}_{hi}^{(m)} = (1, X_{hi}^{(m)})'$, $K_i^{(m)} = K(X_{hi}^{(m)})$, $H_n = \sqrt{n h_n^p}$, $\bar{\theta}_n = H_n (\hat{a}_0 - q_\tau(x), h_n (\hat{a}_1 - \dot{q}(x))')'$, $\theta = H_n (a_0 - q_\tau(x), h_n (a_1 - \dot{q}(x))')'$, $\tilde{\theta} = H_n (\tilde{a}_0 - q_\tau(x), h_n (\tilde{a}_1 - \dot{q}(x))')'$ where $(a_0, a_1)'$, $(\tilde{a}_0, \tilde{a}_1)' \in R^{1+p}$. Let Y_i^* be defined in Lemma 3.1, and set $Y_{ni}^*(\theta) = Y_i^* - \theta' \mathcal{X}_{hi}/H_n$, $T_{ni} = (\dot{q}(x))' X_{hi} h_n$, $U_{ni} = U_{ni}(\theta) = T_{ni} + \theta' \mathcal{X}_{hi}/H_n$. $Y_{ni}^{*(m)}(\theta) = Y_i^{*(m)} - \theta' \mathcal{X}_{hi}^{(m)}/H_n$, $T_{ni}^{(m)} = (\dot{q}(x))' X_{hi}^{(m)} h_n$, $U_{ni}^{(m)} = U_{ni}^{(m)}(\theta) = T_{ni}^{(m)} + \theta' \mathcal{X}_{hi}^{(m)}/H_n$.

The following properties are useful in the discussion.

$$Y_i^* = Y_i - q_\tau(x) - T_{ni}, \quad (\text{A.14})$$

$$Y_i^{*(m)} = Y_i^{(m)} - q_\tau(x) - T_{ni}^{(m)},$$

$$Y_{ni}^*(\theta) = Y_i - q_\tau(x) - U_{ni}(\theta) = Y_i - a_0 - a_1'(X_i - x), \quad (\text{A.15})$$

$$Y_{ni}^{*(m)}(\theta) = Y_i^{(m)} - q_\tau(x) - U_{ni}^{(m)}(\theta) = Y_i^{(m)} - a_0 - a_1'(X_i^{(m)} - x).$$

Since $K(\cdot)$ is a bounded density function with a bounded support,

$$\|X_{hi}\| \leq C, \quad \|\mathcal{X}_{hi}\| \leq C \text{ when } K_i > 0, \quad (\text{A.16})$$

and when $\|\theta\| \leq M$ and $K_i > 0$,

$$|T_{ni}| \leq Ch_n, \quad |U_{ni}| \leq Ch_n + CH_n^{-1} \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (\text{A.17})$$

It follows from (3.2.11) that

$$\bar{\theta}_n = \operatorname{argmin}_{\theta \in R^{1+p}} \sum_{i=1}^n \rho_\tau(Y_{ni}^*(\theta)) K_i \triangleq \operatorname{argmin}_{\theta \in R^{1+p}} G_n(\theta). \quad (\text{A.18})$$

Set

$$V_n(\theta) = H_n^{-1} \sum_{i=1}^n \psi_\tau(Y_{ni}^*(\theta)) \mathcal{X}_{hi} K_i. \quad (\text{A.19})$$

Note that

$$\begin{aligned} V_n(\theta) - V_n(0) &= H_n^{-1} \sum_{i=1}^n [\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)] \mathcal{X}_{hi} K_i \\ &\triangleq H_n^{-1} \sum_{i=1}^n V_{ni}(\theta). \end{aligned} \quad (\text{A.20})$$

Set $V_{ni}(\theta) = (V_{ni}^0(\theta), (V_{ni}^1(\theta))')'$ and $\Delta_i = V_{ni}(\theta) - EV_{ni}(\theta)$, where $V_{ni}^0(\theta) = [\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)] K_i$ and $V_{ni}^1(\theta) = [\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)] \mathcal{X}_{hi} K_i$. And set $V_{ni}^{(m)}(\theta) = (V_{ni}^{0(m)}(\theta), (V_{ni}^{1(m)}(\theta))')'$ and $\Delta_i^{(m)} = V_{ni}^{(m)}(\theta) - EV_{ni}^{(m)}(\theta)$, where $V_{ni}^{0(m)}(\theta) = [\psi_\tau(Y_{ni}^{*(m)}(\theta)) - \psi_\tau(Y_i^{*(m)})] K_i^{(m)}$ and $V_{ni}^{1(m)}(\theta) = [\psi_\tau(Y_{ni}^{*(m)}(\theta)) - \psi_\tau(Y_i^{*(m)})] \mathcal{X}_{hi}^{(m)} K_i^{(m)}$.

Lemma A.4 *Let $V_n(\pi)$ be a vector function that satisfies*

$$(i) \quad -\pi' V_n(\lambda\pi) \geq -\pi' V_n(\pi), \quad \lambda \geq 1,$$

(ii) $\sup_{\|\pi\| \leq M} \|V_n(\pi) + f_{Y|X}(q_\tau(x)|x)D\pi - A_n\| = o_P(1)$,
where $\|A_n\| = O_P(1)$, $0 < M < \infty$, $f_{Y|X}(q_\tau(x)|x) > 0$, and D is a positive-definite
matrix. Suppose that π_n such that $\|V_n(\pi_n)\| = o_P(1)$. Then, $\|\pi_n\| = O_P(1)$ and

$$\pi_n = [f_{Y|X}(q_\tau(x)|x)]^{-1}D^{-1}A_n + o_P(1).$$

Proof. This is Lemma A.4 of [Koenker and Zhao \(1996\)](#).

The proof of Theorem 3.1 is based on the following lemmas in checking the conditions of Lemma A.4. We first prove a lemma which will be applied repeatedly.

Lemma A.5 Under Assumptions A1(ii, iii) and A2 with n large enough,

$$\begin{aligned} E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(\tilde{\theta}))|K_i &\leq CEI_{(|Y_{ni}^*(\tilde{\theta})| < C\|\theta - \tilde{\theta}\|/H_n)} K_i \leq C\|\theta - \tilde{\theta}\|h_n^p/H_n, \\ E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(\tilde{\theta}))|^2K_i^2 &\leq CEI_{(|Y_{ni}^*(\tilde{\theta})| < C\|\theta - \tilde{\theta}\|/H_n)} K_i^2 \leq C\|\theta - \tilde{\theta}\|h_n^p/H_n, \\ E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)|^2K^2((X_i^{(m)} - x)/h_n) &\leq CEI_{(|Y_{ni}^*(\tilde{\theta})| < C\|\theta\|/H_n)} K_i^{(m)2} \leq C\|\theta\|h_n^p/H_n. \end{aligned}$$

for any $\theta, \tilde{\theta} \in \{\theta : \|\theta\| \leq M\}$.

$$\begin{aligned} E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^{*(m)}(\theta))|^2K^2((X_i^{(m)} - x)/h_n) &\leq CL + \frac{C}{L}v_1(m) \\ E|\psi_\tau(Y_i^*) - \psi_\tau(Y_i^{*(m)})|^2K^2((X_i^{(m)} - x)/h_n) &\leq CL + \frac{C}{L}v_1(m), \end{aligned}$$

where $L = L_n \rightarrow 0$

Proof. From (A.16)

$$\begin{aligned} |\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(\tilde{\theta}))|K_i &= |I_{\{Y_{ni}^*(\theta) < 0\}} - I_{\{Y_{ni}^*(\tilde{\theta}) < 0\}}|K_i \\ &= |I_{\{Y_{ni}^*(\tilde{\theta}) < (\theta - \tilde{\theta})'X_{ni}/H_n\}} - I_{\{Y_{ni}^*(\tilde{\theta}) < 0\}}|K_i \leq I_{\{|Y_{ni}^*(\tilde{\theta})| < C\|\theta - \tilde{\theta}\|/H_n\}}K_i. \end{aligned}$$

We have

$$E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(\tilde{\theta}))|K_i \leq CEI_{(|Y_{ni}^*(\tilde{\theta})| < C\|\theta - \tilde{\theta}\|/H_n)} K_i.$$

From (A.15), there exists $0 < \xi < 1$ such that

$$\begin{aligned} EI_{(|Y_{ni}^*(\tilde{\theta})| < C\|\theta - \tilde{\theta}\|/H_n)} K_i &\leq CE[F_{Y|X}(q_\tau(x) + U_{ni}(\tilde{\theta}) + C\|\theta - \tilde{\theta}\|/H_n|X_i) \\ &\quad - F_{Y|X}(q_\tau(x) + U_{ni}(\tilde{\theta}) - C\|\theta - \tilde{\theta}\|/H_n|X_i)]K_i \\ &\leq CE[f_{Y|X}(q_\tau(x) + U_{ni}(\tilde{\theta}) - C\|\theta - \tilde{\theta}\|/H_n + 2\xi C\|\theta - \tilde{\theta}\|/H_n|X_i)]2C\|\theta - \tilde{\theta}\|/H_n K_i \\ &\leq C\|\theta - \tilde{\theta}\|H_n^{-1}E[f_{Y|X}(q_\tau(x) + U_{ni}(\tilde{\theta}) - C\|\theta - \tilde{\theta}\|/H_n + 2\xi C\|\theta - \tilde{\theta}\|/H_n|X_i)]K_i. \end{aligned}$$

for any $\theta, \tilde{\theta} \in \{\theta : \|\theta\| \leq M\}$. Then using assumptions A1(iii) and A2 together with (A.17) and Lemma A2 with n large enough, we have

$$\begin{aligned} E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{hi}^*(\tilde{\theta}))| K_i &\leq C\|\theta - \tilde{\theta}\|H_n^{-1}E[f_{Y|X}(q_\tau(x)|X_i) \\ &\quad \times f_{Y|X}(q_\tau(x) + U_{ni}(\tilde{\theta}) - C\|\theta - \tilde{\theta}\|/H_n + 2\xi C\|\theta - \tilde{\theta}\|/H_n|X_i)/f_{Y|X}(q_\tau(x)|X_i)]K_i \\ &\leq C\|\theta - \tilde{\theta}\|H_n^{-1} [Ef_{Y|X}(q_\tau(x)|X_i)K_i] = C\|\theta - \tilde{\theta}\| O(h_n^p H_n^{-1}). \end{aligned}$$

This is the first inequality of Lemma A5. The second one and third one can be proved similarly. For the fourth inequality, we have

$$\begin{aligned} &E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^{*(m)}(\theta))|^2 K^2((X_i^{(m)} - x)/h_n) \\ &= E|I_{\{Y_{ni}^*(\theta) < 0\}} - I_{\{Y_{ni}^{*(m)}(\theta) < 0\}}|^2 K^2((X_i^{(m)} - x)/h_n) \\ &\leq E|I_{\{Y_{ni}^*(\theta) < 0\}} - I_{\{Y_{ni}^*(\theta) < Y_{ni}^*(\theta) - Y_{ni}^{*(m)}(\theta)\}}|^2 K_i^{(m)2} \\ &\leq CEI_{\{|Y_{ni}^*(\theta)| < |Y_{ni}^*(\theta) - Y_{ni}^{*(m)}(\theta)|\}}^2 K_i^{(m)2} \\ &\leq C\{P(|Y_{ni}^*(\theta)| < |Y_{ni}^*(\theta) - Y_{ni}^{*(m)}(\theta)|, |Y_{ni}^*(\theta)| \leq L) \\ &\quad + P(|Y_{ni}^*(\theta)| < |Y_{ni}^*(\theta) - Y_{ni}^{*(m)}(\theta)|, |Y_{ni}^*(\theta)| > L)\} \\ &\leq C\{P(|Y_{ni}^*(\theta)| \leq L) + P(|Y_{ni}^*(\theta) - Y_{ni}^{*(m)}(\theta)| > L)\} \\ &\leq CL + \frac{C}{L}E|Y_{ni}^*(\theta) - Y_{ni}^{*(m)}(\theta)| \\ &\leq CL + \frac{C}{L}\{E|Y_i - Y_i^m| + E|X_i - X_i^m|\} \\ &\leq CL + \frac{C}{L}v_1(m), \end{aligned}$$

where $L = L_n$.

Lemma A.6 Under the conditions of Lemma 3.1,

$$\sup_{\|\theta\| \leq M} \|V_n(\theta) - V_n(0) - E(V_n(\theta) - V_n(0))\| = o_P(1). \quad (\text{A.21})$$

Proof. The proof is divided into two steps. First we prove that for any fixed $\theta : \|\theta\| \leq M$,

$$\|V_n(\theta) - V_n(0) - E(V_n(\theta) - V_n(0))\| = o_P(1). \quad (\text{A.22})$$

Then from (A.20), the left-hand side of (A.22) is bounded by

$$H_n^{-1} \left| \sum_{i=1}^n (V_{ni}^0(\theta) - EV_{ni}^0(\theta)) \right| + H_n^{-1} \left\| \sum_{i=1}^n (V_{ni}^1(\theta) - EV_{ni}^1(\theta)) \right\| \triangleq V_n^0 + V_n^1. \quad (\text{A.23})$$

It follows from the stationarity and Lemma A.1 that

$$\begin{aligned} E(V_n^0)^2 &= (nh_n^p)^{-1} \left\{ \sum_{j=1}^n E(\Delta_j^0)^2 + 2 \sum_{1 \leq i < j \leq n} E\Delta_i^0 \Delta_j^0 \right\} \\ &= h_n^{-p} E(\Delta_j^0)^2 + 2(nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} E\Delta_i^0 \Delta_j^0 := A_{n1} + 2A_{n2}, \end{aligned} \quad (\text{A.24})$$

In order to bound (A.24), we apply Lemma A.5 with $\tilde{\theta} = 0$; for $\|\theta\| \leq M$,

$$h_n^{-p} E(\Delta_j^0)^2 \leq h_n^{-p} \text{var}(V_{n1}^0(\theta)) \leq h_n^{-p} E(V_{n1}^0)^2 = h_n^{-p} E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)|^2 K_i^2 \leq CH_n^{-1}.$$

Therefore, to complete the proof of this lemma, it suffices to show that $A_{n2} \rightarrow 0$ as $n \rightarrow \infty$. By noticing $E\Delta_i^0 \Delta_j^0 = E\Delta_i^{0(m)} \Delta_j^{0(m)} + E\Delta_i^{0(m)} (\Delta_j^0 - \Delta_j^{0(m)}) + E(\Delta_i^0 - \Delta_i^{0(m)}) \Delta_j^0$, we can further separate A_{n2} into three parts: $A_{n2} = A_{n21} + A_{n22} + A_{n23}$,

$$\begin{aligned} A_{n21} &:= (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} E\Delta_i^{0(m)} \Delta_j^{0(m)} \\ &\leq C(nh_n^{(1+2/a)p})^{-1/2} + CN_n^a \sum_{j=N_n}^{\infty} \alpha(j-m) = o(1) \end{aligned} \quad (\text{A.25})$$

where $N_n = h_n^{-p/a}$, and the equality is due to Lemma A.3 and A.5.

$$\begin{aligned}
A_{n22} &= (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} E \Delta_i^{0(m)} (\Delta_j^0 - \Delta_j^{0(m)}) \\
&\leq (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} \left\{ E(\Delta_i^{0(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j^0 - \Delta_j^{0(m)})^2 \right\}^{1/2} \\
&= (nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{ E(\Delta_i^{0(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j^0 - \Delta_j^{0(m)})^2 \right\}^{1/2}, \quad (\text{A.26})
\end{aligned}$$

and as a result $E(\Delta_i^{0(m)})^2 \leq Ch_n^p H_n^{-1} = o(1)$ and by using the Lipschitz continuity and boundedness of the kernel $K(\cdot)$

$$\begin{aligned}
&E \left(\Delta_j^0 - \Delta_j^{0(m)} \right)^2 \\
&\leq E \left(V_{nj}^0(\theta) - V_{nj}^{0(m)}(\theta) \right)^2 \\
&= E \left\{ [\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*) - \psi_\tau(Y_{nj}^{*(m)}(\theta)) + \psi_\tau(Y_j^{*(m)})] K((X_j^{(m)} - x)/h_n) \right. \\
&\quad \left. + [\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*)] [K((X_j - x)/h_n) - K((X_j^{(m)} - x)/h_n)] \right\}^2 \\
&\leq 2 \left\{ E[\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*) - \psi_\tau(Y_{nj}^{*(m)}(\theta)) + \psi_\tau(Y_j^{*(m)})]^2 K^2((X_j^{(m)} - x)/h_n) \right. \\
&\quad \left. + E[\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*)]^2 [K((X_j - x)/h_n) - K((X_j^{(m)} - x)/h_n)]^2 \right\} \\
&\leq C \left[E|\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_{nj}^{*(m)}(\theta))|^2 K^2((X_j^{(m)} - x)/h_n) \right. \\
&\quad \left. + E|\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})|^2 K^2((X_j^{(m)} - x)/h_n) \right. \\
&\quad \left. + H_n^{-1} h_n^{p-1} E f_{Y|X}(q_\tau(x)|X_i) \left\| X_j - X_j^{(m)} \right\| \right] \\
&\leq C \left[L + \frac{1}{L} v_1(m) + H_n^{-1} h_n^{p-1} v_1(m) \right],
\end{aligned}$$

we have

$$\begin{aligned}
A_{n22} &\leq (nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{ E(\Delta_i^{0(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j^0 - \Delta_j^{0(m)})^2 \right\}^{1/2} \\
&\leq C(nh_n^p)^{-3/4} n^{3/2} [L + v_1(m)L^{-1} + H_n h_n^{p-1} v_1(m)]^{1/2} = o(1), \quad (\text{A.27})
\end{aligned}$$

where the equality is due to Assumption B2 and $L = L_n = v_1(m)$. And similarly to A_{n22} , it can be proved that

$$A_{n23} := (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} E(\Delta_i^0 - \Delta_i^{0(m)}) \Delta_j^0 \rightarrow 0.$$

Therefore,

$$E(V_n^0)^2 = A_{n1} + 2A_{n2} = o(1), \quad (\text{A.28})$$

Similarly to $E(V_n^0)^2$, we have

$$E(V_n^1)^2 = o(1). \quad (\text{A.29})$$

Thus (A.21) follows from (A.23) together with (A.28) and (A.29).

The next step is to use standard chaining argument in Bickel (1975), see also He et al. (1996). We only give a sketch here. Decompose $\{\|\theta\| \leq M\}$ into cubes based on the grid $(j_1\gamma M, \dots, j_{p+1}\gamma M)$, $j_i = 0, \pm 1, \dots, \pm[1/\gamma] + 1$, where $[1/\gamma]$ denotes the integer part of $1/\gamma$, and γ is a small positive number independent of n . Let $R(\theta)$ be the lower vertex of the cube that contains θ . Clearly, $\|R(\theta) - \theta\| \leq C\gamma$ and the number of the elements of $\{R(\theta) : \|\theta\| \leq M\}$ is finite. Then

$$\sup_{\|\theta\| \leq M} \|V_n(\theta) - V_n(0) - E(V_n(\theta) - V_n(0))\| \leq V_{n1}^* + V_{n2}^* + V_{n3}^*, \quad (\text{A.30})$$

where it follows from (A.21) that $V_{n1}^* = \sup_{\|\theta\| \leq M} \|V_n(R(\theta)) - V_n(0) - E(V_n(R(\theta)) - V_n(0))\| = o_P(1)$, and $V_{n2}^* = \sup_{\|\theta\| \leq M} \|V_n(\theta) - V_n(R(\theta))\|$ and $V_{n3}^* = \sup_{\|\theta\| \leq M} \|E(V_n(\theta) - V_n(R(\theta)))\|$. Using (A.16) and for $\|\theta\| \leq M$, applying Lemma A.5 with $\tilde{\theta} = R(\theta)$ with n large, we have

$$\begin{aligned} V_{n3}^* &\leq CH_n^{-1} n \sup_{\|\theta\| \leq M} E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(R(\theta)))| K_i \\ &\leq C \sup_{\|\theta\| \leq M} \|\theta - R(\theta)\| \leq C\gamma. \end{aligned} \quad (\text{A.31})$$

Therefore letting $n \rightarrow \infty$ and then $\gamma \rightarrow 0$ leads to $V_{n3}^* = o(1)$.

Set $B_i(\theta) = I_{(|Y_{ni}^*(\theta)| < C\gamma/H_n)} \|\mathcal{X}_{hi}\| K_i$. Then

$$V_{n2}^* \leq \sup_{\|\theta\| \leq M} \|V_n(\theta) - V_n(R(\theta))\| \leq C \sup_{\|\theta\| \leq M} H_n^{-1} \sum_{i=1}^n B_i(R(\theta)) \leq B_{n1} + B_{n2}, \quad (\text{A.32})$$

where a similar argument to (A.31) leads to $B_{n1} = C \sup_{\|\theta\| \leq M} H_n^{-1} \sum_{i=1}^n E B_i(R(\theta)) = o(1)$, and similarly to (A.24) and (A.25), $B_{n2} = C \sup_{\|\theta\| \leq M} |H_n^{-1} \sum_{i=1}^n (B_i(R(\theta)) - E B_i(R(\theta)))| = o_P(1)$. Thus, $V_{n2}^* = o_P(1)$. Finally (A.20) follows from (A.30).

Lemma A.7 Under Assumptions A1(iii) and A2,

$$\sup_{\|\theta\| \leq M} \|E(V_n(\theta) - V_n(0)) + f_{Y|X}(q_\tau(x)|x)D\theta\| = o(1), \quad (\text{A.33})$$

where $D = f_X(x)\text{diag}(1, \int uu'K(u)du)$.

Proof. It follows from (A.22) and (A.14-A.17) that

$$\begin{aligned} E(V_n(\theta) - V_n(0)) &= H_n^{-1}nE[I_{(Y_i^* < 0)} - I_{(Y_{ni}^*(\theta) < 0)}]\mathcal{X}_{hi}K_i \\ &= H_n h_n^{-p}E[F(q_\tau(x) + T_{ni}|X_i) - F(q_\tau(x) + U_{ni}(\theta)|X_i)]\mathcal{X}_{hi}K_i. \end{aligned}$$

Then similar to the proof of Lemma A.5 with $U_{ni} - T_{ni} = \theta' \mathcal{X}_{hi}/H_n$ and $\|\theta\| \leq M$, there exists a $0 < \xi < 1$ such that

$$\begin{aligned} &\sup_{\|\theta\| \leq M} \|E(V_n(\theta) - V_n(0)) + f_{Y|X}(q_\tau(x)|x)D\theta\| \\ &= \sup_{\|\theta\| \leq M} \|-h_n^{-p}E[f(q_\tau(x) + T_{ni} + \xi\theta' \mathcal{X}_{hi}/H_n|X_i)]\theta' \mathcal{X}_{hi} \mathcal{X}_{hi}' K_i + f_{Y|X}(q_\tau(x)|x)D\theta\| \\ &= \sup_{\|\theta\| \leq M} \|-h_n^{-p}E[f(q_\tau(x) + T_{ni} + \xi\theta' \mathcal{X}_{hi}/H_n|X_i) - f(q_\tau(x)|X_i)]\theta' \mathcal{X}_{hi} \mathcal{X}_{hi}' K_i \\ &\quad - E[h_n^{-p}f(q_\tau(x)|X_i)]\mathcal{X}_{hi} \mathcal{X}_{hi}' K_i - f_{Y|X}(q_\tau(x)|x)D\theta\| \\ &\leq C \sup_{\|\theta\| \leq M} h_n^{-p}E|f(q_\tau(x) + T_{ni} + \xi\theta' \mathcal{X}_{hi}/H_n|X_i) - f(q_\tau(x)|X_i)|\|\mathcal{X}_{hi} \mathcal{X}_{hi}'\| K_i \\ &\quad + C\|E[h_n^{-p}f(q_\tau(x)|X_i)]\mathcal{X}_{hi} \mathcal{X}_{hi}' K_i - f_{Y|X}(q_\tau(x)|x)D\| = o(1), \end{aligned}$$

where the last inequality follows from Assumptions A1(iii) and A2, (A.17) and Lemma A.2.

Lemma A.8 Let $\bar{\theta}_n$ be the minimizer of the function defined in (A.18). Then

$$\|V_n(\bar{\theta}_n)\| \leq \dim(\mathcal{X}_{hi})H_n^{-1} \max_{i \leq n} \|\mathcal{X}_{hi}' K_i\|.$$

Proof. The proof follows from Ruppert and Carroll (1980).

Lemma A.9 Under Assumptions A1 and A2, if $a \geq 1$ and $h_n \rightarrow 0$, then

$$E(c'V_n(0) - c'EV_n(0))^2 \rightarrow \tau(1 - \tau)f_X(x) \int K_c^2(u)du$$

as $n \rightarrow \infty$, where $c = (c_0, c_1)' \in R^{1+p}$.

Proof. Set $v_i = \psi_\tau(Y_i^*)K_c((X_i - x)/h_n)$, $v_i^{(m)} = \psi_\tau(Y_i^{*(m)})K_c((X_i^{(m)} - x)/h_n)$, and $\eta_i = v_i - Ev_i$, $\eta_i^{(m)} = v_i^{(m)} - Ev_i^{(m)}$. A similar argument to (A.24-A.25) leads to

$$\begin{aligned} E(c'V_n(0) - c'EV_n(0))^2 &= (nh_n^p)^{-1} \left\{ \sum_{i=1}^n E\eta_i^2 + 2 \sum_{1 \leq i < j \leq n} E\eta_i \eta_j \right\} \\ &= h_n^{-p} E\eta_1^2 + 2h_n^{-p} \sum_{1 \leq i < j \leq n} E\eta_i \eta_j \\ &\triangleq v_{n1} + 2v_{n2} \end{aligned} \tag{A.34}$$

Note that (A.2) and (A.3) of Lemma A.2 gives

$$\begin{aligned} EI_{(Y_1^* < 0)} K_c^2((X_1 - x)/h_n) &= EF_{Y|X}(q_\tau(x) + \dot{q}_\tau(X_1 - x)|X_1) K_c^2((X_1 - x)/h_n) \\ &\rightarrow \tau f_X(x) \int K_c^2(u) du, \end{aligned}$$

$$EI_{(Y_1^* < 0)} K_c((X_1 - x)/h_n) \rightarrow \tau f_X(x) \int K_c(u) du,$$

which lead to

$$\begin{aligned} h_n^{-p} Ev_1^2 &= E[\tau^2 - 2\tau I_{(Y_1^* < 0)} + I_{(Y_1^* < 0)}] K_c^2((X_1 - x)/h_n) \\ &\rightarrow \tau(1 - \tau) f_X(x) \int K_c^2(u) du, \end{aligned}$$

and

$$h_n^{-p} Ev_1 = E[\tau - I_{(Y_1^* < 0)}] K_c((X_1 - x)/h_n) \rightarrow (\tau - \tau) f_X(x) \int K_c(u) du = 0.$$

Then

$$v_{n1} = h_n^{-p} Ev_1^2 - h_n^{-p} (Ev_1)^2 \rightarrow \tau(1 - \tau) f_X(x) \int K_c^2(u) du. \tag{A.35}$$

Therefore, to complete the proof of this lemma, it suffices to show that $v_{n2} \rightarrow 0$ as $n \rightarrow \infty$. By noticing $E\eta_i \eta_j = E\eta_i^{(m)} \eta_j^{(m)} + E\eta_i^{(m)} (\eta_j - \eta_j^{(m)}) + E(\eta_i - \eta_i^{(m)}) \eta_j$, we can further separate v_{n2} into three parts: $v_{n2} = v_{n21} + v_{n22} + v_{n23}$,

$$\begin{aligned} v_{n21} &:= (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} E\eta_i^{(m)} \eta_j^{(m)} \leq O(h_n) + \epsilon O(h_n^{p(1-1/a)}) + CN_n^a \sum_{j=N_n}^{\infty} \alpha(j - m) \rightarrow 0 \end{aligned} \tag{A.36}$$

Take $N_n = \epsilon h_n^{-p/a}$, where ϵ is a small positive number, and $a \geq 1$.

$$\begin{aligned}
 v_{n22} &= (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} E \eta_i^{(m)} (\eta_j - \eta_j^{(m)}) \\
 &\leq (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} \left\{ E(\eta_i^{(m)})^2 \right\}^{1/2} \left\{ E(\eta_j - \eta_j^{(m)})^2 \right\}^{1/2} \\
 &= (nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{ E(\eta_i^{(m)})^2 \right\}^{1/2} \left\{ E(\eta_j - \eta_j^{(m)})^2 \right\}^{1/2}, \tag{A.37}
 \end{aligned}$$

and as a result $E(\eta_i^{(m)})^2 \leq E(v_i^{(m)})^2 \leq Ch_n^p = o(1)$ and by using the Lipschitz continuity and boundedness of the kernel $K_c(\cdot)$

$$\begin{aligned}
 &E \left(\eta_j - \eta_j^{(m)} \right)^2 \\
 &\leq E \left(v_j - v_j^{(m)} \right)^2 \\
 &= E \left\{ [\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})] K_c((X_j^{(m)} - x)/h_n) \right. \\
 &\quad \left. + [\psi_\tau(Y_j^*)][K_c((X_j - x)/h_n) - K_c((X_j^{(m)} - x)/h_n)] \right\}^2 \\
 &\leq 2 \left\{ E[\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})]^2 K_c^2((X_j^{(m)} - x)/h_n) \right. \\
 &\quad \left. + E[\psi_\tau(Y_j^*)]^2 [K_c((X_j - x)/h_n) - K_c((X_j^{(m)} - x)/h_n)]^2 \right\} \\
 &\leq C \left[E[\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})]^2 K_c^2((X_j^{(m)} - x)/h_n) \right. \\
 &\quad \left. + h_n^{p-1} E F_{Y|X}(q_\tau(x)|X_j) \left\| X_j - X_j^{(m)} \right\| \right] \\
 &\leq C [L + L^{-1}v_1(m) + h_n^{p-1}v_1(m)],
 \end{aligned}$$

we have

$$\begin{aligned}
 v_{n22} &\leq (nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{ E(\eta_i^{(m)})^2 \right\}^{1/2} \left\{ E(\eta_j - \eta_j^{(m)})^2 \right\}^{1/2} \\
 &\leq C(nh_n^p)^{-1/2} n^{3/2} [L + v_1(m)L^{-1} + H_n h_n^{p-1} v_1(m)]^{1/2} = o(1), \tag{A.38}
 \end{aligned}$$

and similarly to v_{n22} , it can be proved that

$$v_{n23} := (nh_n^p)^{-1} \sum_{1 \leq i < j \leq n} E(\eta_i - \eta_i^{(m)}) \eta_j \rightarrow 0.$$

Therefore,

$$v_{n2} = v_{n21} + v_{n22} + v_{n23} = o(1), \quad (\text{A.39})$$

Finally the lemma follows from (A.35) and (A.39).

Proof of Theorem 3.1 We now check the conditions of Lemma A.4, Lemmas A.6 and A.7 lead to (ii) of Lemma A.4 $\|V_n(\bar{\theta}_n)\| = o_P(1)$ follows from Lemma A.8 together with Assumptions A2 and A3. Take $A_n = V_n(\mathbf{0})$. It can be seen from Lemma A.9 that $A_n = O_P(1)$. Since $\psi_\tau(y)$ is an increasing function of y , the function

$$-\theta' V_n(\lambda\theta) = H_n^{-1} \sum_{i=1}^n \psi_\tau(y_i^* - \lambda\theta' \mathcal{X}_{hi}/H_n)(-\theta' \mathcal{X}_{hi})K_i$$

is increasing as a function of λ . Therefore, condition (i) of Lemma A.4 holds and the result follows.

Proof of Theorem 3.2 Define $\bar{\theta}_n = H_n(\hat{a}_0 - q_\tau(x), h_n \hat{a}_1')'$, $\theta = H_n(a_0 - q_\tau(x), h_n a_1')'$, and let Y_i^* be defined as in Lemma 3.2, that is $Y_i^* = Y_i - q_\tau(x)$, and $U_{ni} = U_{ni}(\theta) = \theta' \mathcal{X}_{hi}/H_n$. Then (A.15), (A.16) and (A.18) remain true, while (A.14) and (A.17) are to be changed into $Y_i^* = Y_i - q_\tau(x)$ and $|U_{ni}| \leq CH_n^{-1} \rightarrow 0$ as $n \rightarrow \infty$, respectively. The argument to Lemma 3.2 is similar to that of Lemma 3.1 given in the above except that because now $Y_i^* = Y_i - q_\tau(x)$, not as defined in (A.14), all the T_{ni} 's in the proof of Lemma A.7 need to be canceled.

The proofs of Theorems 3.3-3.6 are based on the local Bahadur representation given in Section 3.3.

Proof of Theorem 3.3 The arguments are similar. We only give the proof of (3.3.3).

Lemma 3.1 entails that

$$\begin{aligned} \hat{q}_\tau(x) - q_\tau(x) &= (nh_n^p)^{-1} \phi_\tau \sum_{i=1}^n \psi_\tau(Y_i^*) K_i + o_P(1/H_n) \\ &= (nh_n^p)^{-1} \phi_\tau \sum_{i=1}^n [\psi_\tau(Y_i^*) K_i - E\psi_\tau(Y_i^*) K_i] + (nh_n^p)^{-1} \phi_\tau \sum_{i=1}^n E\psi_\tau(Y_i^*) K_i + o_P(1/H_n) \\ &\triangleq Q_{n1} + Q_{n2} + o_P(1/H_n). \end{aligned} \quad (\text{A.40})$$

Note that $\tau = F(q_\tau(X_i)|X_i)$ and by (A.16) and Assumption A3 with $v = 1$ that when $K_i > 0$, there exists $0 < \tilde{\xi} < 1$ such that

$$|\Delta_i(x)| \triangleq |q_\tau(X_i) - q_\tau(x) - T_{ni}| = |[\dot{q}(x + \tilde{\xi}X_{hi}h_n) - \dot{q}(x)]'X_{hi}h_n| \quad (\text{A.41})$$

$$\leq Ch_n^{1+\delta}. \quad (\text{A.42})$$

There exists a $0 < \xi < 1$ from (A.42) that,

$$\begin{aligned} |Q_{n2}| &= |h_n^{-p}\phi_\tau E[\tau - I_{(Y_i^* < 0)}]K_i| = h_n^{-p}\phi_\tau E[F(q_\tau(X_i)|X_i) - F(q_\tau(x) + T_{ni}|X_i)]K_i \\ &= |h_n^{-p}\phi_\tau E[f(q_\tau(x) + T_{ni} + \xi\Delta_i(x)|X_i)\Delta_i(x)K_i]| \\ &\leq O(h_n^{1+\delta})h_n^{-p}\phi_\tau E[f(q_\tau(x) + T_{ni} + \xi\Delta_i(x)|X_i)K_i] = O(h_n^{1+\delta}), \end{aligned} \quad (\text{A.43})$$

where the last inequality is derived from (A.17) and Assumption A1(iii).

On the other hand, it easily follows from Lemma A.9 with $c = (1, \mathbf{0}')' \in R^{1+p}$ that,

$$EQ_{n1}^2 = (nh_n^p)^{-1}\phi_\tau^2 E(c'V_n(0) - c'EV_n(0))^2 = O_P((nh_n^p)^{-1}) = O_P(H_n^{-2}),$$

which entails $Q_{n1} = O_P(H_n^{-1})$. The result of this theorem follows from (A.40).

Proof of Theorem 3.4 With the notations given in the proof of Theorem 3.2, this theorem can be proved similarly to the proof of Theorem 3.3.

Proof of Lemma 3.1. Based on the Bhadur representation of Theorem 3.1, the proofs are similar to the arguments in the corresponding proofs for mean regression in Lu and Linton (2007). We first derive the asymptotic variance and expectation, with the Lemma A.3 replacing the corresponding Lemma A.3 in Lu and Linton (2007). Suppose

$$W_n := \begin{pmatrix} w_{n0} \\ w_{n1} \end{pmatrix}, \quad (W_n)_j := (nh_n^p)^{-1} \sum_{i=1}^n \psi_\tau(Y_i^*) \left(\frac{X_i - x}{h_n} \right)_j K \left(\frac{X_i - x}{h_n} \right), \quad (\text{A.44})$$

$j = 0, \dots, p$, with $\left(\frac{X_i - x}{h_n} \right)_0 = 1$.

Denote by $K_j(x) = (x)_j K(x)$ Then it can be noted that

$$\begin{aligned}
 & E \left| (W_n^{(m)})_j - (W_n)_j \right| \\
 &= E \left| (nh_n^p)^{-1} \sum_{i=1}^n \left[\psi_\tau(Y_i^{*(m)}) K_j \left(\frac{X_i^{(m)} - x}{h_n} \right) - \psi_\tau(Y_i^*) K_j \left(\frac{X_i - x}{h_n} \right) \right] \right| \\
 &\leq (nh_n^p)^{-1} \sum_{i=1}^n E \left| \psi_\tau(Y_i^{*(m)}) K_j \left(\frac{X_i^{(m)} - x}{h_n} \right) - \psi_\tau(Y_i^*) K_j \left(\frac{X_i - x}{h_n} \right) \right| \\
 &= h_n^{-p} E \left| \psi_\tau(Y_i^{*(m)}) K_j \left(\frac{X_i^{(m)} - x}{h_n} \right) - \psi_\tau(Y_i^*) K_j \left(\frac{X_i - x}{h_n} \right) \right| \\
 &\leq h_n^{-p} E \left| \psi_\tau(Y_i^{*(m)}) - \psi_\tau(Y_i^*) \right| K_j \left(\frac{X_i^{(m)} - x}{h_n} \right) \\
 &\quad + h_n^{-p} E \left| \psi_\tau(Y_i^*) \right| \left| K_j \left(\frac{X_i^{(m)} - x}{h_n} \right) - K_j \left(\frac{X_i - x}{h_n} \right) \right| \\
 &= O_p(h_n^{-p} v_1(m)) + O_p(h_n^{-p-1} v_1(m)) \\
 &= O_p(h_n^{-p-1} v_1(m)). \tag{A.45}
 \end{aligned}$$

The usual Cramér-Wold device will be adopted. For all $c := (c_0, c_1)' \in \mathbb{R}^{1+p}$, let

$$A_n := (nh_n^p)^{1/2} c' W_n = \phi_\tau(x) \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^n \psi_\tau(Y_i^*) K_c \left(\frac{X_i - x}{h_n} \right),$$

with $K_c(u)$ defined in (3.2.6).

The expectation of the first term on right-hand side of (3.3.1) is as

$$\begin{aligned}
 & E \left[\phi_\tau(x) \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^n \psi_\tau(Y_i^*) \left(\frac{1}{\frac{X_i - x}{h_n}} \right) K \left(\frac{X_i - x}{h_n} \right) \right] \\
 &= \phi_\tau(x) \frac{1}{\sqrt{nh_n^p}} n E \left[\psi_\tau(Y_i^*) \left(\frac{1}{\frac{X_i - x}{h_n}} \right) K \left(\frac{X_i - x}{h_n} \right) \right] \\
 &= \phi_\tau(x) \sqrt{nh_n^p} h_n^{-p} E \left[(F_{Y|X}(q_\tau(X_i)|X_i) - F_{Y|X}(q_\tau(x) + (\dot{q}_\tau(x))'(X_i - x)|X_i)) \left(\frac{1}{\frac{X_i - x}{h_n}} \right) K \left(\frac{X_i - x}{h_n} \right) \right] \\
 &= \sqrt{nh_n^p} \left[(1 + o(1)) \begin{pmatrix} B_0(x) \\ B_1(x) \end{pmatrix} \right] \tag{A.46}
 \end{aligned}$$

Based on Lemma A.9, the variance is as

$$\begin{aligned}\Sigma &:= \text{Var} \left[\phi_\tau(x) \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^n \psi_\tau(Y_i^*) \left(\frac{1}{\frac{X_i - x}{h_n}} \right) K\left(\frac{X_i - x}{h_n}\right) \right] \\ &= \text{Var}(\phi_\tau(x) V_n(0)) \\ &= \phi_\tau^2(x) \tau(1 - \tau) f_X(x) \begin{pmatrix} \int K^2(u) du & \int u' K^2(u) du \\ \int u K^2(u) du & \int u u' K^2(u) du \end{pmatrix}\end{aligned}\quad (\text{A.47})$$

Then

$$\lim_{n \rightarrow \infty} \text{Var}[A_n] = (\phi_\tau(x))^{-2} c' \Sigma c, \quad (\text{A.48})$$

Proof of Lemma 3.2. The fundamental idea to prove the asymptotic normality of $W_n(x)$ is to divide $W_n(x)$ into two parts: with $m = m_n \rightarrow \infty$ (to be specified later),

$$W_n(x) = W_n^{(m)}(x) + [W_n(x) - W_n^{(m)}(x)], \quad (\text{A.49})$$

Then applying the approximation lemma 3.1,

$$(nh_n^p)^{1/2} [W_n(x) - W_n^{(m)}(x)] = O_p(n^{1/2} h_n^{-1/2p-1} v_1(m)) \rightarrow_P 0, \quad (\text{A.50})$$

following from Assumption B2; and similarly

$$(nh_n^p)^{1/2} E[W_n(x) - W_n^{(m)}(x)] \rightarrow 0. \quad (\text{A.51})$$

Therefore, like [Lu and Linton \(2007\)](#), it suffices to prove that

$$(nh_n^p)^{1/2} (c'[W_n^{(m)}(x) - EW_n^{(m)}(x)]/\sigma)$$

is asymptotically standard normal as $n \rightarrow \infty$, which is the main effort we will made in this paper. Define

$$\eta_i^{(m)}(x) := \psi_\tau(Y_i^{*(m)}) K_c((X_i^{(m)} - x)/h_n),$$

$\zeta_{ni}^{(m)} := h_n^{-p/2} (\eta_i^{(m)}(x) - E\eta_i^{(m)}(x))$, and let $S_n^{(m)} := \sum_{i=1}^n \zeta_{ni}^{(m)}$. Then,

$$n^{-1/2} S_n^{(m)} = (nh_n^p)^{1/2} c'(W_n^{(m)}(x) - EW_n^{(m)}(x)) = A_n^{(m)} - EA_n^{(m)},$$

Then, we decompose $n^{-1/2}S_n^{(m)}$ into smaller pieces involving “large” and “small” blocks. More specifically, consider

$$U^{(m)}(1, n, x, k) := \sum_{j=k(p^*+q)+1}^{k(p^*+q)+p^*} \zeta_{nj}^{(m)}(x),$$

$$U^{(m)}(2, n, x, k) := \sum_{j=k(p^*+q)+p^*+1}^{(k+1)(p^*+q)} \zeta_{nj}^{(m)}(x),$$

where $p^* = p_n^*$ and $q = q_n$ are specified in Assumption B3. Without loss of generality, assume that, for some integer $r = r_n$, n is such that $n = r(p^* + q)$, with $r \rightarrow \infty$. For each integer $1 \leq j \leq 2$, define

$$\Upsilon^{(m)}(n, x, j) := \sum_{k=0}^{r-1} U^{(m)}(j, n, x, k).$$

Clearly $S_n^{(m)} = \Upsilon^{(m)}(n, x, 1) + \Upsilon^{(m)}(n, x, 2)$. Note that $\Upsilon^{(m)}(n, x, 1)$ is the sum of the random variables $\zeta_{nj}^{(m)}$ over “large” blocks, whereas $\Upsilon^{(m)}(n, x, 2)$ are sums over “small” blocks. If it is not the case that $n = r(p^* + q)$ for some integer r , then an additional term $\Upsilon^{(m)}(n, x, 3)$, say, containing all the $\zeta_{ni}^{(m)}$ ’s that are not included in the big or small blocks, can be considered. This term will not change the proof much. The general approach consists in showing that, as $n \rightarrow \infty$,

$$Q_1^{(m)} := \left| \text{Eexp}[iu \Upsilon^{(m)}(n, x, 1)] - \prod_{j=0}^{r-1} \text{Eexp}[iu U^{(m)}(1, n, x, k)] \right| \rightarrow 0, \quad (\text{A.52})$$

$$Q_2^{(m)} := n^{-1} \text{E} \left(\Upsilon^{(m)}(n, x, 2) \right)^2 \rightarrow 0, \quad (\text{A.53})$$

$$Q_3^{(m)} := n^{-1} \sum_{k=0}^{r-1} \text{E}[U^{(m)}(1, n, x, k)]^2 \rightarrow \sigma^2, \quad (\text{A.54})$$

$$Q_4^{(m)} := n^{-1} \sum_{k=0}^{r-1} \text{E}[(U^{(m)}(1, n, x, k))^2 I\{|U^{(m)}(1, n, x, k)| > \varepsilon \sigma n^{1/2}\}] \rightarrow 0 \quad (\text{A.55})$$

for every $\varepsilon > 0$. Note that

$$\begin{aligned} [A_n^{(m)} - \text{E}A_n^{(m)}]/\sigma &= (nh_n^p)^{1/2} c'[W_n^{(m)}(x) - \text{E}W_n^{(m)}(x)]/\sigma = S_n^{(m)}/(\sigma n^{1/2}) \\ &= \Upsilon^{(m)}(n, x, 1)/(\sigma n^{1/2}) + \Upsilon^{(m)}(n, x, 2)/(\sigma n^{1/2}). \end{aligned}$$

The term $\Upsilon^{(m)}(n, x, 2)/(\sigma n^{1/2})$ is asymptotically negligible by (A.53). The random variables $U^{(m)}(1, n, x, k)$ are asymptotically mutually independent by (A.52).

The asymptotic normality of $\Upsilon^{(m)}(n, x, 1)/(\sigma n^{1/2})$ follows from (A.54) and the Lindeberg-Feller condition (A.55). The lemma thus follows if we can prove (A.52)-(A.55). This proof is similar to the arguments in the corresponding proofs for mean regression in [Lu and Linton \(2007\)](#), with the different Lemma A.3 and (A.45) established in the above.

Proof of (A.52) See the Proof of A.41 in appendix of [Lu and Linton \(2007\)](#).

Proof of (A.53). The proof follows exactly as in the corresponding proof for mean regression in [Lu and Linton \(2007\)](#), with the Lemma A.3 and (A.45) replacing. Here, we just briefly show the proof.

For notational simplicity, refer to the random variables $U^{(m)}(2, n, x, k)$, $k = 0, 1, \dots, r-1$, as $\widehat{U}_1, \dots, \widehat{U}_r$. We have

$$\mathbb{E}[\Upsilon^{(m)}(n, x, 2)]^2 = \sum_{i=1}^r \text{Var}(\widehat{U}_i) + 2 \sum_{1 \leq i < j \leq r} \text{Cov}(\widehat{U}_i, \widehat{U}_j) := \widehat{V}_1 + \widehat{V}_2, \quad \text{say.} \quad (\text{A.56})$$

Since X_n is stationary,

$$\text{Var}(\widehat{U}_i) = \sum_{i=1}^q \mathbb{E} \left[\left(\zeta_{ni}^{(m)}(x) \right)^2 \right] + \sum_{1 \leq i < j \leq q} \mathbb{E} \left[\zeta_{nj}^{(m)}(x) \zeta_{ni}^{(m)}(x) \right] := \widehat{V}_{11} + \widehat{V}_{12}.$$

From Lemma A.2 and the Lebesgue density theorem (see Chapter 2 of [Devroye and Györfi, 1985](#)),

$$\begin{aligned} \widehat{V}_{11} &= q \text{Var}\{\zeta_{ni}^{(m)}(x)\} = q \{h_n^{-p} \mathbb{E} \left(\Delta_i^{(m)}(x) \right)^2\} \\ &\leq q \{h_n^{-p} \mathbb{E} \left(\psi_\tau(Y_i^{*(m)}) K((x - X_i^{(m)})/h_n) \right)^2\} \\ &\leq q \{h_n^{-p} \mathbb{E} \left(\psi_\tau(Y_i^*) K((x - X_i)/h_n) \right)^2 + O(h_n^{-1-p} v_1(m))\} \\ &= O(q), \end{aligned}$$

where the final equality follows from $h_n^{-1-p} v_1(m) = o(1)$ by Assumption B2.

We then need the cross lemma, Lemma A.3, for \widehat{V}_{12} and then taking $N_n = q$ yields

$$\begin{aligned}\widehat{V}_{12} &= h_n^{-p} \sum_{1 \leq i < j \leq q} \mathbb{E} \left[\Delta_j^{(m)}(x) \Delta_i^{(m)}(x) \right] \\ &\leq Cq h_n^p \left[q + \sum_{t=q}^{\infty} \alpha_m(t) \right] \\ &:= Cq\pi_n.\end{aligned}$$

It follows from Assumption B3 that $\pi_n = O(1)$ and

$$n^{-1}\widehat{V}_1 = n^{-1}r \left(\widehat{V}_{11} + \widehat{V}_{12} \right) \leq n^{-1}rCq[1 + \pi_n] \leq C\left(\frac{q}{p^* + q}\right)[1 + \pi_n]. \quad (\text{A.57})$$

Similarly, we can obtain

$$|\widehat{V}_2| \leq Cn h_n^p \sum_{t=q}^{\infty} \alpha_m(t). \quad (\text{A.58})$$

Condition B4 implies that $qh_n^p = O(1)$ and $\pi_n = O(1)$. Thus, from (A.56), (A.57), and (A.58),

$$n^{-1}\mathbb{E}[\Upsilon^{(m)}(n, x, 2)]^2 \leq C\left(\frac{q}{p^* + q}\right)[1 + \pi_n] + Ch_n^p \left(\sum_{t=q}^{\infty} \alpha_m(t) \right),$$

which tends to zero by $q/p^* \rightarrow 0$ and condition (B4); (A.53) follows.

Proof of (A.54) and (A.55) The main idea of the proof is similar to the Proof in appendix of Lu and Linton (2007). These easily follow by checking the conditions and changing the Lemma A.3 and (A.45). The detail is omitted.

Proof of Theorems 3.5 and 3.6. Based on the Bhadur representation of Theorem 3.1 and Theorem 3.2, the proofs of Theorems 3.5 and 3.6 are similar to the arguments in the corresponding proofs for mean regression in Lu and Linton (2007) and α -mixing condition in Hallin et al. (2009b).

Then consider (3.3.9). Set $v_i = [d_1\phi_{\tau_1}\psi_{\tau_1}(y_i^*(\tau_1)) + d_2\phi_{\tau_2}\psi_{\tau_2}(y_i^*(\tau_2))]K_i$, $\tilde{\tau} =$

$d_1\phi_{\tau_1}\tau_1 + d_2\phi_{\tau_2}\tau_2$ and $I_i(\tau) = I_{(y_i^*(\tau) < 0)}$. Here $d_i \in R^1$, $i = 1, 2$. A simple calculation leads to

$$v_i^2 = \left\{ \tilde{\tau}^2 - 2\tilde{\tau}(d_1\phi_{\tau_1}I_i(\tau_1) + d_2\phi_{\tau_2}I_i(\tau_2)) + d_1^2\phi_{\tau_1}^2I_i(\tau_1) \right. \\ \left. + 2d_1\phi_{\tau_1}d_2\phi_{\tau_2}I_i(\tau_1)I_i(\tau_2) + d_2^2\phi_{\tau_2}^2I_i(\tau_2) \right\} K_i^2. \quad (\text{A.59})$$

It follows from (3.3.7) that

$$\sqrt{nh_n^p}(d_1\hat{q}_{\tau_1}(x) + d_2\hat{q}_{\tau_2}(x) - d_1q_{\tau_1}(x) - d_2q_{\tau_2}(x)) = H_n^{-1} \sum_{i=1}^n v_i + o_P(1) \stackrel{\Delta}{=} D_n + o_P(1).$$

Similar to (A.43),

$$ED_n = H_n^{-1}nEv_1 = H_n(h_n^{-p}Ev_1) = H_nO(h_n^{1+\delta}) = O((nh_n^{p+2(1+\delta)})^{1/2}) \rightarrow 0.$$

as $n \rightarrow \infty$. An analogous argument to (A.34) gives

$$E(D_n - ED_n)^2 \stackrel{\Delta}{=} v_{n1} + 2v_{n2}. \quad (\text{A.60})$$

Similar to the proof of (A.35), (A.59) together with Lemma A.2 ensures

$$v_{n1} = h_n^{-p}Ev_1^2 - h_n^{-p}(Ev_1)^2 \\ \rightarrow [\tilde{\tau}^2 - 2\tilde{\tau}^2 + d_1^2\phi_{\tau_1}^2\tau_1 + 2d_1\phi_{\tau_1}d_2\phi_{\tau_2}\min(\tau_1, \tau_2) + d_2^2\phi_{\tau_2}^2\tau_2]f_X(x) \int K^2(u)du \\ = [d_1^2\phi_{\tau_1}^2\tau_1 + 2d_1\phi_{\tau_1}d_2\phi_{\tau_2}\min(\tau_1, \tau_2) + d_2^2\phi_{\tau_2}^2\tau_2 - \tilde{\tau}^2]f_X(x) \int K^2(u)du. \quad (\text{A.61})$$

In addition, similar argument to (A.38) leads to

$$|v_{n2}| \rightarrow 0. \quad (\text{A.62})$$

Then the asymptotic variance for (3.3.9) follows from (A.60)-(A.62). The other asymptotic variances in (3.3.10) and (3.3.12) can be obtained similarly.

Chapter 4

On Semiparametric Smoothing of Dynamic Functional- coefficient Autoregressive Models for Location-wide Nonstationary Spatio-Temporal Data

4.1 Introduction

Most of spatio-temporal models have been confined to linear and stationary (see [Cressie and Wikle, 2015](#)), which may not be true for real spatial and temporal data. For example, some nonlinear relationship between variables have often been found in real estate markets, energy markets and environmental sciences. Moreover, the price changes in regional real estate markets and energy markets may be location-wide non-stationary processes, in the sense that the mean values and variances of price changes in these markets may change with location. For instance, the regional natural gas prices expected returns depend on the geographic location, which means that they may vary from one location to another. So the model parameters may depend on the locations. The analysis of nonlinear time series data have been widely discussed in previous research (see, e.g., [Robinson, 1983](#); [Tjøstheim and Auestad, 1994](#); [Masry and Fan, 1997](#); [Cai et al., 2000](#); [Fan and Yao, 2003](#)). However, nonlinear modelling of spatio-temporal data is often a challenge

with irregularly observed locations and location-wide non-stationarity ([Al-Sulami et al., 2017a](#)).

In this chapter, a semiparametric family of *Dynamic Functional-coefficient Autoregressive Spatio-Temporal (DyFAST)* models will be proposed to address the difficulty in modelling and analysis of our spatio-temporal data, $Y_t(s_i)$, $t = 1, \dots, T$ and $i = 1, 2, \dots, N$, which are stationary along time but non-stationary over the irregular locations in space. Nonlinear modelling of such spatio-temporal data is often a challenge, with irregularly observed locations and location-wide non-stationarity (c.f., [Lu et al., 2009](#); [Al-Sulami et al., 2017a](#)).

The proposed DyFAST models at least own two significant features. The first feature of the proposed models lies in functional (or varying) coefficient structures with the autoregressive smooth coefficients depending both on concerned regime variable and spatial location. Here the dynamic nature of spatio-temporal data is characterised not only by the autoregressive structure but also by the functional (or varying) coefficient structure, which are popular in traditional statistical analysis of *i.i.d.* or time series data. This kind of models can hence not only well characterise the dynamic regime-switching nature but also overcome the location-wide non-stationarity existent in spatio-temporal data. Our model is a useful extension of spatial autoregressive models ([Ord, 1975](#); [Gao et al., 2006](#); [Kelejian and Prucha, 2010](#); [Su and Jin, 2010](#)) and varying coefficient models ([Zhang and Fan, 1999](#); [Fan and Zhang, 2000](#); [Zhang et al., 2002](#); [Sun et al., 2007](#); [Cheng et al., 2009](#); [Wang and Xia, 2009](#); [Zhang et al., 2009](#); [Li and Zhang, 2011](#); [Hu and Xia, 2012](#)).

To model the dynamic spatial neighbouring temporal-lagged effects with the irregular locations, we consider using spatial weight matrix pre-specified either by experts or by the prior information of spatial locations, which is popular in spatial econometrics. Moreover, both one-step and two-step estimation methods are proposed to estimate the functional coefficients in DyFAST models. In practice, such spatial weight matrix can be pre-specified by the features of the data in many different ways. Therefore, the idea of model selection can be applied to select an optimal weight matrix among the candidates. Both theoretical properties and Monte Carlo simulations are investigated. The empirical applications to energy market data sets further illustrate the usefulness of the models.

In Section 4.2, the Dynamical Functional-coefficient Autoregressive Spatio-Temporal (DyFAST) Models and adapt local linear regression technique for estimation will be presented. Section 4.3 provides DyFAST models and the asymptotic properties

of the estimators. In Section 4.4, the discussion about the selection of the weight matrix will also be provided. Section 4.5 investigates the finite-sample properties via a simulation study. In Section 4.6, the methodology is demonstrated to investigate diesel prices in relation to crude oil price in the EU countries.

4.2 Dynamical Functional-coefficient Autoregressive Spatio-Temporal (DyFAST) Models

Let $Y_t(s)$ be a spatio-temporal process observed at discrete time points, $t = 1, \dots, T$ and at irregularly positioned locations, s_i , $i = 1, 2, \dots, N$, on a spatial domain $S \subset \mathbb{R}^2$. In practice, the data are usually non-stationary across space, such as certain changes of retail gasoline prices along time considered in the real data section below. Here, along time, the considered quantity at a return level, such as the log-return, that is the change of the logarithmic prices in a given horizon of time, is often thought as stationary over time, while it is usually uneasy to turn the data into stationary one along space. This chapter is therefore concerned with modelling of the spatio-temporal data, $Y_t(s_i)$, which are stationary along time but non-stationary over the irregular locations in space. How to model this kind of data that are with irregularly observed locations and of location-wide non-stationarity is a challenge.

To deal with the irregularly observed locations, a popular idea in spatial econometrics (c.f., [Anselin, 2013](#)) is to model the spatial neighbouring effects in linear modelling of $Y_t(s_i)$ by using a spatial weight matrix $W = (w_{ij})_{N \times N}$, following [Cliff and Ord \(1972\)](#) to define a spatial lag variable

$$Y_t^{SL}(s_i) = \sum_{j=1}^N w_{ij} Y_t(s_j),$$

where, as in the above mentioned literature, W is supposed to be row-wise standardised, which is satisfying $w_{ij} \geq 0$ and $\sum_{j=1}^N w_{ij} = 1$. With a pre-specified spatial weight matrix W either by experts knowledge or by using location-related information, a linear spatio-temporal autoregressive model from the forecasting

perspective can be formulated as follows:

$$Y_t(s_i) = \beta_0(s_i) + \sum_{j=1}^p \beta_j(s_i) Y_{t-j}^{\text{SL}}(s_i) + \sum_{l=1}^q \beta_{p+l}(s_i) Y_{t-l}(s_i) + \varepsilon_t(s_i), \quad (4.2.1)$$

$$i = 1, \dots, N, \quad t = r + 1, \dots, T.$$

This kind of autoregressive spatio-temporal model can well characterise the location-wide non-stationarity existent in the data (c.f., [Subba Rao, 2008](#)). However, at each location s_i , it is still a constant coefficient linear time series model.

In the sequel, this study further considers the dynamic impact of a regime-switching covariate variable $X_t(s)$ that denotes a spatio-temporal covariate vector of dimension d which may impact $Y_t(s)$. Without loss of generality, this study consider only the case $d = 1$. For multivariate $X_t(s)$, the modelling procedure and the related theory for the univariate case continue to hold, but further, more complicated notations are involved. Assume that both processes are observed at the time points $t = 1, \dots, T$ and at N spatial locations $s_i = (u_i, v_i)' \in S$ for $i = 1, \dots, N$ on a possibly irregular grid. By extending (4.2.1), a semiparametric class of Dynamical Functional-coefficient Autoregressive Spatio-Temporal (DyFAST) model is therefore proposed in the form

$$Y_t(s_i) = \beta_0(X_t(s_i), s_i) + \sum_{j=1}^p \beta_j(X_t(s_i), s_i) Y_{t-j}^{\text{SL}}(s_i) + \sum_{l=1}^q \beta_{p+l}(X_t(s_i), s_i) Y_{t-l}(s_i) + \varepsilon_t(s_i), \quad (4.2.2)$$

$$i = 1, \dots, N, \quad t = r + 1, \dots, T,$$

where $\beta_j(x, s)$, $j = 0, 1, \dots, p + q$, are unknown functional coefficients that need to be estimated.

This semiparametric family takes account of salient features of dynamic spatio-temporal data. It covers a wide range of existent models in the literature. For example, it extends, to the spatio-temporal network setting, the nonlinear functional-coefficient autoregressive time series model at each fixed location s_i , which is very popular in the literature of time series data analysis (c.f., [Chen and Tsay, 1993](#); [Cai et al., 2000](#)). Our DyFAST family also extends the semiparametric spatio-temporal models of [Subba Rao \(2008\)](#), [Lu et al. \(2009\)](#) and [Al-Sulami et al. \(2017a\)](#), among others. For example, [Subba Rao \(2008\)](#) only considers the temporal lag effect without taking the spatial neighbouring effects into account. Also, [Lu et al. \(2009\)](#) only apply to the case when the data are observed regularly on a lattice with the

same small number of neighbouring variables able to be well specified for each site, which is impossible to be implemented with the irregularly positioned data considered in this paper. Furthermore, Al-Sulami et al. (2017a) is only a special case of model (4.2.2) with the β coefficients (except β_0) depending on the location site only, which actually forms a constant-coefficient time series model at each given site s_i , unable to capture the dynamic regime-switching effects in reality. Clearly, our model (4.2.2) can well overcome all these shortcomings.

This proposed family of DyFAST models at least consists of two significant features. (i) The functional (or varying) coefficient structures that are popular in traditional statistical analysis of *i.i.d.* or time series data can be well used to characterise not only the dynamic regime-switching nature but also the location-wide non-stationarity in the autoregressive coefficients that depend on both the regime variable and the spatial location. (ii) The spatial weight matrix pre-specified either by experts or by the prior information of spatial locations is considered to model the dynamic spatial neighbouring temporal-lagged effects with the irregular locations, which is popular in spatial econometrics. Moreover, both one-step and two-step estimation methods are proposed to estimate the functional coefficients in DyFAST models. The idea of model selection can be applied to select an optimal weight matrix among the different pre-specified spatial weight matrices.

4.3 Estimation

With a pre-specified spatial weight matrix W given, the dynamic functional-coefficient spatio-temporal autoregressive model given in (4.2.2) can be rewritten as

$$Y_t(s_i) = Z_t(s_i)' \beta(X_t(s_i), s_i) + \varepsilon_t(s_i), \quad (4.3.1)$$

for $t = r + 1, \dots, T$, with $r = \max\{p, q\}$, and $i = 1, \dots, N$, where, through this chapter, the notation A' denotes the transpose of a vector or matrix A ,

$$Z_t(s_i) = (1, Y_{t-1}^{\text{SL}}(s_i), \dots, Y_{t-p}^{\text{SL}}(s_i), Y_{t-1}(s_i), \dots, Y_{t-q}(s_i))'$$

denoting the vector of spatio-temporally lagged variables, and

$$\beta(x, s) = (\beta_0(x, s), \beta_1(x, s), \dots, \beta_p(x, s), \beta_{p+1}(x, s), \dots, \beta_{p+q}(x, s))'$$

denotes the corresponding vector of functional autoregressive coefficients.

4.3.1 One-step estimation method

We first need to estimate the coefficients $\beta(\cdot, \cdot)$ in (4.3.1). Differently from the two-step estimation procedure (c.f., [Lu et al., 2009](#); [Al-Sulami et al., 2017a](#)) separating utilisation of spatial and temporal information into two steps, this study proposes a one-step estimation procedure in the spatio-temporal setting by considering $\beta_j(X_t(s_i), s_i)$ as a function of $X_t(s_i)$ and s_i . It can be seen as an extension of the time series estimation of [Cai et al. \(2000\)](#). In this proposed procedure, all the data across space and time are utilised together to estimate the coefficients $\beta(\cdot, \cdot)$ in one step.

By the idea of local linear fitting (c.f., [Fan and Gijbels, 1996](#)), we can approximate the unknown function $\beta_j(X_t(s_i), s_i)$ by a local linear function. For a given $s = (u, v)'$, we denote $(\partial\beta(x, s)/\partial u, \partial\beta(x, s)/\partial v)'$ by $\dot{\beta}^s(x, s)$, where $\partial\beta(x, s)/\partial u = (\partial\beta_0(x, s)/\partial u, \dots, \partial\beta_{p+q}(x, s)/\partial u)'$. For any $X_t(s_i)$ and s_i in the neighbourhood of x and s , we have

$$\begin{aligned}\beta_j(X_t(s_i), s_i) &\approx \beta_j(x, s) + \dot{\beta}_j^x(x, s)(X_t(s_i) - x) + \dot{\beta}_j^s(x, s)'(s_i - s) \\ &\equiv b_{0,j} + b_{1,j}(X_t(s_i) - x) + b'_{2,j}(s_i - s).\end{aligned}\quad (4.3.2)$$

Locally, estimating $(\beta(x, s), \dot{\beta}^x(x, s), \dot{\beta}^s(x, s)')$ is equivalent to estimating (b_0, b_1, b'_2) . This motivates us to define an estimator by setting $\hat{\beta}_j(x, s) \equiv \hat{b}_{0,j}$, $\hat{\beta}_j^x(x, s) \equiv \hat{b}_{1,j}$, and $\hat{\beta}_j^s(x, s) \equiv \hat{b}_{2,j}$, which minimise

$$\sum_{i=1}^N \sum_{t=r+1}^T \left[Y_t(s_i) - \sum_{j=0}^{p+q} \{b_{0,j} + b_{1,j}(X_t(s_i) - x) + b'_{2,j}(s_i - s)\} Z_{t,j}(s_i) \right]^2 K_{it}, \quad (4.3.3)$$

with respect to $b_{0,j}$, $b_{1,j}$ and $b_{2,j}$ for $j = 0, 1, \dots, (p+q)$, where $K_{it} = K_{h_1}(X_t(s_i) - x)L_{h_2}(s_i - s)$, $K_{h_1}(x) = h_1^{-1}K(x/h_1)$, with K a kernel function on \mathbb{R}^1 and $h_1 > 0$ a bandwidth, and $L_{h_2}(s) = h_2^{-2}L(s/h_2)$ with L a kernel function on \mathbb{R}^2 and $h_2 > 0$ another bandwidth.

4.3.2 Asymptotic property

We provide the asymptotic property for the estimator $\hat{\beta}(x, s)$ in this subsection. Let $Y = (Y_{r+1}(s_1), \dots, Y_T(s_1), \dots, Y_{r+1}(s_N), \dots, Y_T(s_N))'$ denote a NT_0 -dimensional vector, with $T_0 = T - r$. The local linear estimators can be expressed

as

$$\begin{pmatrix} \hat{b}_0 \\ h_1 \hat{b}_1 \\ h_2 \hat{b}_2 \end{pmatrix} = U_{TN}^{-1} V_{TN},$$

where

$$U_{TN} = (NT_0)^{-1} \tilde{Z}' \tilde{W} \tilde{Z} \quad (4.3.4)$$

and

$$V_{TN} = (NT_0)^{-1} \tilde{Z}' \tilde{W} Y, \quad (4.3.5)$$

where \tilde{Z} denotes an $NT_0 \times 4(p+q+1)$ matrix with $\left(Z_t(s_i)', \frac{X_t(s_i)-x}{h_1} Z_t(s_i)', \left(\frac{s_i-s}{h_2} \otimes Z_t(s_i) \right)' \right)$ as its $(i \times T_0 + t - r)$ th row, and

$$\tilde{W} = \text{diag} \left\{ K_h(X_{r+1}(s_1) - x) L_h(s_1 - s), \dots, K_h(X_T(s_1) - x) L_h(s_1 - s), \right. \\ \left. \dots, K_h(X_T(s_N) - x) L_h(s_N - s) \right\},$$

and $H_{it} = \left(1, \frac{X_t(s_i)-x}{h_1}, \left(\frac{s_i-s}{h_2} \right)' \right)'$. Then,

$$U_{TN} = (T_0 N)^{-1} \sum_{i=1}^N \sum_{t=r+1}^T H_{it} H_{it}' \otimes Z_t(s_i) Z_t(s_i)' K_{it}, \quad (4.3.6)$$

and

$$V_{TN} = (T_0 N)^{-1} \sum_{i=1}^N \sum_{t=r+1}^T H_{it} \otimes Z_t(s_i) Y_t(s_i) K_{it}. \quad (4.3.7)$$

Thus, with $e_1 = (1, 0, 0, 0) \otimes \text{diag}(\mathbf{1}_{p+q+1}')$, the local linear estimator of $\beta(x, s)$ is given by

$$\hat{\beta}(x, s) = \hat{b}_0 = e_1 U_{TN}^{-1} V_{TN}. \quad (4.3.8)$$

$$\begin{aligned}
& \left(\hat{b}_0(x, s) - b_0(x, s) \right) = \left(\hat{\beta}(x, s) - \beta(x, s) \right) \\
& = e_1 U_{TN}^{-1} \left\{ V_{TN} - U_{TN} \begin{pmatrix} b_0 \\ h_1 b_1 \\ h_2 b_2 \end{pmatrix} \right\} \\
& := e_1 U_{TN}^{-1} \mathcal{W}_{TN},
\end{aligned} \tag{4.3.9}$$

$$\mathcal{W}_{TN,j} := (T_0 N)^{-1} \sum_{i=1}^N \sum_{t=r+1}^T (H_{it})_j \otimes Y_t^*(s_i) Z_t(s_i) K_{it}, \quad j = 1, 2, 3, \tag{4.3.10}$$

where $Y_t^*(s_i) := Y_t(s_i) - [a(x, s) + b_1(x, s)(X_t(s_i) - x) + b_2(x, s)'(s_i - s)]' Z_t(s_i)$.

Lemma 4.1. *Under Assumptions in Section 4.8,*

$$U_{TN} \xrightarrow{P} U := f_X(x, s) f_S(s) \begin{pmatrix} \Omega(x, s) & 0 & 0 \\ 0 & \Omega(x, s) \mu_2^K & 0 \\ 0 & 0 & \mu_2^L \otimes \Omega(x, s) \end{pmatrix}$$

as $T_0, N \rightarrow \infty$, where $f_X(x, s)$ represents the marginal probability density function of X given location s , $f_S(s)$ represents the sampling density function, $\Omega(x, s) = E[Z_t(s) Z_t(s)' | X_t(s) = x]$, $\mu_j^K = \int u^j K(u) du$, $\mu_2^L = \int u u' L(u) du$.

Next, we consider the asymptotic behaviour of $E[\mathcal{W}_{TN}]$.

Lemma 4.2. *Under Assumptions in Section 4.8,*

$$\begin{aligned}
E[\mathcal{W}_{TN,1}] &= f_X(x, s) f_S(s) \Omega(x, s) B_0 + o(h_1^2 + h_2^2), \\
\lim_{N, T_0 \rightarrow \infty} E[\mathcal{W}_{TN,j}] &= 0, \quad j = 2, 3
\end{aligned} \tag{4.3.11}$$

where

$$\begin{aligned}
B_0 &= \frac{h_1^2}{2} \mu_1(x, s) + \frac{h_2^2}{2} \mu_2(x, s), \\
\mu_1(x, s) &= \frac{\partial^2 \beta(x, s)}{\partial x^2} \int u^2 K(u) du, \\
\mu_2(x, s) &= \left(\text{tr} \left\{ \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\}, \dots, \text{tr} \left\{ \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right)'.
\end{aligned}$$

The following lemma provides the asymptotic variance of $\mathcal{W}_{TN,1}$.

Lemma 4.3. *Under Assumptions in Section 4.8,*

$$\text{Var}[\mathcal{W}_{TN,1}] = (T_0 h_1 N h_2^2)^{-1} \Sigma_1 (1 + o(1)) + T_0^{-1} \Sigma_2 (1 + o(1)), \quad (4.3.12)$$

where

$$\begin{aligned} \Sigma_1 &= \sigma^2(s) \frac{\Omega(x, s)}{f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz, \\ \Sigma_2 &= \Gamma(s, s) \frac{\Omega^*(x, s, s)}{f_X^2(x, s)} q(x, x; s), \end{aligned}$$

with $\Omega(x, s) = E[Z_t(s)Z_t(s)'|X_t(s) = x]$, $\Omega^*(x, s_i, s_j) = E[Z_t(s_i)Z_t(s_j)'|X_t(s_i) = x, X_t(s_j) = x]$, $\Gamma(s_i, s_j) = E(\varepsilon_t(s_i)\varepsilon_t(s_j)|X_t(s_i) = x, X_t(s_j) = x)$ and $\sigma^2(s) = \text{Var}(\varepsilon_1(s))$, $q(\cdot, \cdot; s)$ defined in Assumption (A1).

We now turn to the main consistency and asymptotic normality result as follows.

Theorem 4.1. *Let Assumptions in Section 4.8 hold. Then,*

$$\hat{\beta}(x, s) - \beta(x, s) - B_0 = \left\{ (T_0 h_1 N h_2^2)^{-\frac{1}{2}} \Theta_1 + T_0^{-\frac{1}{2}} \Theta_2 \right\} \eta(s) (1 + o_p(1)), \quad (4.3.13)$$

as $T_0, N \rightarrow \infty$, where B_0 is defined in Lemma 4.2, where $\eta(s)$ is a $p + q + 1$ random vector with zero mean and identity variance matrix, and Θ_1 and Θ_2 are two $(p + q + 1) \times (p + q + 1)$ matrices, satisfying

$$\begin{aligned} \Theta_1 \Theta_1' &= \sigma^2(s) \frac{\Omega^{-1}(x, s)}{f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz, \\ \Theta_2 \Theta_2' &= \Gamma(s, s) \frac{\Omega^{-1}(x, s) \Omega^*(x, s, s) \Omega^{-1}(x, s)}{f_X^2(x, s)} q(x, x; s) \end{aligned}$$

with $\Omega(x, s) = E[Z_t(s)Z_t(s)'|X_t(s) = x]$, $\Omega^*(x, s_i, s_j) = E[Z_t(s_i)Z_t(s_j)'|X_t(s_i) = x, X_t(s_j) = x]$, $\Gamma(s_i, s_j) = E(\varepsilon_t(s_i)\varepsilon_t(s_j)|X_t(s_i) = x, X_t(s_j) = x)$ and $\sigma^2(s) = \text{Var}(\varepsilon_1(s))$.

The proofs of the above lemmata and Theorem 4.1 will be given in Section 4.8.

Remark 1: For the one-step method, the optimal bandwidths for estimating $\beta_j(x, s)$ can be obtained by minimizing the squared bias plus variance, which is as follow.

$$\min_{h_1, h_2} \left[(c_2 h_1^2 + c_3 h_2^2)^2 + \frac{c_4}{T_0 N h_1 h_2^2} \right]. \quad (4.3.14)$$

This leads to the optimal bandwidths

$$h_1 = \left(\frac{c_3 c_4}{4c_2^2} \frac{1}{T_0 N} \right)^{\frac{1}{7}},$$

$$h_2 = \left(\frac{2^{\frac{3}{2}} c_2^{\frac{3}{2}}}{c_3^{\frac{5}{2}}} \frac{c_4}{T_0 N} \right)^{\frac{1}{7}},$$

where $c_2 = \frac{1}{2} \frac{\partial^2 \beta_j(x, s)}{\partial x^2} \int u^2 K(u) du$, $c_3 = \frac{1}{2} \left\| \text{tr} \left\{ \frac{\partial^2 \beta_j(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right\|$, and

$$c_4 = \frac{\sigma^2(s) e_j' \Omega^{-1}(x, s) e_j}{f_X(x, s) f_S(s)} \int K^2(u) du \int L^2(u) du.$$

Remark 2: Based on the optimal bandwidths, we have $h_1 N h_2^2 = O(N/(T_0 N)^{3/7}) = O((N^4/T_0^3)^{1/7})$. Therefore, if $h_1 N h_2^2 = O((N^4/T_0^3)^{1/7}) \rightarrow O(1)$ or $h_1 N h_2^2 = O((N^4/T_0^3)^{1/7}) \rightarrow \infty$, the rate of the convergence for $\hat{\beta}(x, s)$ is $T^{\frac{1}{2}}$. Otherwise, the rate of the convergence for $\hat{\beta}(x, s)$ is $(T_0 h_1 N h_2^2)^{\frac{1}{2}} = o_p(T_0^{\frac{1}{2}})$.

4.3.3 Bandwidth and Order Selection

In the one-step estimation procedure above, there are two bandwidths (h_1, h_2) used in model (4.3.3) and two orders (p, q) in model (4.2.2), which need to be selected in application. For the bandwidths h_1 and h_2 , we will use the cross-validation method (Li and Racine, 2004) to choose the optimal bandwidth in this study.

$$\text{CV}(h_1, h_2) = \frac{1}{T_0 N} \sum_{t=r+1}^T \sum_{i=1}^N \left\{ Y_t(s_i) - \sum_{j=0}^{p+q} \check{\beta}_{j, h_1, h_2}(X_{-t}, s_{-i}) Z_{t, j}(s_i) \right\}^2. \quad (4.3.15)$$

Then, we let

$$h_{1, \text{opt}}^{h_2} = \arg \min_{h_1} \text{CV}(h_1, h_2) \quad (4.3.16)$$

for given h_2 . Thus, we can select $h_{2, \text{opt}} = \arg \min_{h_2} \text{CV}(h_{1, \text{opt}}^{h_2}, h_2)$, and $h_{1, \text{opt}} = h_{1, \text{opt}}^{h_{2, \text{opt}}}$.

Moreover, we suggest a procedure by extending the method proposed by Hurvich et al. (1998) and Al-Sulami et al. (2017a) to develop the following nonparametric version of the Akaike Information Criterion with correction (AICc) to determine

the orders of temporally lagged variables, p and q ,

$$\text{AIC}_c(p, q) = \log(\hat{\sigma}^2) + \frac{1 + (T_0 N)^{-1} \text{tr}(H)}{1 - (T_0 N)^{-1} \{\text{tr}(H) + 2\}}, \quad (4.3.17)$$

where $\hat{\sigma}^2 = (T_0 N)^{-1} \sum_{t=r+1}^T \sum_{i=1}^N \left\{ Y_t(s_i) - \hat{Y}_t(s_i) \right\}^2$ and the hat matrix H is an $NT_0 \times NT_0$ matrix, such that $\hat{Y} = HY$ with

$$\overbrace{\begin{pmatrix} \hat{Y}(s_1) \\ \hat{Y}(s_2) \\ \vdots \\ \hat{Y}(s_N) \end{pmatrix}}^{\hat{Y}} = \overbrace{\begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NN} \end{pmatrix}}^H \overbrace{\begin{pmatrix} Y(s_1) \\ Y(s_2) \\ \vdots \\ Y(s_N) \end{pmatrix}}^Y.$$

Here, the $T_0 \times T_0$ sub-matrix $H_{jk} = 0$ for $i \neq k$, $i, k = 1, \dots, N$ and $\hat{Y}(s_i) = H_{ii}Y(s_i)$, with $\hat{Y}(s_i) = (\hat{Y}_{r+1}(s_i), \dots, \hat{Y}_T(s_i))'$ and $Y(s_i) = (Y_{r+1}(s_i), \dots, Y_T(s_i))'$. Here, $(t-r)$ th row of H_{ii} is of the form $e' \left\{ \tilde{Z}(X_t(s_i), s_i)' W(X_t(s_i), s_i) \tilde{Z}(X_t(s_i), s_i) \right\}^{-1} \tilde{Z}(X_t(s_i), s_i)' W(X_t(s_i), s_i)$, for $t = r + 1, \dots, T$.

We now outline the algorithm to select the order and bandwidth as follows:

Step 1: For given p and q , minimizing $\text{CV}(h_1, h_2)$, giving the best selection of bandwidth $\hat{h}_{1,opt}$ and $\hat{h}_{2,opt}$.

Step 2: Compare the $\text{AIC}_c(p, q)$ for different p and q to find the minimum value.

4.3.4 Discussion on comparison with a two-step procedure

In the above estimation, all the data across space and time are used in one step, which is hence called one-step estimation. Alternatively, we can develop a two-step procedure for estimating the unknown function $\beta(\cdot, \cdot)$ as in [Lu et al. \(2009\)](#) and [Al-Sulami et al. \(2017a\)](#), which is sketched for comparison as follows.

Step 1: (Time-series based estimation) This step just follows the estimation procedure in [Cai et al. \(2000\)](#). For each location s_i , the coefficient functions $\beta_j(X_t(s_i), s_i)$ are estimated by a local linear fitting method.

At a fixed location s_i , we estimate $\beta_j(x, s_i)$ for $X_t(s_i)$ in the neighbourhood of x as follows

$$\begin{aligned}\beta_j(X_t(s_i), s_i) &\approx \beta_j(x, s_i) + \dot{\beta}_j(x, s_i)(X_t(s_i) - x) \\ &\equiv a_j + b_j(X_t(s_i) - x).\end{aligned}\quad (4.3.18)$$

Then, we define an estimator by setting $\hat{\beta}_j(x, s_i) \equiv \hat{a}_j$ and $\hat{\dot{\beta}}_j(x, s_i) \equiv \hat{b}_j$, which minimise:

$$\sum_{t=r+1}^T \left[Y_t(s_i) - \sum_{j=0}^{p+q} \{a_j + b_j(X_t(s_i) - x)\} Z_{t,j}(s_i) \right]^2 K_{h_1}(X_t(s_i) - x), \quad (4.3.19)$$

w.r.t. $a_j, b_j, j = 0, 1, \dots, p+q$, where $K_{h_1}(x) = h_1^{-1}K(x/h_1)$, with K the kernel function on \mathbb{R}^1 and $h_1 > 0$ denoting the temporal bandwidth.

Then the local linear estimators can be expressed as

$$\hat{\beta}(x, s_i) = \bar{e}(A_i(x)' B_i(x) A_i(x))^{-1} A_i(x)' B_i(x) Y(s_i), \quad (4.3.20)$$

where $\bar{e} = (1, 0) \otimes \text{diag}(\mathbf{1}_{p+q+1})$. $A_i(x)$ denotes an $T_0 \times 2(p+q+1)$ matrix with $(Z_t(s_i)', (X_t(s_i) - x)Z_t(s_i)')$ as its $t - r$ th row for $t = r+1, \dots, T$, and

$$B_i(x) = \text{diag} \left\{ K_{h_1}(X_{r+1}(s_i) - x), \dots, K_{h_1}(X_T(s_i) - x) \right\}.$$

Step 2: (Spatial smoothing) The estimators based on Step-1 procedure can be improved by pooling the information from neighbouring locations by spatial smoothing (Lu et al., 2009). At a location $s \in S$, where S has the spatial sampling intensity function f_S (c.f., Assumptions in 4.8), the spatial smoothing estimators of $\beta(\cdot)$ can be obtained by

$$\tilde{\beta}(x, s) = \sum_{i=1}^N \hat{\beta}(x, s_i) \tilde{L}_{h_2, i}^*(s), \quad (4.3.21)$$

where $\tilde{L}_{h_2, i}^*(s) = \tilde{e}_1' (C'DC)^{-1} C'De_i$ denotes a weight function on \mathbb{R}^2 , which is a local linear fitting equivalent kernel, where $\tilde{e}_1 = (1, 0, 0)' \in \mathbb{R}^3$, e_i denotes a $N \times 1$ unit vector with 1 at the i -th position, C denotes an $N \times 3$ matrix with the i th-row $(1, (s_i - s)' / h_2)$, and $D = \text{diag} \{L_{h_2}(s_i - s)\}_{i=1}^N$ with $L_{h_2}(\cdot) = (h_2)^{-2}L(\cdot/h_2)$ and $L(\cdot)$ a kernel function on \mathbb{R}^2 .

Theorem 4.2. Under Assumptions in Section 4.8, for $s \in S$, as $T_0, N \rightarrow \infty$,

$$\tilde{\beta}(x, s) - \beta(x, s) - \frac{h_1^2}{2}\mu_1(x, s) - \frac{h_2^2}{2}\mu_2(x, s) = \left\{ (T_0 h_1 N h_2^2)^{-\frac{1}{2}} \Theta_1 + T_0^{-\frac{1}{2}} \Theta_2 \right\} \eta(s) (1 + o_p(1)),$$

where $\eta(s)$ is a $p + q + 1$ random vector with zero mean and identity variance matrix,

$$\mu_1(x, s) = \frac{\partial^2 \beta(x, s)}{\partial x^2} \int u^2 K(u) du,$$

$$\mu_2(x, s) = \left(\text{tr} \left\{ \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\}, \dots, \text{tr} \left\{ \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right)',$$

and Θ_1 and Θ_2 are two $(p + q + 1) \times (p + q + 1)$ matrices, satisfying

$$\Theta_1 \Theta_1' = \sigma^2(s) \frac{\Omega(x, s)^{-1}}{f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz,$$

$$\Theta_2 \Theta_2' = \Gamma(s, s) \frac{\Omega(x, s)^{-1} \Omega^*(x, s, s) \Omega(x, s)^{-1}}{f_X^2(x, s)} q(x, x; s)$$

with $\Omega(x, s) = E[Z_t(s) Z_t(s)' | X_t(s) = x]$, $\Omega^*(x, s_i, s_j) = E[Z_t(s_i) Z_t(s_j)' | X_t(s_i) = x, X_t(s_j) = x]$, $\Gamma(s_i, s_j) = E[\varepsilon_t(s_i) \varepsilon_t(s_j) | X_t(s_i) = x, X_t(s_j) = x]$ and $\sigma^2(s) = \text{Var}(\varepsilon_1(s))$.

Remark 1: The optimal bandwidth h_1 for time series estimation is similar to Cai et al. (2000), which is $h_1 = O(T_0^{-1/5})$. Then, by minimizing the squared bias plus variance of $\tilde{\beta}(x, s)$, we can find $h_2 = O((NT_0)^{-1/6})$.

Remark 2: If $h_1 N h_2^2 \rightarrow O(1)$ or $h_1 N h_2^2 \rightarrow \infty$, the rate of the convergence for $\tilde{\beta}(x, s)$ is $T_0^{\frac{1}{2}}$. Otherwise, the rate of the convergence for $\tilde{\beta}(x, s)$ is $(T_0 h_1 N h_2^2)^{\frac{1}{2}} = o_p(T_0^{\frac{1}{2}})$.

Remark 3: It is interesting to discuss the comparison between the one-step and the two-step procedures. With the same bandwidths h_1 and h_2 , we can find that the one-step and two-step methods have the same asymptotic bias and variances in Theorems 4.1 and 4.2. However, the optimal bandwidths $h_1 = O(T_0^{-1/5})$ and $h_2 = O((NT_0)^{-1/6})$ based on two-step method could not minimize the mean squared estimation error of $\beta(x, s)$. Compared with the two-step method, the one-step method has a more direct estimation procedure and could also obtain relatively efficient estimators for fixed T and infinite N . See the detail on the comparison of both methods in the simulations of Section 4.5.

4.4 Discussion

In real applications, one can often specify different spatial weight matrices based on different features of the real data, such as distance based spatial weight matrix W_1 or contiguity based spatial weight matrix W_2 , and so on (c.f., [Anselin, 2013](#)). A significant problem is which spatial weight matrix should be used. Therefore, the model selection procedures, using cross-validation method and Akaike Information Criterion with correction (AICc)(Section 4.3.3), can be employed to select an optimal weight matrix among the candidates. Moreover, we can also apply the combination of spatial weight matrices and estimate the model based on the procedure in Section 4.3. For simplicity, let us just look at two spatial weight matrices W_1 and W_2 well pre-specified (the case for more than two spatial weight matrices is easily extended). We here develop a new spatial weight matrix (W) by combining two individual row-wise standardised spatial weight matrices, W_1 and W_2 , say

$$W = a_1 W_1 + a_2 W_2, \quad (4.4.1)$$

with $a_1 + a_2 = 1$ for identifiability.

Therefore, we can extend model (4.2.2) with W as follows:

$$\begin{aligned} Y_t(s_i) = & \beta_0(X_t(s_i), s_i) + \sum_{j=1}^p \beta_j(X_t(s_i), s_i) \left(a_1 Y_{t-j}^{\text{SL}(1)}(s_i) + a_2 Y_{t-j}^{\text{SL}(2)}(s_i) \right) \\ & + \sum_{l=1}^q \beta_{p+l}(X_t(s_i), s_i) Y_{t-l}(s_i) + \varepsilon_t(s_i), \\ & i = 1, \dots, N, t = r + 1, \dots, T, \end{aligned} \quad (4.4.2)$$

where $\beta_j(x, s)$, $j = 0, 1, \dots, p + q$, are unknown functional coefficients as specified in model (4.2.2), and

$$Y_t^{\text{SL}(k)}(s_i) = \sum_{j=1}^N w_{ij,k} Y_t(s_j), \quad k = 1, 2,$$

with $w_{ij,k}$ is the (i, j) th component of the weight matrix W_k , satisfying $w_{ij,k} \geq 0$ and $\sum_{j=1}^N w_{ij,k} = 1$ for $k = 1, 2$. This idea of combining weight matrices, in our setting of nonlinear modelling, is different from, and more convenient than, the usual idea of spatial weights matrix selection and model averaging (c.f., [Zhang and Yu, 2018](#), in the linear model setting), which first need to establish different

models with individual spatial weight matrix W_k , respectively. Clearly, this latter idea is more involved from the model selection perspective. Actually, in our combining procedure, for example, if $(a_1, a_2) = (1, 0)$, then the model with W_1 is automatically selected, while if $(a_1, a_2) = (0, 1)$, so is the model with W_2 . In the usual case of $0 < a_1, a_2 < 1$, then it indicates that combining different features in W_1 and W_2 is important in our DyFAST models. We need to estimate the (a_1, a_2) based on the procedure in Section 4.3 as follows.

Step 1: We first consider the following model.

$$\begin{aligned} Y_t(s_i) = & \beta_0(X_t(s_i), s_i) + \sum_{j=1}^p \beta_j^{(1)}(X_t(s_i), s_i) Y_{t-j}^{\text{SL}(1)}(s_i) \\ & + \sum_{j=1}^p \beta_j^{(2)}(X_t(s_i), s_i) Y_{t-j}^{\text{SL}(2)}(s_i) + \sum_{l=1}^q \beta_{p+l}(X_t(s_i), s_i) Y_{t-l}(s_i) + \varepsilon_t(s_i), \end{aligned} \quad (4.4.3)$$

$$i = 1, \dots, N, t = r + 1, \dots, T,$$

where $\beta_j^{(k)}(X_t(s_i), s_i) = a_k \beta_j(X_t(s_i), s_i)$ for $k = 1, 2$. And we can follow the one-step estimation procedure in Section 4.3 to get the estimation results of $\beta_j^{(k)}(X_t(s_i), s_i)$, $j = 0, \dots, p + q$, $k = 1, 2$, $i = 1, \dots, N$, $t = r + 1, \dots, T$.

Step 2: As $a_1 + a_2 = 1$, based on (4.4.2) and (4.4.3), we have

$$\begin{aligned} \beta_j(X_t(s_i), s_i) &= \beta_j^{(1)}(X_t(s_i), s_i) + \beta_j^{(2)}(X_t(s_i), s_i), \quad j = 1, 2, \dots, p \\ \sum_{j=1}^p \sum_{i=1}^N \sum_{t=r+1}^T \beta_j^{(k)}(X_t(s_i), s_i) &= a_k \sum_{j=1}^p \sum_{i=1}^N \sum_{t=r+1}^T \beta_j(X_t(s_i), s_i), \end{aligned}$$

and can hence obtain the estimation of a_k by

$$\hat{a}_k = \frac{\sum_{j=1}^p \sum_{i=1}^N \sum_{t=r+1}^T \hat{\beta}_j^{(k)}(X_t(s_i), s_i)}{\sum_{l=1}^2 \sum_{j=1}^p \sum_{i=1}^N \sum_{t=r+1}^T \hat{\beta}_j^{(l)}(X_t(s_i), s_i)}, \quad k = 1, 2. \quad (4.4.4)$$

Then, we can obtain a new spatial weight matrix $W = a_1 W_1 + a_2 W_2$.

Step 3: Based on model (4.4.2), $\beta_j(x, s)$, $j = 0, 1, \dots, p + q$, can be estimated by using the DyFAST in Section 4.3 and new spatial weight matrix W .

This procedure is an alternative model selection method to the cross-validation method and Akaike Information Criterion with correction (AICc). In next section, we will consider using the simulation to evaluate this combination procedure.

4.5 Simulation Study

In this section, we will study the finite-sample performance of our proposed method in a simulation study. We will consider the spatial varying exogenous variable $X_t(s_i)$ in this section. For simplicity, we consider the following generative model:

$$Y_t(s_i) = \beta_0(X_t(s_i), s_i) + \sum_{j=1}^2 \beta_j(X_t(s_i), s_i) Y_{t-j}^{SL}(s_i) + \beta_3(X_t(s_i), s_i) Y_{t-1}(s_i) + \varepsilon_t(s_i), \quad (4.5.1)$$

where $Y_t^{SL}(s_i) = \sum_{k=1}^N W_{ik} Y_t(s_k)$, W_{ik} from a pre-specified spatial weight matrix W , satisfying $W_{ik} \geq 0$ and $\sum_{k=1}^N W_{ik} = 1$. As in the Section 4.6 below, $s_i = (u_i, v_i)$ is the centroid consisting of the latitude and longitude of the i th country, $i = 1, \dots, 23$, in the EU. The covariate process $X_t(s_i)$ follows $X_t(s_i) = \alpha(s_i) X_{t-1} + e_t(s_i)$, where $\alpha(s_i) = 0.9 + 0.05 \times \cos(u_i \times v_i)$ for $s_i = (u_i, v_i)'$ is the latitude and longitude of the i -th country, $e_t(s_i)$ follows i.i.d. $N(0, 1)$, innovation $\varepsilon_t(s_i)$ follows i.i.d. $N(0, 1)$.

$$\begin{aligned} \beta_0(X_t(s_i), s_i) &= 0.2 + 0.05 \times X_t(s_i) + b(s_i) \\ \beta_1(X_t(s_i), s_i) &= 0.2 + 0.1 \times \sin(X_t(s_i) + 1) + b(s_i) \\ \beta_2(X_t(s_i), s_i) &= 0.2 + 0.1 \times \cos(X_t(s_i) - 1) + b(s_i) \\ \beta_3(X_t(s_i), s_i) &= 0.3 + 0.1 \times \cos(X_t(s_i) + 1) + b(s_i) \end{aligned} \quad (4.5.2)$$

where $b(s_i) = 0.05 \sin(u_i \times v_i)$ for $s_i = (u_i, v_i)'$ is the latitude and longitude of the i -th country.

We generate data from model (4.5.1) as follows. At each location s_i for $i = 1, \dots, 23$, the initial values of $Y_0(s_i)$ are set to zero. Then we generate $Y_t(s_i)$ for $t = 1, 2, \dots, T + 50$. The first 50 time points are discarded and the next T time points are saved, denoted as $\{(X_t(s_i), Y_t(s_i))$ for $t = 1, \dots, T$, and $i = 1, \dots, N\}$. We consider three time series lengths: $T = 200$, $T = 400$ and $T = 600$. To assess the estimate of the unknown functions $\beta(x(s_i), s_i)$, we select 50 points of x from $(-2, 2)$ (approximately 10th and 90th percentiles of the covariate $X_t(s_i)$). The temporal bandwidth h_1 and the spatial bandwidth h_2 are 0.4 and 7, respectively.

The performance of estimation will be assessed by defining a squared estimation error (SEE) as a measure of the accuracy of estimation at a location s (c.f., Lu

et al., 2009). That is, for each location s , we define

$$\text{SEE}(\hat{\beta}_j(\cdot, s)) = \frac{1}{50} \sum_{k=1}^{50} \left\{ \hat{\beta}_j(x_k, s) - \beta_j(x_k, s) \right\}^2, \quad j = 0, 1, 2, 3 \quad (4.5.3)$$

where x_k for $k = 1, \dots, 50$ are 50 points that equally partition the interval between $(-2, 2)$. Similarly, to measure the accuracy of estimation for $\hat{\beta}_j(\cdot)$, we define

$$\text{SEE}(\hat{\beta}_j(\cdot)) = \frac{1}{50 \times 23} \sum_{i=1}^{23} \sum_{k=1}^{50} \left\{ \hat{\beta}_j(x_k, s_i) - \beta_j(x_k, s_i) \right\}^2, \quad j = 0, 1, 2, 3, \quad (4.5.4)$$

where N_s denotes the number of simulation times. We then consider the finite-sample performance of our proposed method by using the spatial varying exogenous variable $X_t(s_i)$.

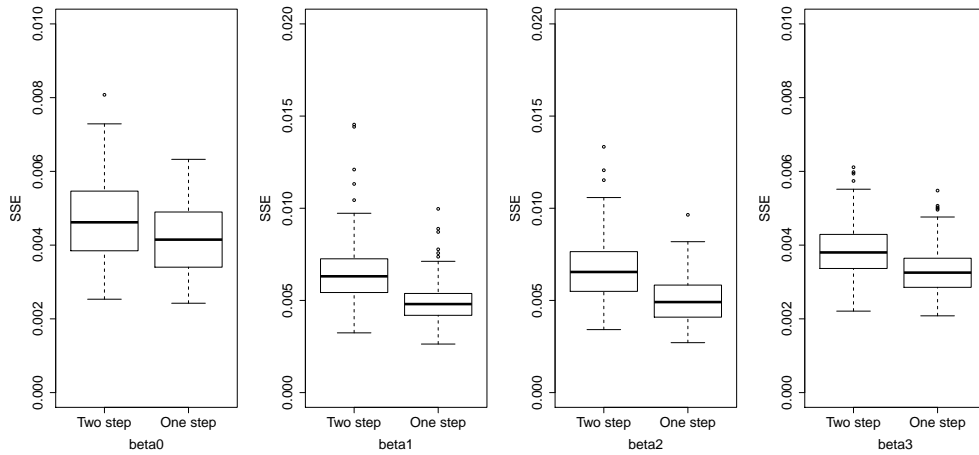


Figure 4.1: Boxplots of 100 times squared estimation error (SEE) for the estimation of $\beta_j(\cdot)$, $j = 0, 1, 2, 3$, for Two-step (left box) and One-step (right box) with time series length $T = 400$ in each plot.

The most commonly used specifications are binary contiguity matrix (W_1) and distance function matrices (W_2) (LeSage and Pace, 2009). For W_2 , units are considered as $w_{ij} = 1/d_{ij}$, where d_{ij} is the Euclidean distance between the centroids of two countries s_i and s_j , and $w_{jj} = 0$. Both spatial weight matrices are row-standardized weights matrices.

Firstly, we want to compare the estimation results of $\beta_j(\cdot)$ by One-step estimation and Two-step estimation procedures. We consider the spatial weight matrix as W_1 . We repeat the simulation 100 times with time series length $T = 400$, and thus have 100 values for each SEE of $\beta_j(\cdot)$, $j = 0, 1, 2, 3$, with the One-step and Two-step

methods, summarized in the boxplots in Figures 4.1. The results show that the One-step method has smaller SSE compared with Two-step method, indicating that the One-step method can get more accurate estimation results than Two-step method for large T and small N .

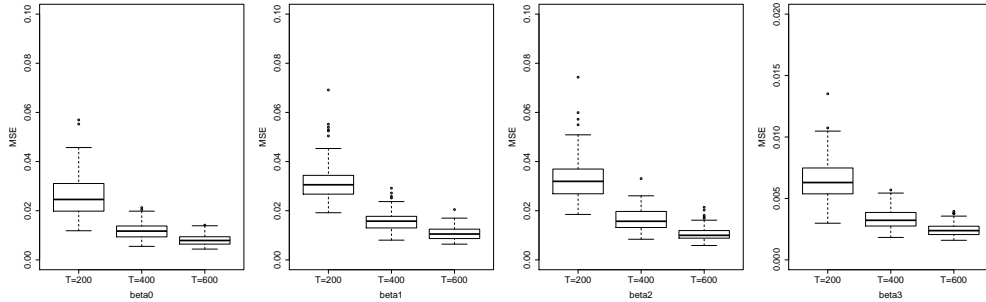


Figure 4.2: Boxplots of 100 times squared estimation error (SEE) for $\beta_j(\cdot)$, $j = 0, 1, 2, 3$ by the DyFAST, and for $T = 200$ time points (left box), $T = 400$ time series lengths (middle box) and $T = 600$ time series lengths (right box) in each plot.

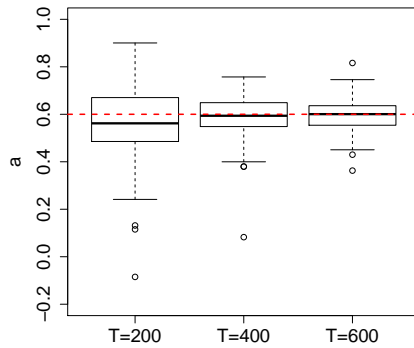


Figure 4.3: Boxplots of 100 times estimation of the combining weight a for $T = 200$ time points (left box), $T = 400$ time series lengths (middle box) and $T = 600$ time series lengths (right box) in each plot. The red dashed line shows the true value $a = 0.6$.

Secondly, we then consider three time series lengths: $T = 200$, $T = 400$, and $T = 600$. We repeat the simulation 100 times, and thus obtain 100 values for each SEE of $\beta_j(\cdot)$ using the DyFAST with One-step estimation method, $j = 0, 1, 2, 3$, summarized in the boxplots in Figures 4.2. As the sample size increasing, the accuracy of estimate for the $\beta_j(\cdot)$ by DyFAST method apparently improve, which indicates the properties of estimators are asymptotic efficiency.

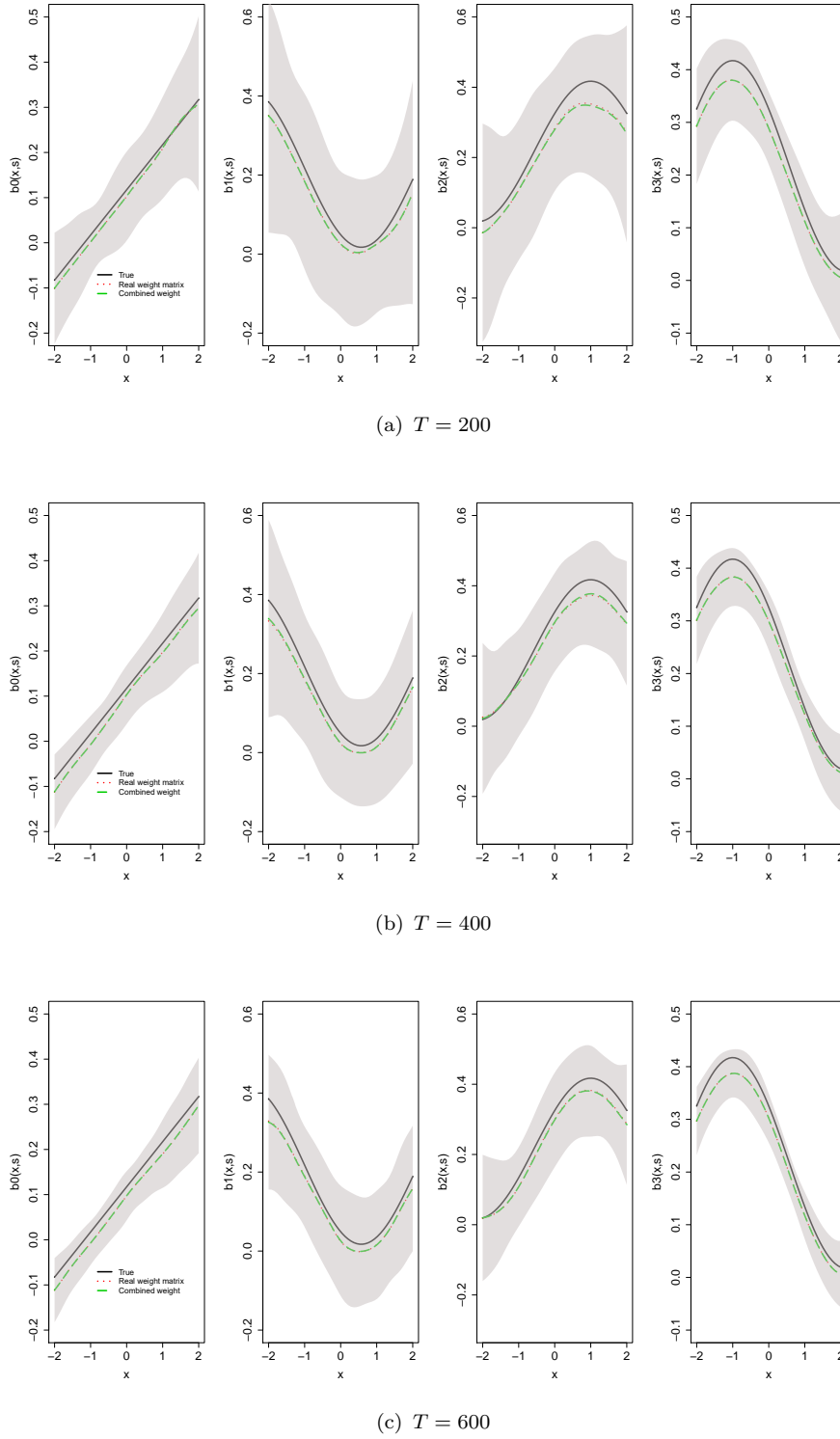


Figure 4.4: Simulation results: the mean in 100 replications at location $i = 1$ of the DyFAST with true spatial weight matrix (red lines) and with combined spatial weight matrix (green lines) of the coefficient function $\beta_j(\cdot, s_1)$, for $j = 0, 1, 2, 3$ under $T = 200$ time points (top plots), $T = 400$ time series lengths (middle plots) and $T = 600$ time series lengths (bottom plots). In each plot, the solid line shows the true coefficient function. Shaded areas correspond to 95% confidence intervals of estimation of DyFAST with combined spatial weight matrix.

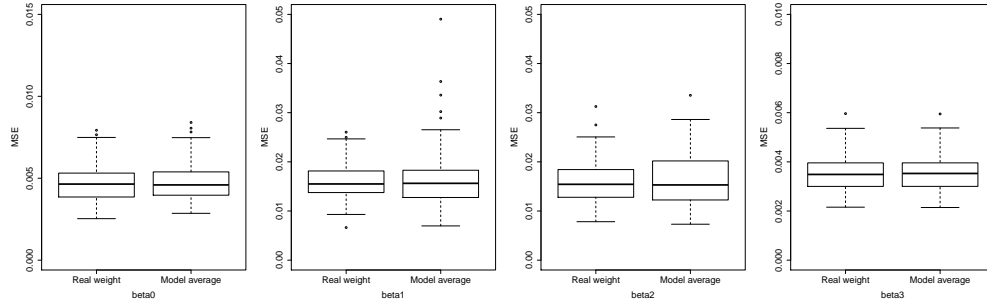


Figure 4.5: Boxplots of 100 times squared estimation error (SEE) of the DyFAST with real spatial weight matrix and combined spatial weight matrix for the estimation of $\beta_j(\cdot)$, for $j = 0, 1, 2, 3$ under $T = 400$ time series length. In each plot, the results the right box are using the true spatial weight matrix $W = 0.6W_1 + 0.4W_2$ and the results in the left box are considering the $\hat{W} = \hat{a}_1W_1 + \hat{a}_2W_2$.

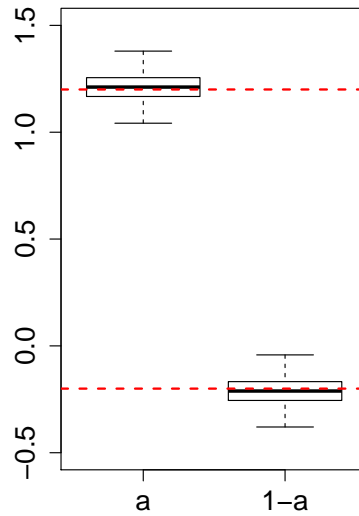


Figure 4.6: Boxplots of 100 times estimation of the combining weight a and $1 - a$ for $T = 400$ time series lengths. The red dashed line shows the true value $a = 1.2$ and $1 - a = -0.2$.

Finally, we then show that the combination procedure of spatial weight matrices in Section 4.4 is efficient and can be used to estimate the combining weight a . Here, we consider the true spatial weight matrix $W = 0.6W_1 + 0.4W_2$, and also consider three time series lengths: $T = 200$, $T = 400$, and $T = 600$. Figure 4.3 shows the squared estimation error (SEE) for the estimation of a , in $T = 200$, $T = 400$, and $T = 600$ time points, which appears acceptable. Figure 4.5 displays the squared estimation error (SEE) of the DyFAST with real spatial weight matrix and combined spatial weight matrix for the estimation of $\beta_j(\cdot)$, $j = 0, 1, 2, 3$ under $T = 400$ time series length. Figure 4.4 presents the mean estimated $\beta_j(\cdot)$, $j = 0, 1, 2, 3$ by DyFAST with true spatial weight matrix W and combined spatial weight matrix (W_1, W_2) , compared with the true curves in solid lines. We observe that our estimates $\beta_j(\cdot)$, $j = 0, 1, 2, 3$ mimic the corresponding true curves quite well. The estimation results indicate that with our combined spatial weight matrix, our DyFAST method could still accurately estimate the coefficient functions. Moreover, we want to show the efficiency of estimation with the true spatial weight matrix containing negative combining weight ($W = 1.2W_1 - 0.2W_2$). Figure 4.6 shows the squared estimation error (SEE) for the estimation of a and $1 - a$ for $T = 400$ time series lengths. These results indicate that the proposed procedure could also accurately identify the negative combining weight for the candidate spatial weight matrices. In sum, these findings indicate that our proposed DyFAST method works very well.

4.6 Real Data Example

4.6.1 Background

Retail gasoline prices are considerably fluctuating over time and different across countries within the Europe, due to many different factors, such as environmental policy, cost of crude oil, and macroeconomics. Numerous studies have examined the impact of crude oil price on retail gasoline prices, particularly focusing on the asymmetries in price transmission (c.f. Wlazlowski et al., 2009; Clerides et al., 2010; Blair et al., 2017; Bettendorf et al., 2009; Atil et al., 2014). The large majority of these literature, however, has neglected the possible cross-market dependency in retail gasoline markets (Wlazlowski et al., 2009). Note that the price disequilibria of retail gasoline in neighbour countries can cause “fuel travels” and create significantly cross-national spatial price spillovers in EU countries (Banfi

et al., 2005; Wlazlowski et al., 2009). Negligence of the spatial dependency in retail gasoline markets may cause the biased result in estimating the effect of crude oil price on the retail gasoline prices. To better measure the price transmission of retail gasoline in the EU, we will consider our DyFAST in this study.

4.6.2 Data

Table 4.1: Descriptive statistics for data.

	Mean	Stdev	ADF
Austria	1074.374	67.7683	0.9048
Belgium	1129.325	65.8963	0.9163
Bulgaria	1038.685	107.6421	0.6884
Czech.Republic	1077.506	80.3771	0.8801
Germany	1130.429	73.153	0.9008
Denmark	1222.557	75.0411	0.9348
Estonia	1056.022	59.5608	0.5301
Spain	1065.106	76.8082	0.9051
Finland	1252.319	75.4393	0.9699
France	1127.715	60.0909	0.8572
Greece	1123.077	75.5325	0.8575
Hungary	1112.495	81.8031	0.9698
Italy	1342.72	82.5838	0.8749
Lithuania	1009.299	85.6321	0.6109
Luxembourg	971.863	70.1195	0.9667
Latvia	1004.726	76.0957	0.9289
Netherlands	1183.517	76.3279	0.8308
Poland	1008.342	88.1633	0.9015
Portugal	1159.995	56.5434	0.7757
Romania	1131.182	82.3587	0.8656
Sweden	1357.833	55.4936	0.7956
Slovenia	1125.709	72.286	0.7642
Slovakia	1084.011	73.062	0.7509
Brent oil	47.9548	8.4755	0.7826

Note: This table displays summary statistics for weekly Brent crude oil price and regional diesel prices. The sample period is from 26/01/2015 to 19/12/2016. ADF is the p-value of augmented Dickey-Fuller test for unit root, which alternative hypothesis is stationary.

Our data consist of panel data of weekly spot prices for Brent crude oil price (X_t) and diesel prices (Y_{it}) in 23 European countries from 26/01/2015 to 19/12/2016. The data are obtained from European Commission³. All commodity prices are expressed in EUR and the diesel prices are inclusive of duties and taxes (Wlazlowski

³Source from: <https://ec.europa.eu/energy/en/data-analysis/weekly-oil-bulletin>

et al., 2009). The descriptive statistics for the raw data are reported in Table 4.1. The ADF test results indicate that all series are nonstationary. Therefore, we compute the Brent crude oil price return (X_t) and the diesel prices return (Y_{it}) by taking the difference in the logarithm of two consecutive weekly prices.

4.6.3 Result

In estimating the spatial neighbouring effect for diesel prices, there are different weight matrices. In this study, for example, we consider pre-specified spatial weight matrices W_1 and W_2 , two distance based spatial weight matrices. Let centroid distances from each spatial unit i to all other units $j \neq i$ be ranked as follows: $d_{ij(1)} \leq d_{ij(2)} \leq \dots \leq d_{ij(n-1)}$, where d_{ij} is the Euclidean distance between the centroids of two countries s_i and s_j . Then for each $k = 1, \dots, n-1$, the set $N_k(i) = \{j(1), j(2), \dots, j(k)\}$ contains the k closest units to i . Then for W_1 , we consider $k = 2$, and the set $N_2(i) = \{j(1), j(2)\}$ contains the 2 closest units to i . The spatial weights are set as $w_{ij} = 1/d_{ij}, j \in N_2(i)$, and $w_{ij} = 0$ otherwise. For W_2 , the elements are defined as $w_{ij} = 1/d_{ij}$, for $j \in N_{22}(i)$, and $w_{ij} = 0$ otherwise. All spatial weight matrices are row-standardized weights matrices. For the case with more spatial weight matrices, we just need to combine our combined weight matrix $W = a_1W_1 + a_2W_2$ with the new spatial weight matrices.

Table 4.2: Akaike Information Criterion with correction (AICc)

$\begin{matrix} \backslash & q \\ p & \end{matrix}$	1	2	3	4	5
1	1.6994	1.6975	1.6980	1.6948	1.7195
2	1.6905	1.7061	1.7099	1.7082	1.7380
3	1.6970	1.6912	1.7104	1.7152	1.7444
4	1.6956	1.6990	1.7118	1.7237	1.7473
5	1.7360	1.7270	1.7377	1.7484	1.7546

First, we can use the DyFAST to estimate the spatial neighbouring effect for diesel price returns. The DyFAST with W_1, W_2 is as follow.

$$\begin{aligned}
 Y_t(s_i) = & \beta_0(X_t, s_i) + \sum_{j=1}^p \beta_j(X_t, s_i) Y_{t-j}^{\text{SL}(1)}(s_i) + \sum_{j=1}^p \beta_{p+j}(X_t, s_i) Y_{t-j}^{\text{SL}(2)}(s_i) \\
 & + \sum_{l=1}^q \beta_{2p+l}(X_t(s_i), s_i) Y_{t-l}(s_i) + \varepsilon_t(s_i),
 \end{aligned} \tag{4.6.1}$$

The orders of temporal lagged autoregressive effects and temporal lagged spatial neighbouring effects are determined by Akaike Information Criterion with correction (AICc). Table 4.2 shows that $p = 2$ and $q = 1$ are the optimal orders. Moreover, the bandwidths are decided by cross validation, which are $h_1 = 3.23$ and $h_2 = 0.526$, respectively. Figure 4.7 shows the results of model (4.6.1). We can find that the top-middle ($\beta_1(X_t(s_i), s_i)$) and top-right ($\beta_2(X_t(s_i), s_i)$) sub-figures in Figure 4.7 fluctuate around zero. These findings indicate that the W_1 has much less impacts compared with W_2 . Then, we would like to further verify that for two pre-specified spatial weights matrices, which one is preferred?

Therefore, we will consider the DyFAST by combining spatial weight matrices W_1 and W_2 to see whether the combining weight a_1 is statistically significant from zero. The DyFAST with combining spatial weight matrices is of the form:

$$Y_t(s_i) = \beta_0(X_t(s_i), s_i) + \sum_{j=1}^2 \beta_j(X_t(s_i), s_i) \left(\sum_{k=1}^2 a_k Y_{t-j}^{\text{SL}(k)}(s_i) \right) + \sum_{l=1}^1 \beta_{2+l}(X_t(s_i), s_i) Y_{t-l}(s_i) + \varepsilon_t(s_i), \quad (4.6.2)$$

where a_1 and a_2 are two constants, satisfying $a_1 + a_2 = 1$. Moreover, the bandwidths are decided by cross validation, which are $h_1 = 3.294$ and $h_2 = 0.378$, respectively. Then, we can obtain the combining weight vector ($a_1 = -0.22$ and $a_2 = 1.22$) and functional coefficients $\beta_j(X_t(s_i), s_i)$, $j = 0, 1, 2$ (Figure 4.8). To test whether the a_k is statistically significant, we will consider the bootstrap approach. Given the observed sample $\{(X_t(s_i), Y_t(s_i)), t = 1, \dots, T \text{ and } i = 1, \dots, N\}$, the method proceeds as follows:

Algorithm (Bootstrap for DyFAST).

1. Let $\hat{\beta}(X_t(s_i), s_i)$ be defined as in (4.3.1), and obtain residuals $\hat{\varepsilon}_t(s_i) = Y_t(s_i) - Z_t(s_i)' \hat{\beta}(X_t(s_i), s_i)$ for $i = 1, \dots, N, t = r, \dots, T$ by DyFAST, where $r = \max(p, q)$, $W = \sum_{k=1}^2 a_k W_k$, $Z_t(s_i) = (1, Y_{t-1}^{\text{SL}}(s_i), \dots, Y_{t-p}^{\text{SL}}(s_i), Y_{t-1}(s_i), \dots, Y_{t-q}(s_i))'$ denoting the vector of spatio-temporally lagged variables, and $\beta(x, s) = (\beta_0(x, s), \beta_1(x, s), \dots, \beta_p(x, s), \beta_{p+1}(x, s), \dots, \beta_{p+q}(x, s))'$ is the corresponding vector of functional autoregressive coefficients.
2. Notice that the $\hat{\varepsilon}_t(s_i)$ estimated by our sample data appear to be independent across time but dependent across space. Therefore, we denote $\hat{\varepsilon}_t = (\hat{\varepsilon}_t(s_1), \dots, \hat{\varepsilon}_t(s_N))'$, and then generate bootstrap residuals $\varepsilon_1^*, \dots, \varepsilon_T^*$ randomly from the elements of $\hat{\varepsilon}_t$, $t = r, \dots, T$ with replacement.

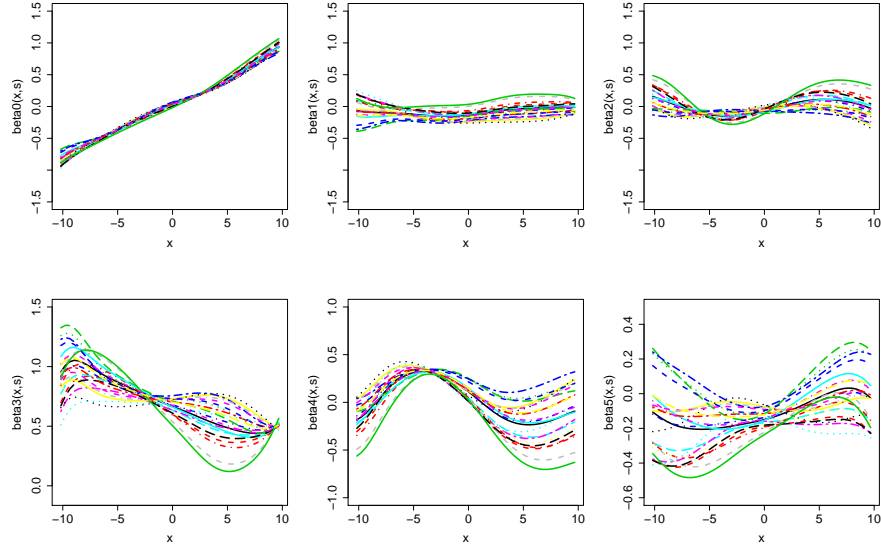


Figure 4.7: The impacts of Brent crude oil price return on diesel price returns in 23 EU countries by DyFAST with W_1 and W_2 .

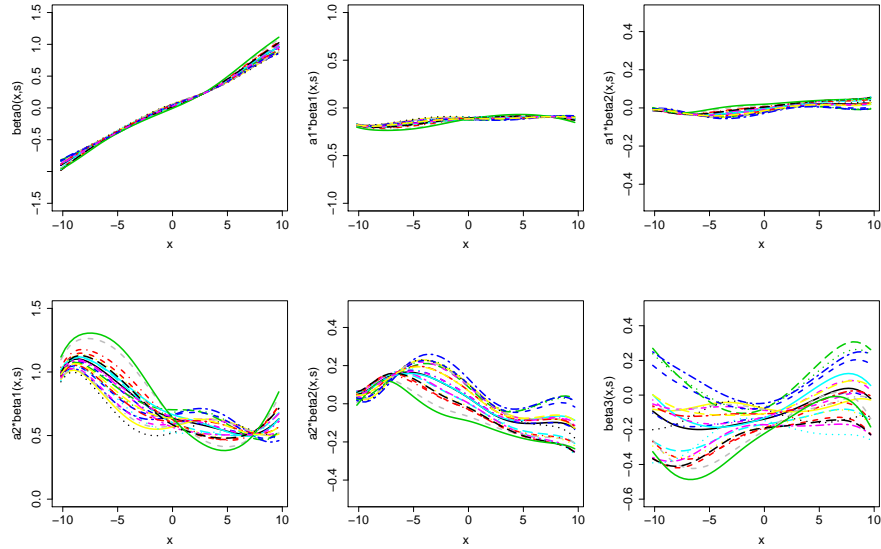


Figure 4.8: The impacts of Brent crude oil price return on diesel price returns in 23 EU countries by DyFAST with combined weight matrix.

3. Let the initial values of $Y_t^*(s_i)$, $t = 1, \dots, r$ equal to the values of $Y_t(s_i)$, $t = 1, \dots, r$. Then generate the bootstrap observation $Y_{r+1}^* = (Y_{r+1}^*(s_1), \dots, Y_{r+1}^*(s_N))'$ to $Y_T^* = (Y_T^*(s_1), \dots, Y_T^*(s_N))'$ through $Y_t^*(s_i) = Z_t^*(s_i)' \hat{\beta}(X_t(s_i), s_i) + \varepsilon_t^*(s_i)$, $t = r+1, \dots, T$, where $Z_t^*(s_i) = (1, Y_{t-1}^{SL*}(s_i), \dots, Y_{t-p}^{SL*}(s_i), Y_{t-1}^*(s_i), \dots, Y_{t-q}^*(s_i))'$ by iteration.
4. Finally, construct the estimators $\hat{\beta}(X_t(s_i), s_i)$ and a_k by using the bootstrap sample $\{(X_t(s_i), Y_t^*(s_i))\}$.

By repeating above bootstrap method 1000 times, we can obtain the t -value for a_1 and a_2 based on 1000 replications, which are -0.4956 and 2.7506 , respectively. This result shows that the a_1 is not statistically significant at 1% significance level, which indicates that our method selects W_2 as the optimal weight matrix ($a_1 = 0$ and $a_2 = 1$). The top-middle and top-right sub-figures in Figure 4.8 further describes that the $a_1\beta_1(X_t(s_i), s_i)$ and $a_1\beta_2(X_t(s_i), s_i)$ fluctuate around zero. Therefore, we consider W_2 to investigate the impacts of Brent crude oil price return on diesel price returns in 23 EU countries, which is as follow.

$$Y_t(s_i) = \beta_0(X_t, s_i) + \sum_{j=1}^2 \beta_j(X_t, s_i) Y_{t-j}^{SL(2)}(s_i) + \beta_3(X_t, s_i) Y_{t-1}(s_i) + \varepsilon_t(s_i), \quad (4.6.3)$$

Figure 4.9 shows the local linear estimates of $\beta_j(X_t, s_i)$ as a function of the weekly return of crude oil price for the i th country, where $j = 0, 1, 2, 3$. We can find varying coefficient with changing crude oil price return. In particular, for temporal lag-1 of spatial neighbouring effect, the impacts are positive. And, based on the temporal lag-2 of spatial neighbouring effect ($\beta_2(X_t(s_i), s_i)$), we can divide the countries into three groups, which is described in Table 4.3 and Figure 4.10. From Figure 4.10, we can find that the group 1 includes Belgium, Germany, Denmark, Luxembourg, Netherlands, Portugal and Sweden, which are located in North-Western Europe (except Portugal). The group 3 contains Bulgaria, Greece, Hungary and Romania, which are located in Balkans. The temporal lag effects in group 3 are positive and larger compared with group 1 and 2 (Figure 4.9). Moreover, temporal lag effects are negative in group 2, and near zero in group 1. These empirical findings can help governments and energy users to mitigate the negative impacts from the expected or unexpected fluctuations in the oil and the neighbouring retail gasoline markets and better manage energy risk. Moreover, local state governments may formulate relevant energy policies on the basis of their geographical location .

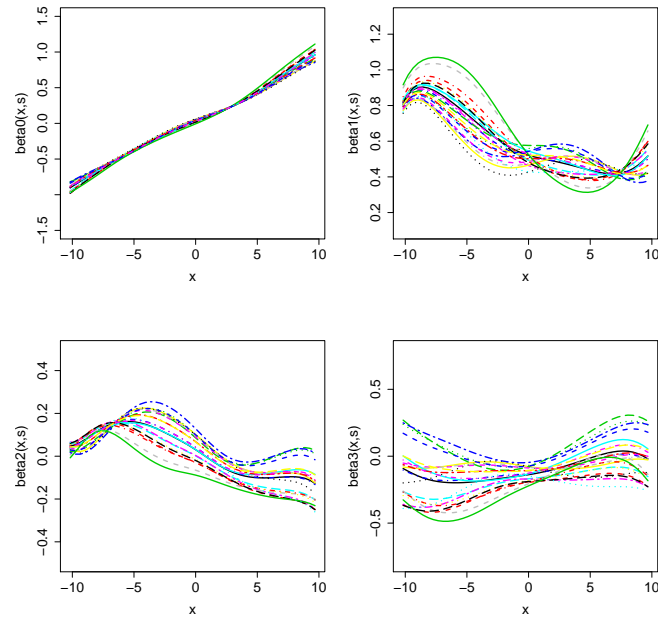


Figure 4.9: The impacts of Brent crude oil price return on diesel price returns in 23 EU countries by DyFAST with W_2 .

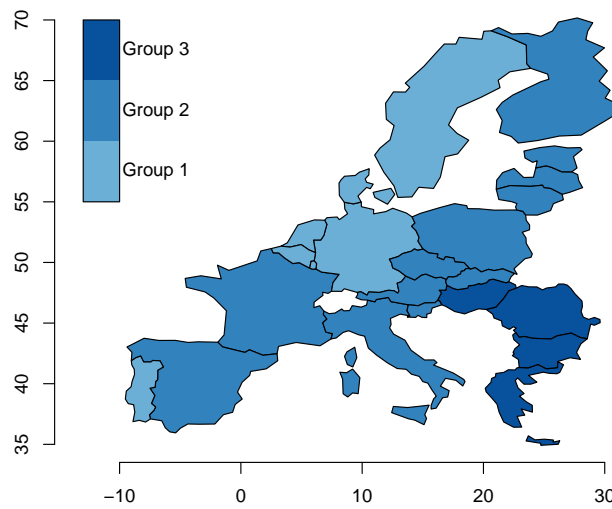


Figure 4.10: Three EU country groups based on the results of DyFAST with W_2 .

Table 4.3: Country Groups

Group 1	Belgium	Germany	Denmark	Luxembourg	Netherlands	Portugal	Sweden	
Group 2	Austria	Czech.Republic	Estonia	Spain	Finland	France	Italy	Lithuania
	Latvia	Poland	Slovenia	Slovakia				
Group 3	Bulgaria	Greece	Hungary	Romania				

Note: The groups are classified based on Figure 4.9.

To further evaluate our methods, we can compare the prediction performances of linear model and DyFAST. In addition, due to the fact that nonparametric model may not provide optimal prediction, we also consider the prediction results of β with nonlinear threshold forms. Therefore, based on the Figure 4.9 and 4.10, we construct the nonlinear threshold functions of $\beta_j(x, s)$, $j = 0, 1, 2, 3$ for 3 groups as following.

- Group 1

$$\begin{aligned}
\beta_0(x, s) &= b_0^0(s) + b_1^0(s)x \\
\beta_1(x, s) &= \{b_{10}^1(s) + b_{11}^1(s)x\} \mathcal{I}(x < -8) + \{b_{20}^1(s) + b_{21}^1(s)x\} \mathcal{I}(-8 \leq x) \\
\beta_2(x, s) &= \{b_{10}^2(s) + b_{11}^2(s)x\} \mathcal{I}(x < -5) + \{b_{20}^2(s) + b_{21}^2(s)x\} \mathcal{I}(-5 \leq x) \\
\beta_3(x, s) &= \{b_{10}^3(s) + b_{11}^3(s)x\} \mathcal{I}(x < 8) + \{b_{20}^3(s) + b_{21}^3(s)x\} \mathcal{I}(8 \leq x)
\end{aligned}$$

- Group 2

$$\begin{aligned}
\beta_0(x, s) &= b_0^0(s) + b_1^0(s)x \\
\beta_1(x, s) &= \{b_{10}^1(s) + b_{11}^1(s)x\} \mathcal{I}(x < -7) + \{b_{20}^1(s) + b_{21}^1(s)x\} \mathcal{I}(-7 \leq x < 0) \\
&\quad + \{b_{30}^1(s) + b_{31}^1(s)x\} \mathcal{I}(0 \leq x < 8) + \{b_{40}^1(s) + b_{41}^1(s)x\} \mathcal{I}(8 \leq x) \\
\beta_2(x, s) &= \{b_{10}^2(s) + b_{11}^2(s)x\} \mathcal{I}(x < -5) + \{b_{20}^2(s) + b_{21}^2(s)x\} \mathcal{I}(-5 \leq x < 5) \\
&\quad + \{b_{30}^2(s) + b_{31}^2(s)x\} \mathcal{I}(5 \leq x) \\
\beta_3(x, s) &= \{b_{10}^3(s) + b_{11}^3(s)x\} \mathcal{I}(x < -7) + \{b_{20}^3(s) + b_{21}^3(s)x\} \mathcal{I}(-7 \leq x < 0) \\
&\quad + \{b_{30}^3(s) + b_{31}^3(s)x\} \mathcal{I}(0 \leq x < 8) + \{b_{40}^3(s) + b_{41}^3(s)x\} \mathcal{I}(8 \leq x)
\end{aligned}$$

- Group 3

$$\begin{aligned}\beta_0(x, s) &= b_0^0(s) + b_1^0(s)x \\ \beta_j(x, s) &= \{b_{10}^j(s) + b_{11}^j(s)x\} \mathcal{I}(x < -3) + \{b_{20}^j(s) + b_{21}^j(s)x\} \mathcal{I}(-3 \leq x < 3) \\ &\quad + \{b_{30}^j(s) + b_{31}^j(s)x\} \mathcal{I}(3 \leq x), \quad j = 1, 2 \\ \beta_3(x, s) &= \{b_{10}^3(s) + b_{11}^3(s)x\} \mathcal{I}(x < 0) + \{b_{20}^3(s) + b_{21}^3(s)x\} \mathcal{I}(0 \leq x)\end{aligned}$$

where $I(\cdot)$ is an indicator function, $b^l(s)$, $l = 0, 1, 2, 3$, are the spatially-varying linear coefficients.

Table 4.4: Comparison of forecasting results I

	Model (4.6.1)	Model (4.6.2)	Model (4.6.3)
$T = 15$	1.649	1.570	1.566

Note: This table shows the mean squared prediction error (MSPE) of the one-step ahead prediction for DyFAST with W_1, W_2 (Model (4.6.1)), DyFAST with combined spatial weight matrix (Model (4.6.2)) and DyFAST with W_2 (Model (4.6.3)).

Table 4.5: Comparison of forecasting results II

	Linear model	Nonlinear threshold model
$T = 15$	1.875	1.519

Note: This table shows the mean squared prediction error (MSPE) of the one-step ahead prediction for traditional linear model and DyFAST with nonlinear threshold functions.

We set aside the last 15 data for evaluation and use first $T = 76$ data as training set. Then, we can obtain the mean squared prediction error (MSPE) of the one-step ahead prediction for DyFAST with W_1, W_2 (Model (4.6.1)), DyFAST with combined spatial weight matrix (Model (4.6.2)) and DyFAST with W_2 (Model (4.6.3)), which are 1.649, 1.570 and 1.566, respectively. Compared to the DyFAST with W_1, W_2 , the performance under DyFAST with W_2 outperforms in prediction, with a relative improvement of 5%. This result also indicates that our method can appropriately estimate the underlying relationships in empirical analysis by selecting the optimal weight matrix. Moreover, the MSPE values for traditional linear model and DyFAST with nonlinear threshold functions are 1.875 and 1.519, respectively. The prediction result shows that the threshold parametrization of $\beta(x, s)$ performs the best in prediction, with a relative improvement of 19% compared with traditional linear model. These results further illustrate that DyFAST with model averaging can help to uncover the effect of crude oil price return on

retail gasoline price returns, which is related to the crude oil market price fluctuations.

4.7 Conclusions

In real estate markets, energy markets and environmental sciences, the real structures of some spatial and temporal data may not be linear and stationary. This chapter proposes a semiparametric family of *Dynamic Functional-coefficient Autoregressive Spatio-Temporal (DyFAST)* models to overcome the difficulty in non-linear modelling and analysis of spatio-temporal data with irregularly observed locations and location-wide non-stationarity. Moreover, both one-step and two-step estimation methods are proposed to estimate the functional coefficients in DyFAST models. Both theoretical properties and Monte Carlo simulations are investigated. Our model is applied to investigate the spatial dependency in retail gasoline markets in EU. The results illustrate that DyFAST can help to uncover the effect of crude oil price return on retail gasoline price returns, which is more complex than linear.

4.8 Appendix

A1: Regularity Conditions

Suppose that a spatial sampling locations set $S = \{s_1, \dots, s_N\}$ has a sampling density function $f_S(s)$ in the spatial domain $S \subset \mathbb{R}^2$.

For stationary process $\mathbf{X}_t = (X_t(s_1), \dots, X_t(s_N))$, $t = 0, \pm 1, \pm 2, \dots$, if

$$\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |P(A)P(B) - P(AB)| \longrightarrow 0, \quad \text{as } k \rightarrow \infty$$

where \mathcal{F}_i^j is σ -algebra generated by $\{\mathbf{X}_t\}_{i \leq t \leq j}$. We call \mathbf{X}_t the α -mixing process.

In the following, we will first provide the regularity assumptions for Lemmata and Theorems.

Assumption:

(A1) (i) For $t = 0, \pm 1, \pm 2, \dots$, $\mathbf{X}_t = (X_t(s_1), \dots, X_t(s_N))$ is α -mixing process

(ii) For some real positive number $\delta > 2$ and $c > 1 - 2/\delta$, the α -mixing coefficient $\alpha(\cdot)$ satisfies

$$\lim_{k \rightarrow \infty} k^c \sum_{k=d_N}^{\infty} \{\alpha(k)\}^{1-2/\delta} = 0$$

.

(iii) $X_i(s)$ and $X_j(s)$ have a joint probability density function $p(x_i, x_j; s)$, which is continuous and bounded for all $i \neq j$. For each t , the joint probability density function of $X_t(s_1)$ and $X_t(s_2)$ satisfies the following limit: $\lim_{s_1, s_2 \rightarrow s} p(x_1, x_2; s_1, s_2) = q(x_1, x_2; s)$, where $q(x_1, x_2; s)$ is continuous and bounded with respect to both x_1 and x_2 .

(iv) The marginal density of X given location s , $f_X(x, s)$, is continuous, and $0 < f_X(x, s) \leq M < \infty$.

(A2) The functions $\beta_j(x, s)$, $j = 1, \dots, p + q + 1$ are continuous at all x and twice differentiable w.r.t. x and $s \in S$.

- (A3) (i) At each location $s \in S$, the innovations $\varepsilon_t(s)$ are *i.i.d.* random variables independent of $\{X_t(s)\}$. Moreover, for each $t > r$, $\{\varepsilon_t(s)\}_{s \in S}$ are independent of $\{Y_{t-i}^{\text{SL}}(s)\}_{s \in S}$ for $i = 1, \dots, p$, and $\{Y_{t-l}(s)\}_{s \in S}$ for $l = 1, \dots, q$. $E[|\varepsilon_t(s)|^{2+\delta}] < \infty$ and $E[|Y_t(s)|^{2+\delta}] < \infty$, for some $\delta > 2$.
- (ii) For each $t > r$, the spatial covariance function $\Gamma_t(s_l, s_k) \equiv \text{Cov}[\varepsilon_t(s_l), \varepsilon_t(s_k)]$ is bounded over $S \times S$.
- (iii) For each $t \geq r + 1$ and $s \in S$, $\varepsilon_t(s) = \varepsilon_{1,t}(s) + \varepsilon_{2,t}(s)$, where $\{\varepsilon_{1,t}(s)\}$ and $\{\varepsilon_{2,t}(s)\}$ are two independent processes and both satisfy the condition (A3)(i). Further, $\Gamma_{1t}(s_j, s_k) \equiv \text{Cov}[\varepsilon_{1,t}(s_1), \varepsilon_{1,t}(s_2)]$ is continuous in (s_1, s_2) and $\Gamma_{2t}(s_1, s_2) \equiv \text{Cov}[\varepsilon_{2,t}(s_1), \varepsilon_{2,t}(s_2)] = 0$ if $s_1 \neq s_2$ and $\Gamma_{2t}(s_1, s_2) = \sigma_2^2(s_j) > 0$ is continuous in s_1 .
- (iv) For each t , the $\Omega(x, s_1, s_2) = E[Z_t(s_1)Z_t(s_2)|x] = \Omega_1(x, s_1, s_2) + \Omega_2(x, s_1, s_2)$, where $\Omega_1(x, s_1, s_2)$ is continuous in (s_1, s_2) , and $\Omega_2(x, s_1, s_2) = 0$ if $s_1 \neq s_2$ and $\Omega_2(x, s_1, s_1) = \Omega_2(x, s_1) > 0$ is continuous in s_1 .
- (A4) For any measurable set $A \subset S \subset \mathbb{R}^2$, $N^{-1} \sum_{j=1}^N \mathcal{I}(s_j \in A) \longrightarrow \int_A f_S(s) ds$, $N \rightarrow \infty$, where the density function f satisfies $f > 0$ in a neighbourhood of $s \in S$.
- (A5) (i) The kernel function $K(\cdot)$ is symmetric, uniformly bounded by some constant, and integrable. Further, $\int K(u) du = 1$ and $\int u^2 K(u) du < \infty$.
- (ii) The kernel function $L(\cdot)$ is symmetric and satisfies $\int_{\mathbb{R}^2} L(z) dz = 1$, $\int_{\mathbb{R}^2} zL(z) dz = 0$ and $\int_{\mathbb{R}^2} zz' L(z) dz < \infty$.
- (A6) (i) The temporal bandwidth $h_1 \rightarrow 0$ in such a way that $T_0 h_1 \rightarrow \infty$. The spatial smoothing bandwidth $h_2 \rightarrow 0$, $N h_2^2 \rightarrow \infty$, as $N \rightarrow \infty$.
- (ii) There exists a sequence of positive integers q_T such that $q_T \rightarrow \infty$, $q_T = o\left((T_0 h_1)^{1/2}\right)$, and $(T_0/h_1)^{1/2} \alpha(q_T) \rightarrow 0$, as $T_0 \rightarrow \infty$.
- (iii) $T_0^{1/2-\delta/4} h_1^{\delta/\delta^*-1/2-\delta/4} h_2^{2\delta/\delta^*-\delta} = O(1)$ where $\delta^* > \delta > 2$.
- (A7) $\Omega(x, s)$ is positive definite and continuous in a neighbourhood of x and s .

Assumption (A1)(i)&(ii) assume that the process $X_t(s_i)$ is α -mixing, which is standard in nonlinear time series analysis. Assumption (A1)(iii) assumes the boundedness of the joint probability density functions, which can ensure the consistency w.r.t. the different times and locations. Assumption (A1)(iv) assumes marginal density of $X_t(s_i)$. Assumption (A2) is a standard condition on the estimated function in nonlinear modelling, which ensures the smoothness of the estimated function. Assumption (A3) assumes the independence of innovation, boundedness

of spatial covariance function, and the finite moment of $Y_t(s)$. Such moment conditions are usually necessary for establishing asymptotic normality in nonparametric modelling. See, for example, Assumption (C4) of Al-Sulami et al. (2017b). For normal distributions (e.g., $\varepsilon_t(s) \sim N(0, 1)$) and t -distributions with $v > 4$ degrees of freedom, these conditions can be satisfied. But if the errors have heavier tails, say, Cauchy distributions with the probability density function $f(x) = \frac{1}{\pi(1+x^2)}$ not having any moments satisfying A3(i), or the density function $f(x) = 3(2+x^2)^{-\frac{5}{2}}$ not having the fourth moment. Assumption (A4) assumes the spatial sampling density, which is a standard condition in previous literature (e.g., Lu et al., 2009; Al-Sulami et al., 2017a). This indicates that the asymptotics belong to the category of the fixed-domain or in-fill asymptotics. Assumption (A5) assumes standard conditions for the kernel functions, which are commonly used in nonparametric kernel estimation. Assumption (A6)(i) assumes the conditions on the bandwidths h_1 and h_2 , which are standard in nonparametric time series or spatial analysis. Assumption (A6)(ii)&(iii) are mainly required to establish the asymptotic normality. Assumption (A7) is the necessary and sufficient condition to ensure that $\beta(x, s)$ is identified. These conditions imposed on the spatio-temporal series in (A1)–(A7) are fairly mild and used in the literature (c.f. Cai et al., 2000; Fan and Yao, 2003; Lu et al., 2009; Al-Sulami et al., 2017a).

A2: Proof for Section 4.3

Lemma A.1. (i) Let $\mathcal{L}_r(\mathcal{F})$ denote the class of \mathcal{F} -measurable random variables ξ satisfying $\|\xi\|_r := (E|\xi|^r)^{1/r} < \infty$. Let $X \in \mathcal{L}_r(\mathcal{F}_1)$ and $Y \in \mathcal{L}_s(\mathcal{F}_2)$. Then, for any $1 \leq r, s, h < \infty$ such that $r^{-1} + s^{-1} + h^{-1} = 1$,

$$\left| E[XY] - E[X]E[Y] \right| \leq C \|X\|_r \|Y\|_s [\alpha(\mathcal{F}_1, \mathcal{F}_2)]^{1/h}, \quad (\text{A.1})$$

where $\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup_{A \in \mathcal{F}_1, B \in \mathcal{F}_2} |P(AB) - P(A)P(B)|$.

(ii) If moreover $|X|$ and $|Y|$ are P-a.s. bounded, the right-hand side of (A.1) can be replaced with $C\alpha(\mathcal{F}_1, \mathcal{F}_2)$.

Proof: This lemma is borrowed from Deo (1973), where we refer to for a proof.

Lemma A.2. *Under Assumptions (A4) and (A6), for any function $\varphi(s_1, s_2)$ defined on $\mathbb{R}^2 \times \mathbb{R}^2$, which is continuous at (s_0, s_0) , we have*

$$\sum_{i=1}^N \sum_{j=1, j \neq i}^N \varphi(s_i, s_j) L_{h_2}(s_i - s_0) L_{h_2}(s_j - s_0) = N^2 \varphi(s_0, s_0) f^2(s_0) (1 + o(1)) \quad (\text{A.2})$$

Proof: This lemma is borrowed from [Lu et al. \(2009\)](#), where we refer to for a proof.

Proof of Lemma 4.1. This lemma easily follows from [Lu \(2001\)](#). The detail is omitted.

Proof of Lemma 4.2.

We first consider the $\text{EW}_{TN,1}$.

$$\text{EW}_{TN,1} := \frac{1}{N h_1 h_2^2} \sum_{i=1}^N \text{E} \varsigma_{i1}, \quad (\text{A.3})$$

where

$$\varsigma_{i1} = Y_1^*(s_i) Z_1(s_i) K\left(\frac{X_1(s_i) - x}{h_1}\right) L\left(\frac{s_i - s}{h_2}\right), \quad (\text{A.4})$$

and

$$Y_1^*(s_i) = m(X_1(s_i), Z_1(s_i)) + \varepsilon_1(s_i), \quad (\text{A.5})$$

$$\begin{aligned} & m(X_1(s_i), Z_1(s_i)) \\ &= \frac{1}{2} (X_1(s_i) - x)^2 Z_1(s_i)' \frac{\partial^2 \beta(\xi_1, \xi_2)}{\partial \xi_1^2} + (X_1(s_i) - x) Z_1(s_i)' \frac{\partial^2 \beta(\xi_1, \xi_2)}{\partial \xi_1 \partial \xi_2'} (s_i - s) \\ & \quad + \frac{1}{2} \sum_{j=0}^{p+q} (s_i - s)^j \frac{\partial^2 \beta_j(\xi_1, \xi_2)}{\partial \xi_2 \partial \xi_2'} (s_i - s) Z_{1,j}, \end{aligned} \quad (\text{A.6})$$

where $\xi_1 = x + \eta_1 (X_1(s_i) - x)$ with $|\eta_1| < 1$ and $\xi_2 = s + \eta_2 (s_i - s)$ with $|\eta_2| < 1$.

$$\begin{aligned}
& \frac{1}{Nh_1h_2^2} \sum_{i=1}^N E_{\zeta_{i1}} \\
&= \frac{1}{Nh_1h_2^2} \sum_{i=1}^N E \left[Z_1(s_i) Y_1^*(s_i) K\left(\frac{X_1(s_i) - x}{h_1}\right) L\left(\frac{s_i - s}{h_2}\right) \right] \\
&= \frac{1}{Nh_1h_2^2} \sum_{i=1}^N E \left[Z_1(s_i) Y_1^*(s_i) L\left(\frac{s_i - s}{h_2}\right) K\left(\frac{X_1(s_i) - x}{h_1}\right) \right] \\
&= \frac{1}{Nh_1h_2^2} \sum_{i=1}^N E \left[Z_1(s_i) L\left(\frac{s_i - s}{h_2}\right) K\left(\frac{X_1(s_i) - x}{h_1}\right) \left(\frac{1}{2} (X_1(s_i) - x)^2 Z_1(s_i)' \frac{\partial^2 \beta(\xi_1, \xi_2)}{\partial \xi_1^2} \right. \right. \\
&\quad \left. \left. + (X_1(s_i) - x) Z_1(s_i)' \frac{\partial^2 \beta(\xi_1, \xi_2)}{\partial \xi_1 \partial \xi_2'} (s_i - s) \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \sum_{j=0}^{p+q} (s_i - s)' \frac{\partial^2 \beta_j(\xi_1, \xi_2)}{\partial \xi_2 \partial \xi_2'} (s_i - s) Z_{1,j} + \varepsilon_t(s_i) \right) \right] \\
&= I_0 + I_1 + I_2.
\end{aligned} \tag{A.7}$$

By using Assumptions (A1) and (A5), we can obtain

$$\begin{aligned}
I_0 &= \frac{1}{Nh_1h_2^2} \sum_{i=1}^N E \left[Z_1(s_i) L\left(\frac{s_i - s}{h_2}\right) K\left(\frac{X_1(s_i) - x}{h_1}\right) \frac{1}{2} (X_1(s_i) - x)^2 Z_1(s_i)' \frac{\partial^2 \beta(\xi_1, \xi_2)}{\partial \xi_1^2} \right] \\
&= \frac{1}{Nh_1h_2^2} \sum_{i=1}^N E \left[Z_1(s_i) Z_1(s_i)' L\left(\frac{s_i - s}{h_2}\right) K\left(\frac{X_1(s_i) - x}{h_1}\right) \frac{1}{2} (X_1(s_i) - x)^2 \frac{\partial^2 \beta(x, s)}{\partial x^2} \right] (1 + o(1)) \\
&= \frac{1}{2} h_1^2 f_X(x, s) f_S(s) \Omega(x, s) \frac{\partial^2 \beta(x, s)}{\partial x^2} \int u^2 K(u) du + o(h_1^2).
\end{aligned} \tag{A.8}$$

Similarly, based on (A.7) and Assumptions (A1) and (A5), we can obtain

$$\begin{aligned}
I_1 &= o(h_1^2 + h_2^2), \\
I_2 &= \frac{1}{2} h_2^2 f_X(x, s) f_S(s) \Omega(x, s) \left(\text{tr} \left\{ \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\}, \dots, \right. \\
&\quad \left. \text{tr} \left\{ \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right)' + o(h_2^2).
\end{aligned} \tag{A.9}$$

Then, for $\mathcal{W}_{TN,j}$, $j = 2, 3$, based on Assumptions (A1) and (A5) and the symmetry of kernel functions, we can get

$$\lim_{N,T \rightarrow \infty} E(\mathcal{W}_{TN,j}) = 0, \quad j = 2, 3.$$

The proof of the lemma is completed.

Proof of Lemma 4.3.

We now derive the asymptotic variance, let

$$\begin{aligned} Q_{it} &= Y_t^*(s_i) Z_t(s_i) K\left(\frac{X_t(s_i) - x}{h_1}\right) L\left(\frac{s_i - s}{h_2}\right) \\ &= (m(X_t(s_i), Z_t(s_i)) + \varepsilon_t(s_i)) Z_t(s_i) K\left(\frac{X_t(s_i) - x}{h_1}\right) L\left(\frac{s_i - s}{h_2}\right) \\ &= Q_{it,1} + Q_{it,2}, \end{aligned} \quad (\text{A.10})$$

and $\zeta_{it} = Q_{it,1} - EQ_{it,1}$, $EQ_{it,2} = 0$, $\Delta_t = \sum_{i=1}^N \zeta_{it} + \sum_{i=1}^N Q_{it,2} = \Delta_{t,1} + \Delta_{t,2}$. Then $\mathcal{W}_{TN,1}$ can be expressed as

$$\mathcal{W}_{TN,1} = (T_0 N h_1 h_2^2)^{-1} \sum_{t=r+1}^T \sum_{i=1}^N Q_{it}, \quad (\text{A.11})$$

and therefore

$$\begin{aligned} \text{Var}(\mathcal{W}_{TN,1}) &= (T_0 N h_1 h_2^2)^{-2} \sum_{t=r+1}^T E(\Delta_t \Delta_t') + 2(T_0 N h_1 h_2^2)^{-2} \sum_{r+1 \leq i < j \leq T} E(\Delta_i \Delta_j') \\ &= T_0^{-1} (h_1 N h_2^2)^{-2} E(\Delta_1 \Delta_1') \\ &\quad + T_0^{-1} (h_1 N h_2^2)^{-2} \sum_{i=1}^{T-r-1} \left(1 - \frac{i}{T-r}\right) E(\Delta_1 \Delta_{1+i}') \end{aligned} \quad (\text{A.12})$$

For the first part in (A.12), we have

$$\begin{aligned} &T_0^{-1} (h_1 N h_2^2)^{-2} E(\Delta_1 \Delta_1') \\ &= T_0^{-1} (h_1 N h_2^2)^{-2} E((\Delta_{1,1} + \Delta_{1,2})(\Delta_{1,1} + \Delta_{1,2})') \\ &= T_0^{-1} (h_1 N h_2^2)^{-2} \{E(\Delta_{1,1} \Delta_{1,1}') + 2E(\Delta_{1,1} \Delta_{1,2}) + E(\Delta_{1,2} \Delta_{1,2}')\} \\ &= T_0^{-1} (h_1 N h_2^2)^{-2} E(\Delta_{1,1} \Delta_{1,1}') + T_0^{-1} (h_1 N h_2^2)^{-2} E(\Delta_{1,2} \Delta_{1,2}'). \end{aligned} \quad (\text{A.13})$$

First, we consider the second part in (A.13).

$$\begin{aligned}
 & T_0^{-1}(h_1 N h_2^2)^{-2} \mathbb{E} (\Delta_{1,2} \Delta'_{1,2}) \\
 &= T_0^{-1}(h_1 N h_2^2)^{-2} \mathbb{E} \left(\left(\sum_{i=1}^N Q_{i1,2} \right) \left(\sum_{i=1}^N Q_{i1,2} \right)' \right) \\
 &= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \mathbb{E} (Q_{i1,2} Q'_{i1,2}) + T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq l}^N \mathbb{E} (Q_{l1,2} Q'_{k1,2})
 \end{aligned} \tag{A.14}$$

By using the conditions in Assumption (A3), (A4) and (A5), we have

$$\begin{aligned}
 & T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \mathbb{E} (Q_{i1,2} Q'_{i1,2}) \\
 &= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \mathbb{E} \left[Z_1(s_i) Z_1(s_i)' \varepsilon_1^2(s_i) K^2 \left(\frac{X_t(s_i) - x}{h_1} \right) L^2 \left(\frac{s_i - s}{h_2} \right) \right] \\
 &= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \sigma^2(s_i) \mathbb{E} \left[\mathbb{E} [Z_1(s_i) Z_1(s_i)' | X_1(s_i)] K^2 \left(\frac{X_t(s_i) - x}{h_1} \right) L^2 \left(\frac{s_i - s}{h_2} \right) \right] \\
 &= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \mathbb{E} [Z_1(s_i) Z_1(s_i)' | x] \sigma^2(s_i) \int K^2 \left(\frac{u - x}{h_1} \right) L^2 \left(\frac{s_i - s}{h_2} \right) f_X(u, s_i) du \\
 &= (T_0 h_1)^{-1} (N h_2^2)^{-2} \sum_{i=1}^N \Omega(x, s_i) \sigma^2(s_i) f_X(x, s_i) L^2 \left(\frac{s_i - s}{h_2} \right) \int K^2(u) du (1 + o(1)) \\
 &= (T_0 h_1 N h_2^2)^{-1} \sigma^2(s) \Omega(x, s) f_X(x, s) \int K^2(u) du \int L^2(u) du (1 + o(1)).
 \end{aligned} \tag{A.15}$$

and, under Assumption (A1)(iii) and Lemma A.2,

$$\begin{aligned}
& T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq i}^N \mathbb{E} (Q_{l1,2} Q'_{k1,2}) \\
&= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq i}^N \mathbb{E} \left[Z_1(s_l) Z_1(s_k)' \varepsilon_1(s_l) \varepsilon_1(s_k) K\left(\frac{X_1(s_l) - x}{h_1}\right) L\left(\frac{s_l - s}{h_2}\right) \right. \\
&\quad \left. \times K\left(\frac{X_1(s_k) - x}{h_1}\right) L\left(\frac{s_k - s}{h_2}\right) \right] \\
&= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq i}^N \mathbb{E} \left\{ \mathbb{E} \left[Z_1(s_l) Z_1(s_k)' \varepsilon_1(s_l) \varepsilon_1(s_k) | X_1(s_l), X_1(s_k) \right] \right. \\
&\quad \left. \times K\left(\frac{X_1(s_l) - x}{h_1}\right) K\left(\frac{X_1(s_k) - x}{h_1}\right) \right\} L\left(\frac{s_l - s}{h_2}\right) L\left(\frac{s_k - s}{h_2}\right) \\
&= T_0^{-1}(N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq i}^N \Gamma(s_l, s_k) \mathbb{E} \left[Z_1(s_l) Z_1(s_k)' | X_1(s_l) = x, X_1(s_k) = x \right] \\
&\quad \times q(x, x; s_l, s_k) L\left(\frac{s_l - s}{h_2}\right) L\left(\frac{s_k - s}{h_2}\right) (1 + o(1)) \\
&= T_0^{-1} f_S^2(s) \Gamma(s, s) \mathbb{E} \left[Z_1(s) Z_1(s)' | x, x \right] q(x, x; s) (1 + o(1)) \\
&= T_0^{-1} f_S^2(s) \Gamma(s, s) \Omega^*(x, s, s) q(x, x; s) (1 + o(1)) \tag{A.16}
\end{aligned}$$

Then, from above results, we can obtain the

$$\begin{aligned}
& T_0^{-1}(h_1 N h_2^2)^{-2} \text{Var}(\Delta_1) \\
&= (T_0 N h_1 h_2^2)^{-1} \sigma^2(s) \Omega(x, s) f_X(x, s) f_S(s) \int K^2(u) du \int L^2(u) du (1 + o(1)) \\
&\quad + T_0^{-1} f_S^2(s) \Gamma(s, s) \Omega^*(x, s, s) q(x, x; s) (1 + o(1)) \\
&= (T_0 h_1 N h_2^2)^{-1} \Sigma_1 (1 + o(1)) + T_0^{-1} \Sigma_2 (1 + o(1)) \\
&= \Sigma (1 + o(1)) \tag{A.17}
\end{aligned}$$

If $N h_1 h_2^2 \rightarrow 0$, then we have $\Sigma = (T_0 h_1 N h_2^2)^{-1} \Sigma_1$. If $N h_1 h_2^2 \rightarrow O(1)$, then $\Sigma = (T_0 h_1 N h_2^2)^{-1} \Sigma_1 + T_0^{-1} \Sigma_2$. If $N h_1 h_2^2 \rightarrow \infty$, then $\Sigma = T_0^{-1} \Sigma_2$.

Then, we consider the first part in (A.13). By Assumption (A1), (A5), and Lemma 4.2, we have

$$\begin{aligned}
& T_0^{-1}(h_1 N h_2^2)^{-2} \mathbb{E} (\Delta_{1,1} \Delta'_{1,1}) \\
&= T_0^{-1}(h_1 N h_2^2)^{-2} \mathbb{E} \left(\left(\sum_{i=1}^N \zeta_{i1} \right) \left(\sum_{i=1}^N \zeta_{i1} \right)' \right) \\
&= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \mathbb{E} (\zeta_{i1} \zeta'_{i1}) + T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq l}^N \mathbb{E} (\zeta_{l1} \zeta'_{k1}) \\
&= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \mathbb{E} ((Q_{i1,1} - \mathbb{E} Q_{i1,1})(Q_{i1,1} - \mathbb{E} Q_{i1,1})') \\
&\quad + T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq l}^N \mathbb{E} ((Q_{l1,1} - \mathbb{E} Q_{l1,1})(Q_{k1,1} - \mathbb{E} Q_{k1,1})') \\
&= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N \mathbb{E} (Q_{i1,1} Q'_{i1,1} + \mathbb{E} Q_{i1,1} \mathbb{E} Q'_{i1,1} - Q_{i1,1} \mathbb{E} Q'_{i1,1} - \mathbb{E} Q_{i1,1} Q'_{i1,1}) \\
&\quad + T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq l}^N \mathbb{E} ((Q_{l1,1} - \mathbb{E} Q_{l1,1})(Q_{k1,1} - \mathbb{E} Q_{k1,1})') \\
&= T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^N (\mathbb{E} Q_{i1,1} Q'_{i1,1} - \mathbb{E} Q_{i1,1} \mathbb{E} Q'_{i1,1}) \\
&\quad + T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{l=1}^N \sum_{k=1, \neq l}^N (\mathbb{E} (Q_{l1,1} Q'_{k1,1}) - \mathbb{E} Q_{l1,1} \mathbb{E} Q'_{k1,1}) \\
&= (T_0 N h_1 h_2^2)^{-1} O(h_1^4 + 2h_1^2 h_2^2 + h_2^4) + T_0^{-1} O(h_1^4 + 2h_1^2 h_2^2 + h_2^4) = o(\Sigma) \quad (\text{A.18})
\end{aligned}$$

Next, we will consider the second part $T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^{T-r-1} (1 - \frac{i}{T-r}) \mathbb{E} (\Delta_1 \Delta'_{1+i})$ in (A.12)

$$\begin{aligned}
& \mathbb{E} (\Delta_1 \Delta'_{1+i}) \\
&= \mathbb{E} ((\Delta_{1,1} + \Delta_{1,2})(\Delta_{1+i,1} + \Delta_{1+i,2})') \\
&= \mathbb{E} (\Delta_{1,1} \Delta'_{1+i,1}) + \mathbb{E} (\Delta_{1,1} \Delta'_{1+i,2}) + \mathbb{E} (\Delta_{1,2} \Delta'_{1+i,1}) + \mathbb{E} (\Delta_{1,2} \Delta'_{1+i,2}). \quad (\text{A.19})
\end{aligned}$$

For $\mathbb{E} (\Delta_{1,1} \Delta'_{1+i,1})$, based on Assumption (A1) and (A5), let d_T be a sequence of positive integers. Define

$$J_1 = T_0^{-1}(h_1 N h_2^2)^{-2} \sum_{i=1}^{d_T-1} |\mathbb{E} (\Delta_{1,1} \Delta'_{1+i,1})|,$$

and

$$J_2 = T_0^{-1} (h_1 N h_2^2)^{-2} \sum_{i=d_T}^{T-r-1} |\mathbb{E} (\Delta_{1,1} \Delta'_{1+i,1})|.$$

For $1 \leq i \leq d_T - 1$, by Assumption (A1), (A3), (A5) and Lemma A.2, we have

$$\begin{aligned} & |\mathbb{E} (\Delta_{1,1} \Delta'_{1+i,1})| \\ & \leq \sum_{l=1}^N \sum_{k=1}^N |\mathbb{E} (\zeta_{l,1} \zeta'_{k,1+i})| \\ & = \sum_{l=1}^N \sum_{k=1}^N |\mathbb{E} ((Q_{l,1,1} - \mathbb{E} Q_{l,1,1}) (Q_{k,1+i,1} - \mathbb{E} Q_{k,1+i,1})')| \\ & = \sum_{l=1}^N \sum_{k=1}^N |\mathbb{E} Q_{l,1,1} Q'_{l,1+i,1} - \mathbb{E} Q_{l,1,1} \mathbb{E} Q'_{k,1+i,1}| \\ & = (N h_1 h_2^2)^2 O(h_1^4 + 2 h_1^2 h_2^2 + h_2^4) \\ & = (N h_1 h_2^2)^2 O(\max(h_1^4, h_2^4)). \end{aligned} \tag{A.20}$$

Then, we can obtain

$$J_1 \leq T_0^{-1} d_T O(\max(h_1^4, h_2^4)).$$

We next consider the upper bound of J_2 . By using Lemma A.1, we obtain, for $i \geq 1$

$$\begin{aligned} & |\mathbb{E} (\Delta_{1,1} \Delta'_{1+i,1})| \\ & \leq \sum_{l=1}^N \sum_{k=1}^N |\mathbb{E} (\zeta_{l,1} \zeta'_{k,1+i})| \\ & = \sum_{l=1}^N \sum_{k=1}^N |\mathbb{E} ((Q_{l,1,1} - \mathbb{E} Q_{l,1,1}) (Q_{l,1+i,1} - \mathbb{E} Q_{k,1+i,1})')| \\ & = \sum_{l=1}^N \sum_{k=1}^N |\text{cov} (Q_{l,1,1}, Q_{k,1+i,1})| \\ & \leq C \sum_{l=1}^N \sum_{k=1}^N [\alpha(i)]^{1-2/\delta} \left[\mathbb{E} |Q_{l,1,1}|^\delta \right]^{1/\delta} \left[\mathbb{E} |Q_{k,1+i,1}|^\delta \right]^{1/\delta}. \end{aligned} \tag{A.21}$$

By Assumption (A1), (A3), (A5), we can obtain

$$\begin{aligned} E \left[|Q_{l,1,1}|^\delta \right] &\leq CE \left[|Z_1(s_l)|^\delta |m(X_1(s_l), Z_1(s_l))|^\delta K^\delta \left(\frac{X_1(s_l) - x}{h_1} \right) L^\delta \left(\frac{s_l - s}{h_2} \right) \right] \\ &\leq Ch_1 h_2^2 (h_1^{2\delta} + h_1^\delta h_2^\delta + h_2^{2\delta}). \end{aligned} \quad (\text{A.22})$$

Then, a combination of above two equations leads to

$$\begin{aligned} J_2 &\leq CT_0^{-1} (h_1 h_2^2)^{2/\delta-2} (h_1^{2\delta} + h_1^\delta h_2^\delta + h_2^{2\delta})^{2/\delta} \sum_{i=d_T}^{\infty} [\alpha(i)]^{1-2/\delta} \\ &\leq CT_0^{-1} (h_1 h_2^2)^{2/\delta-2} \max(h_1^4, h_2^4) d_T^{-c} \sum_{i=d_T}^{\infty} i^c [\alpha(i)]^{1-2/\delta} \\ &\leq CT_0^{-1} \max(h_1^3 h_2^{-2}, h_2^2 h_1^{-1}) (h_1 h_2^2)^{2/\delta-1} d_T^{-c} \sum_{i=d_T}^{\infty} i^c [\alpha(i)]^{1-2/\delta}. \end{aligned} \quad (\text{A.23})$$

Then, with Assumption (A1) and (A6), if $Nh_1 h_2^2 \rightarrow 0$, it follows that

$$\begin{aligned} J_1 + J_2 &\leq C \left(T_0^{-1} d_T h_1 h_2^2 \max(h_1^3 h_2^{-2}, h_2^2 h_1^{-1}) + T_0^{-1} \max(h_1^3 h_2^{-2}, h_2^2 h_1^{-1}) (h_1 h_2^2)^{2/\delta-1} d_T^{-c} \sum_{i=d_T}^{\infty} i^c [\alpha(i)]^{1-2/\delta} \right) \\ &= o(T_0^{-1}). \end{aligned} \quad (\text{A.24})$$

We can choose d_T such that $(h_1 h_2^2)^{1-2/\delta} d_T^c \rightarrow C$, $c > 1 - 2/\delta$ and $\delta > 2$, satisfying $d_T h_1 h_2^2 \rightarrow 0$.

If $Nh_1 h_2^2 \rightarrow 0$, by choosing d_T such that $d_T h_1 h_2^2 \rightarrow 0$, it follows that

$$J_1 + J_2 \leq o(T_0^{-1}) \leq o\left((T_0 N h_1 h_2^2)^{-1}\right). \quad (\text{A.25})$$

Finally, by Assumption (A1), (A3) and (A5), $E(\Delta_{1,1} \Delta'_{1+i,2}) = E(\Delta_{1,2} \Delta'_{1+i,1}) = 0$

For $E(\Delta_{1,2}\Delta'_{1+i,2})$, by Assumption (A1), (A3) and (A5), we have

$$\begin{aligned}
& E(\Delta_{1,2}\Delta'_{1+i,2}) \\
& \leq \sum_{l=1}^N \sum_{k=1}^N |E(Q_{l,1,2}Q'_{k,1+i,2})| \\
& = \sum_{l=1}^N \sum_{k=1}^N C |E \left[Z_1(s_l)Z_{1+i}(s_k)' \varepsilon_1(s_l)\varepsilon_{1+i}(s_k) \right. \\
& \quad \times K\left(\frac{X_1(s_l) - x}{h_1}\right)L\left(\frac{s_l - s}{h_2}\right)K\left(\frac{X_{1+i}(s_k) - x}{h_1}\right)L\left(\frac{s_k - s}{h_2}\right) \Big]| \\
& = 0
\end{aligned} \tag{A.26}$$

The proof of the lemma is completed.

Proof of Theorem 4.1

To prove this Theorem, we use the small-block and large-block technique, partitioning $\{1, \dots, T_0\}$ into $r(p+q)$ subsets for some integer $r = r_T \rightarrow 0$ with large block of size $p = p_T \in \mathbb{Z}$ and small block of size $q = q_T \in \mathbb{Z}$.

First, we consider under the condition $Nh_1h_2^2 \rightarrow 0$. There is a sequence of positive constants $\gamma_T \rightarrow \infty$ such that

$$\gamma_T q_T = o(\sqrt{T_0 h_1})$$

and

$$\gamma_T (T_0/h_1)^{1/2} \alpha(q_T) \rightarrow 0$$

Without loss of generality, let the large-block size p_T by $p = p_T = \lfloor (T_0 h_1)^{1/2} / \gamma_T \rfloor$ such that

$$q_T/p_T \rightarrow 0, \quad T/p_T \rightarrow \infty, \quad T_0 p_T^{-1} \alpha(q_T) \rightarrow 0. \tag{A.27}$$

We now use the Cramér-Wold device to derive the asymptotic normality. For any unit vector $\mathbf{c} \in \mathbb{R}^{p+q+1}$, let

$$S_{TN} = (T_0)^{1/2} \mathbf{c}' (\mathcal{W}_{TN,1} - E\mathcal{W}_{TN,1}) = (T_0)^{-1/2} \sum_{t=r+1}^T \Phi_t, \tag{A.28}$$

where $\Phi_t = (Nh_1h_2^2)^{-1} \mathbf{c}'\Delta_t$. Then based on Lemma 4.3, we have

$$\text{Var}(\Phi_1) = \Sigma(1 + o(1)),$$

where $\Sigma = (h_1Nh_2^2)^{-1} \Sigma_1 + \Sigma_2$ when $Nh_1h_2^2 \rightarrow O(1)$, and $\Sigma = \Sigma_2$ if $Nh_1h_2^2 \rightarrow \infty$.

$$\sum_{r+1 \leq i < j \leq T} |\text{cov}(\Phi_i, \Phi_j)| = o(1).$$

Now, let us decompose S_{TN} into smaller pieces involving “large” and “small” blocks. More specifically, for $0 \leq j \leq r-1$, consider

$$U(1, j) := \sum_{t=j(p+q)+1}^{j(p+q)+p} \Phi_t, \quad (\text{A.29})$$

$$U(2, j) := \sum_{t=j(p+q)+p+1}^{(j+1)(p+q)} \Phi_t. \quad (\text{A.30})$$

Then

$$\begin{aligned} A_{TN} &= \frac{1}{\sqrt{T_0}} \left\{ \sum_{j=0}^{r-1} U(1, j) + \sum_{j=0}^{r-1} U(2, j) \right\} \\ &\equiv \frac{1}{\sqrt{T_0}} \{ \Upsilon_{T,1} + \Upsilon_{T,2} \} \end{aligned} \quad (\text{A.31})$$

If it is not the case that $T_0 = r(p+q)$ for some integer r , then an additional term $\Upsilon_{T,3}$, say, containing all the Φ_t 's that are not included in the large or small blocks, can be considered. This term will not change the proof much. Then we show that, as $T_0, N \rightarrow \infty$

$$\frac{1}{T_0} E [\Upsilon_{T,2}]^2 \rightarrow 0, \quad (\text{A.32})$$

$$\left| E [\exp (iu \Upsilon_{T,1})] - \prod_{j=0}^{r_T-1} E [\exp (iu U(1, j))] \right| \rightarrow 0 \quad (\text{A.33})$$

$$\frac{1}{T_0} \sum_{j=0}^{r_T-1} E (U^2(1, j)) \rightarrow \Sigma \quad (\text{A.34})$$

$$\frac{1}{T_0} \sum_{j=0}^{r_T-1} E \left[U(1, j) I \left\{ |U(1, j)| \geq \varepsilon \Sigma^{1/2} \sqrt{T_0} \right\} \right] \rightarrow 0 \quad (\text{A.35})$$

for every $\varepsilon > 0$. (A.32) implies that $\Upsilon_{T,2}$ is asymptotically negligible in probability. (A.33) indicates that $U(1, j)$ are asymptotically mutually independent. The asymptotic normality of $\Upsilon_{T,1}$ follows from (A.34) and the Lindeberg–Feller condition (A.35).

First, we proof (A.32).

$$\begin{aligned} E [\Upsilon_{T,2}]^2 &= \sum_{j=0}^{r_T-1} \text{var} (U(2, j)) + 2 \sum_{0 \leq i < j \leq r_T-1} \text{cov} (U(2, i), U(2, j)) \\ &\equiv I_1 + I_2. \end{aligned} \quad (\text{A.36})$$

Based on the stationarity and Lemma 4.3, we have

$$\begin{aligned} I_1 &= r_T \text{var} (U(2, 1)) = r_T \text{var} \left(\sum_{j=1}^{q_T} \Phi_j \right) \\ &= r_T q_T [\Sigma + o(1)]. \end{aligned}$$

Next consider the second term I_2 . Let $p_i^* = i(p_T + q_T)$ and $p_j^* = j(p_T + q_T)$, then $p_j^* - p_i^* \geq p_T$ for all $j > i$, we thus have

$$\begin{aligned} |I_2| &\leq 2 \sum_{0 \leq i < j \leq r_T-1} \sum_{j_1=1}^{q_T} \sum_{j_2=1}^{q_T} \left| \text{cov} \left(\Phi_{p_i^*+p_T+j_1}, \Phi_{p_j^*+p_T+j_2} \right) \right| \\ &\leq 2 \sum_{j_1=1}^{T_0-p_T} \sum_{j_2=j_1+p_T}^{T_0} |\text{cov} (\Phi_{j_1}, \Phi_{j_2})|. \end{aligned}$$

By stationarity and Lemma 4.3, one obtains

$$|I_2| \leq 2T_0 \sum_{j=p_T+1}^{T_0} |\text{cov}(\Phi_1, \Phi_j)| = o(T_0).$$

Hence, we have

$$\frac{1}{T_0} E [\Upsilon_{T,2}]^2 = O(r_T q_T T_0^{-1}) + o(1) = o(1).$$

For (A.34), by stationarity, (A.27), and Lemma 4.3, it is easily seen that

$$\frac{1}{T_0} \sum_{j=0}^{r_T-1} E(U^2(1, j)) = \frac{r_T}{T_0} E(U^2(1, 1)) = \frac{r_T p_T}{T_0} \cdot \frac{1}{p_T} \text{var} \left(\sum_{j=1}^{p_T} \Phi_j \right) \rightarrow \Sigma.$$

For (A.33), using (A.27) and the Lemma 1.1 in Volkonskii and Rozanov (1959), we have

$$\left| E[\exp(iu\Upsilon_{T,1})] - \prod_{j=0}^{r_T-1} E[\exp(iuU(1, j))] \right| \leq 16(T_0/p_T) \alpha(q_T) \rightarrow 0 \quad (\text{A.37})$$

To establish (A.35), we use theorem 4.1 in Shao and Yu (1996) with $\delta^* > \delta$ and Assumptions (A6) and obtain

$$E \left[U(1, j) I \left\{ |U(1, j)| \geq \varepsilon \Sigma^{1/2} \sqrt{T_0} \right\} \right] \quad (\text{A.38})$$

$$\leq CT_0^{1-\delta/2} E(|U(1, j)|^\delta) \leq CT_0^{1-\delta/2} (p_T N)^{\delta/2} \left\{ E \left(\left| (Nh_1 h_2^2)^{-1} \mathbf{c}' Q_{11} \right|^{\delta^*} \right) \right\}^{\delta/\delta^*} \quad (\text{A.39})$$

Following (A.22),

$$\begin{aligned} & E \left(\left| (Nh_1 h_2^2)^{-1} \mathbf{c}' Q_{11} \right|^{\delta^*} \right) \\ & \leq C (Nh_1 h_2^2)^{-\delta^*} E(|\mathbf{c}' Q_{11}|^{\delta^*}) \\ & \leq C (Nh_1 h_2^2)^{-\delta^*} h_1 h_2^2 \\ & \leq CN^{-\delta^*} (h_1 h_2^2)^{1-\delta^*} \end{aligned} \quad (\text{A.40})$$

Thus, by (A.27) and Assumption (A6), we obtain

$$\begin{aligned}
& \frac{1}{T_0} \sum_{j=0}^{r_T-1} E \left[U(1, j) I \left\{ |U(1, j)| \geq \varepsilon \Sigma^{1/2} \sqrt{T_0} \right\} \right] \\
& \leq C T_0^{1-\delta/2} p_T^{\delta/2-1} N^{\delta/2} N^{-\delta} (h_1 h_2^2)^{(1-\delta^*)\delta/(\delta^*)} \\
& \leq C T_0^{1-\delta/2} N^{-\delta/2} \left[(T_0 h_1)^{1/2} / \gamma_T \right]^{\delta/2-1} (h_1 h_2^2)^{(1-\delta^*)\delta/(\delta^*)} \\
& \leq C \gamma_T^{1-\delta/2} T_0^{1/2-\delta/4} h_1^{\delta/\delta^*-1/2-\delta/4} h_2^{2\delta/\delta^*-\delta} (N h_1 h_2^2)^{-\delta/2} \rightarrow 0,
\end{aligned} \tag{A.41}$$

because $\gamma_T \rightarrow \infty$. This completes the proof of the theorem.

Then, we consider under the condition $N h_1 h_2^2 \rightarrow 0$. There is a sequence of positive constants $\gamma_T \rightarrow \infty$ such that

$$\gamma_T q_T = o(\sqrt{T_0 h_1})$$

and

$$\gamma_T (T_0/h_1)^{1/2} \alpha(q_T) \rightarrow 0$$

Without loss of generality, let the large-block size p_T by $p = p_T = \left\lfloor (T_0 h_1)^{1/2} / \gamma_T \right\rfloor$ such that

$$q_T/p_T \rightarrow 0, \quad T_0/p_T \rightarrow \infty, \quad T_0 p_T^{-1} \alpha(q_T) \rightarrow 0. \tag{A.42}$$

We now use the Cramér-Wold device to derive the asymptotic normality. For any unit vector $\mathbf{c} \in \mathfrak{R}^{p+q+1}$, let

$$S_{TN} = (T_0 N h_1 h_2^2)^{1/2} \mathbf{c}' (\mathcal{W}_{TN,1} - E \mathcal{W}_{TN,1}) = (T_0)^{-1/2} \sum_{t=r+1}^T \Phi_t, \tag{A.43}$$

where $\Phi_t = (N h_1 h_2^2)^{-1/2} \mathbf{c}' \Delta_t$. Then based on Lemma 4.3, we have

$$\text{Var}(\Phi_1) = \Sigma(1 + o(1)),$$

with $\Sigma = \Sigma_1$, and

$$\sum_{r+1 \leq i < j \leq T} |\text{cov}(\Phi_i, \Phi_j)| = o(1).$$

Now, let us decompose S_{TN} into smaller pieces involving “large” and “small” blocks. More specifically, for $0 \leq j \leq r-1$, consider

$$U(1, j) := \sum_{t=j(p+q)+1}^{j(p+q)+p} \Phi_t, \quad (\text{A.44})$$

$$U(2, j) := \sum_{t=j(p+q)+p+1}^{(j+1)(p+q)} \Phi_t. \quad (\text{A.45})$$

Then

$$\begin{aligned} A_{TN} &= \frac{1}{\sqrt{T_0}} \left\{ \sum_{j=0}^{r-1} U(1, j) + \sum_{j=0}^{r-1} U(2, j) \right\} \\ &\equiv \frac{1}{\sqrt{T_0}} \{ \Upsilon_{T,1} + \Upsilon_{T,2} \} \end{aligned} \quad (\text{A.46})$$

Then we show that, as $T_0, N \rightarrow \infty$

$$\frac{1}{T_0} E [\Upsilon_{T,2}]^2 \rightarrow 0, \quad (\text{A.47})$$

$$\left| E [\exp(iu\Upsilon_{T,1})] - \prod_{j=0}^{r_T-1} E [\exp(iuU(1, j))] \right| \rightarrow 0 \quad (\text{A.48})$$

$$\frac{1}{T_0} \sum_{j=0}^{r_T-1} E (U^2(1, j)) \rightarrow \Sigma \quad (\text{A.49})$$

$$\frac{1}{T_0} \sum_{j=0}^{r_T-1} E \left[U(1, j) I \left\{ |U(1, j)| \geq \varepsilon \Sigma^{1/2} \sqrt{T_0} \right\} \right] \rightarrow 0 \quad (\text{A.50})$$

for every $\varepsilon > 0$.

The proofs for (A.47)-(A.50) are similar to the condition $Nh_1h_2^2 \rightarrow 0$. Here, we just show the proof for (A.50).

To establish (A.50), we use theorem 4.1 in Shao and Yu (1996) with $\delta^* > \delta$ and Assumptions (A6) and obtain

$$E \left[U(1, j) I \left\{ |U(1, j)| \geq \varepsilon \Sigma^{1/2} \sqrt{T_0} \right\} \right] \quad (\text{A.51})$$

$$\leq CT_0^{1-\delta/2} E \left(|U(1, j)|^\delta \right) \leq CT_0^{1-\delta/2} (p_T N)^{\delta/2} \left\{ E \left(\left| (Nh_1 h_2^2)^{-1/2} \mathbf{c}' Q_{11} \right|^{\delta^*} \right) \right\}^{\delta/\delta^*} \quad (\text{A.52})$$

Following (A.22),

$$\begin{aligned} & E \left(\left| (Nh_1 h_2^2)^{-1/2} \mathbf{c}' Q_{11} \right|^{\delta^*} \right) \\ & \leq C (Nh_1 h_2^2)^{-\delta^*/2} E \left(|\mathbf{c}' Q_{11}|^{\delta^*} \right) \\ & \leq C (Nh_1 h_2^2)^{-\delta^*/2} h_1 h_2^2 \\ & \leq CN^{-\delta^*/2} (h_1 h_2^2)^{1-\delta^*/2} \end{aligned} \quad (\text{A.53})$$

Thus, by (A.27) and Assumption (A6), we obtain

$$\begin{aligned} & \frac{1}{T_0} \sum_{j=0}^{r_T-1} E \left[U(1, j) I \left\{ |U(1, j)| \geq \varepsilon \Sigma^{1/2} \sqrt{T_0} \right\} \right] \\ & \leq CT_0^{1-\delta/2} p_T^{\delta/2-1} N^{\delta/2} N^{-\delta/2} (h_1 h_2^2)^{(2-\delta^*)\delta/(2\delta^*)} \\ & \leq CT_0^{1-\delta/2} \left[(T_0 h_1)^{1/2} / \gamma_T \right]^{\delta/2-1} (h_1 h_2^2)^{(2-\delta^*)\delta/(2\delta^*)} \\ & \leq C \gamma_T^{1-\delta/2} T_0^{1/2-\delta/4} h_1^{\delta/\delta^*-1/2-\delta/4} h_2^{2\delta/\delta^*-\delta} \rightarrow 0, \end{aligned} \quad (\text{A.54})$$

because $\gamma_T \rightarrow \infty$.

Proof of Theorem 4.2.

For kernel function $\tilde{L}_{h_2, i}^*(s)$ in Section 4.3, it is straightforward to verify that, under Assumption (A4),

$$\sum_{i=1}^N \left(\frac{s_i - s}{h_2} \right) \tilde{L}_{h_2, i}^*(s) = 0 \quad (\text{A.55})$$

and

$$\sum_{i=1}^N \left(\frac{s_i - s}{h_2} \right) \left(\frac{s_i - s}{h_2} \right)' \tilde{L}_{h_2, i}^*(s) \longrightarrow \int z z' \tilde{L}(z) dz. \quad (\text{A.56})$$

First, considering the time series estimation, for $s_i \in S$, based on the Theorem 1 in [Cai et al. \(2000\)](#), for $s_i \in S$

$$\begin{aligned} & \hat{\beta}^*(x, s_i) - \beta(x, s_i) \\ &= \frac{\Omega(x, s_i)^{-1}}{f_X(x, s_i)T_0h_1} \sum_{t=r+1}^T Z_t(s_i)\varepsilon_t(s_i)K\left(\frac{X_t(s_i) - x}{h_1}\right) \\ &+ \frac{h_1^2}{2} \frac{\partial^2 \beta(x, s_i)}{\partial x^2} \int u^2 K(u) du + o_p(h_1^2). \end{aligned} \quad (\text{A.57})$$

Then, it follows from (A.57) that, as $T_0 \rightarrow \infty$,

$$\begin{aligned} & \tilde{\beta}(x, s) - \beta(x, s) \\ &= \sum_{i=1}^N \left\{ \hat{\beta}^*(x, s_i) - \beta(x, s_i) \right\} \tilde{L}_{h_2, i}^*(s) + \sum_{i=1}^N \left\{ \beta(x, s_i) - \beta(x, s) \right\} \tilde{L}_{h_2, i}^*(s) \\ &= \sum_{i=1}^N \frac{\Omega^*(x, s_i)^{-1}}{f_X(x, s_i)T_0h_1} \sum_{t=r+1}^T Z_t(s_i)\varepsilon_t(s_i)K\left(\frac{X_t(s_i) - x}{h_1}\right) \tilde{L}_{h_2, i}^*(s) \{1 + o(1)\} \\ &+ \sum_{i=1}^N \frac{h_1^2}{2} \frac{\partial^2 \beta(x, s_i)}{\partial x^2} \int u^2 K(u) du \tilde{L}_{h_2, i}^*(s) \{1 + o(1)\} \\ &+ \sum_{i=1}^N \left\{ \beta(x, s_i) - \beta(x, s) \right\} \tilde{L}_{h_2, i}^*(s) \{1 + o(1)\} \\ &= I_1\{1 + o(1)\} + I_2\{1 + o(1)\} + I_3\{1 + o(1)\}, \end{aligned} \quad (\text{A.58})$$

where

$$\begin{aligned} I_2 &= \sum_{i=1}^N \frac{h_1^2}{2} \frac{\partial^2 \beta(x, s_i)}{\partial x^2} \int u^2 K(u) du \tilde{L}_{h_2, i}^*(s) \\ &= \frac{h_1^2}{2} \frac{\partial^2 \beta(x, s)}{\partial x^2} \int \int u^2 K(u) du \frac{1}{h_2^2 f_S(s)} L\left(\frac{v - s}{h_2}\right) f_S(v) dv \{1 + o(1)\} \\ &= \frac{h_1^2}{2} \frac{\partial^2 \beta(x, s)}{\partial x^2} \int u^2 K(u) du \{1 + o(1)\}. \end{aligned} \quad (\text{A.59})$$

Also, from (A.55), (A.56), and by Taylor's expansion, we have

$$\begin{aligned}
 I_3 &= \sum_{i=1}^N \{\beta(x, s_i) - \beta(x, s)\} \tilde{L}_{h_2, i}^*(s) \\
 &= \sum_{i=1}^N \frac{\partial \beta(x, s)}{\partial s'} (s_i - s) \tilde{L}_{h_2, i}^*(s) + (1/2) \left(\sum_{i=1}^N (s_i - s)' \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} (s_i - s) \tilde{L}_{h_2, i}^*(s), \dots, \right. \\
 &\quad \left. \sum_{i=1}^N (s_i - s)' \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} (s_i - s) \tilde{L}_{h_2, i}^*(s) \right)' \\
 &= \frac{h_2^2}{2} \left(\text{tr} \left\{ \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\}, \dots, \text{tr} \left\{ \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right)' \{1 + o(1)\}.
 \end{aligned} \tag{A.60}$$

Further, $E[I_1] = 0$ and

$$\begin{aligned}
 E[I_1^2] &= (T_0 h_1)^{-2} \sum_{i=1}^N \frac{\Omega(x, s_i)^{-1} \Omega(x, s_i)^{-1}}{(f_X(x, s_i))^2} \sum_{t=r+1}^T E[Z_t(s_i) Z_t(s_i)' \varepsilon_t^2(s_i) K^2 \left(\frac{X_t(s_i) - x}{h_1} \right) \\
 &\quad \times \left\{ \tilde{L}_{h_2, i}^*(s) \right\}^2] + (T_0 h_1)^{-2} \sum_{k=1}^N \sum_{i=1, i \neq k}^N \frac{\Omega(x, s_i)^{-1} \Omega(x, s_k)^{-1}}{f_X(x, s_i) f_X(x, s_k)} \\
 &\quad \times \sum_{t=r+1}^T E[Z_t(s_i) \varepsilon_t(s_i) \varepsilon_t(s_k) Z_t(s_k)' K \left(\frac{X_t(s_i) - x}{h_1} \right) K \left(\frac{X_t(s_k) - x}{h_1} \right) \tilde{L}_{h_2, i}^*(s) \tilde{L}_{h_2, k}^*(s)] \\
 &= I_{11} + I_{12}.
 \end{aligned} \tag{A.61}$$

In particular,

$$\begin{aligned}
 I_{11} &= (T_0 h_1)^{-2} \sum_{i=1}^N \frac{\Omega(x, s_i)^{-1} \Omega(x, s_i)^{-1}}{(f_X(x, s_i))^2} \sum_{t=r+1}^T \sigma^2(s_i) E[Z_t(s_i) Z_t(s_i)' | X_t(s_i) = x] \\
 &\quad \times E K^2 \left(\frac{X_t(s_i) - x}{h_1} \right) \left\{ \tilde{L}_{h_2, i}^*(s) \right\}^2 \\
 &= (T_0 h_1)^{-1} \sum_{i=1}^N \sigma^2(s_i) \frac{\Omega(x, s_i)^{-1}}{f_X(x, s_i)} \int K^2(u) du \\
 &\quad \times \left[\left\{ N h_2^2 f_S(s) \right\}^{-1} L \left(\frac{s_i - s}{h_2} \right) \right]^2 \{1 + o(1)\} \\
 &= \{T_0 h_1 N h_2^2\}^{-1} \sigma^2(s) \frac{\Omega(x, s)^{-1}}{f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz \{1 + o(1)\}, \tag{A.62}
 \end{aligned}$$

and, under Assumptions and Lemma A.2,

$$\begin{aligned}
 I_{12} &= (T_0 h_1)^{-2} \sum_{k=1}^N \sum_{i=1, \neq k}^N \frac{\Omega(x, s_i)^{-1} \Omega(x, s_k)^{-1}}{f_X(x, s_i) f_X(x, s_k)} \sum_{t=r+1}^T E[Z_t(s_i) \varepsilon_t(s_i) \varepsilon_t(s_k) Z_t(s_k)'] \\
 &\quad \times K\left(\frac{X_t(s_i) - x}{h_1}\right) K\left(\frac{X_t(s_k) - x}{h_1}\right) \tilde{L}_{h_2, i}^*(s) \tilde{L}_{h_2, k}^*(s) \\
 &= T_0^{-1} \sum_{k=1}^N \sum_{i=1, \neq k}^N \frac{\Omega(x, s_i)^{-1} \Omega(x, s_k)^{-1}}{f_X(x, s_i) f_X(x, s_k)} \\
 &\quad \times E[Z_t(s_i) Z_t(s_k)' \varepsilon_t(s_i) \varepsilon_t(s_k) | X_t(s_i) = x, X_t(s_k) = x] q(x, x; s_i, s_k) \tilde{L}_{h_2, i}^*(s) \tilde{L}_{h_2, k}^*(s) \\
 &= T_0^{-1} \Gamma(s, s) \frac{\Omega(x, s)^{-1} \Omega^*(x, s, s)^{-1} \Omega(x, s)^{-1}}{f_X^2(x, s)} q(x, x; s) \{1 + o(1)\}. \tag{A.63}
 \end{aligned}$$

Thus, from (A.62) and (A.63), we have

$$\begin{aligned}
 &\{T_0 h_1\}^{-1} \left[\sigma^2(s) \frac{\Omega(x, s)^{-1}}{N h_2^2 f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz \right. \\
 &\quad \left. + h_1 \Gamma(s, s) \frac{\Omega(x, s)^{-1} \Omega^*(x, s, s)^{-1} \Omega(x, s)^{-1}}{f_X^2(x, s)} q(x, x; s) \right] \{1 + o(1)\}. \tag{A.64}
 \end{aligned}$$

Together with (A.59) and (A.60), the proof for asymptotic variance and bias is completed. Finally, the asymptotic normality follows from (A.58) by letting $T_0 \rightarrow \infty$ first and then $N \rightarrow \infty$. The proof is completed.

Chapter 5

Semiparametric Dynamic Varying Coefficients Spatio-temporal Quantile Regression

5.1 Introduction

In Chapter 4, we proposed the dynamic functional-coefficient auto-regressive spatio-temporal model, which permits flexibility in nonlinear modelling and analysis of spatio-temporal data with irregularly observed locations and location-wide non-stationarity. However, spatio-temporal data often do not comply with the normal distribution and have obvious peak and fat tails in practice. The classical conditional mean regression, like the DyFAST, will hence yield poor and biased estimation results. In time series case, to circumvent this difficulty, quantile regression ([Koenker and Bassett Jr, 1978](#)) has been used in the literature. Compared with the classical conditional mean regression, the quantile regression is more robust for the skewed data (see [Yu et al., 2003](#), for a review). In addition, quantile regression can be used to analyse the tail behaviour of distributions and characterize the entire conditional distribution, with no need of distributional assumptions ([Sherwood and Wang, 2016](#)). Therefore, this chapter proposes a semiparametric dynamic varying coefficients spatio-temporal quantile regression model to analyse the nonlinear relationship in non-normal spatio-temporal data, which can be flexibly applied in many research fields.

This proposed model applies the quantile regression to estimate the varying coefficient structures with the autoregressive smooth coefficients depending both on concerned regime variable and spatial location. Therefore, this model can hence not only well characterise the entire conditional distribution of the dynamic regime-switching nature but also overcome the location-wide non-stationarity existent in spatio-temporal data. This model is a useful extension of DyFAST models and varying coefficient quantile regression models (Honda, 2004; Cai and Xu, 2008). Both theoretical properties and Monte Carlo simulations are investigated.

In empirical application, this model is applied to identify the price transmission of retail gasoline in the EU. As argued by Banfi et al. (2005) and Wlazlowski et al. (2009), the price dis-equilibria of retail gasoline in neighbour countries can cause “fuel travels” and create significantly cross-national spatial price spillovers in EU countries. Moreover, retail gasoline price returns data in some countries may be non-Gaussian and have obvious peak and fat tails (Hung et al., 2008). Correctly estimating the tail behaviour of distributions is of crucial importance in the energy risk management and has policy implications. Therefore, this data set can be used to further illustrate the usefulness of this proposed model.

5.2 Modelling Procedure

Let $Y_t(s)$ and $X_t(s)$ denote two spatio-temporal processes at discrete time points $t = -r, \dots, T$, where $r = \max\{p, q\}$ ($\{p, q\}$ will be defined later), and locations s in a spatial domain $S \subset \mathbb{R}^2$. The relationship of between Y and X is of interest, with Y denoting the response variable and X the covariate vector of dimension d , respectively. Without loss of generality, we consider only the case $d = 1$. Assume that both processes are observed at $T + r$ time points $t = -r, \dots, T$ and at N spatial sampling locations $s_i = (u_i, v_i)' \in S$ for $i = 1, \dots, N$ on a possibly irregular grid. That is, the data comprise $\{(Y_t(s_i), X_t(s_i)): t = -r, \dots, T \text{ and } i = 1, \dots, N\}$.

Following Chapter 4, let define a spatial lag variable

$$Y_t^{SL}(s_i) = \sum_{j=1}^N w_{ij} Y_t(s_j),$$

where $W = (w_{ij})_{N \times N}$ is a pre-specified spatial weight matrix either by experts knowledge or by using location-related information and row-wise standardised, which is satisfying $w_{ij} \geq 0$ and $\sum_{j=1}^N w_{ij} = 1$. And let $Z_t(s_i) = (1, \dots, Y_{t-1}^{sl}(s_i), \dots,$

$Y_{t-p}^{sl}(s_i), Y_{t-1}(s_i), \dots, Y_{t-q}(s_i))'$ denotes the vector of spatio-temporally lagged variables and $\beta(x, s_i) = (g(x, s_i), \dots, \lambda_1(x, s_i), \dots, \lambda_p(x, s_i), \alpha_1(x, s_i), \dots, \alpha_q(s_i))'$ denote the corresponding vector of functional coefficients. By extending the DyFAST in Chapter 4, we therefore propose a class of location-dependent semiparametric varying coefficients spatio-temporal quantile regression models in the form of

$$\begin{aligned} & Q_{Y_t(s_i)}(\tau | X_t(s_i), Y_{t-1}^{sl}(s_i), \dots, Y_{t-p}^{sl}(s_i), Y_{t-1}(s_i), \dots, Y_{t-q}(s_i)) \\ &= g_\tau(X_t(s_i), s_i) + \sum_{j=1}^p \lambda_{\tau,j}(X_t(s_i), s_i) Y_{t-j}^{sl}(s_i) + \sum_{l=1}^q \alpha_{\tau,l}(X_t(s_i), s_i) Y_{t-l}(s_i), \quad (5.2.1) \\ & Y_t^{sl}(s_i) = \sum_{k=1}^N W_{ik} Y_t(s_k), \quad i = 1, \dots, N, t = 1, \dots, T. \end{aligned}$$

Model given in (5.2.1) can be rewritten as

$$q_\tau(X_t(s_i), Z_t(s_i)) = Z_t(s_i)' \beta_\tau(X_t(s_i), s_i), \quad (5.2.2)$$

where $t = 1, \dots, T$ and $i = 1, \dots, N$.

For ease of notation, we may simply denote the β_τ as β in the following. Then, we are considering to estimate the unknown function $\beta(\cdot)$ by a local linear fitting. For a given $s = (u, v)'$, we denote $\dot{\beta}^u(x, s) = \partial \beta(x, s) / \partial u$ and $\dot{\beta}^v(x, s) = \partial \beta(x, s) / \partial v$. Then, for any $X_t(s_i)$ and s_i in the neighbourhood of x and s , we have

$$\begin{aligned} \beta(X_t(s_i), s_i) &\approx \beta(x, s) + \dot{\beta}^x(x, s)(X_t(s_i) - x) + \dot{\beta}^u(x, s)(u_i - u) + \dot{\beta}^v(x, s)(v_i - v) \\ &\equiv a_0 + a_1(X_t(s_i) - x) + a_2(u_i - u) + a_3(v_i - v). \end{aligned} \quad (5.2.3)$$

Locally, estimating $(\beta(x, s), \dot{\beta}^x(x, s), \dot{\beta}^u(x, s), \dot{\beta}^v(x, s))$ is equivalent to estimating (a_0, a_1, a_2, a_3) . This motivates us to define an estimator by setting $\widehat{\beta}(x, s) \equiv \widehat{a}_0$, $\widehat{\dot{\beta}^x}(x, s) \equiv \widehat{a}_1$, $\widehat{\dot{\beta}^u}(x, s) \equiv \widehat{a}_2$, and $\widehat{\dot{\beta}^v}(x, s) \equiv \widehat{a}_3$ and considering

$$\begin{aligned} & \min_{a_0, a_1, a_2, a_3} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(Y_t(s_i) - Z_t(s_i)'(a_0 + a_1(X_t(s_i) - x) + a_2(u_i - u) + a_3(v_i - v))) \\ & \quad \times K_h(X_t(s_i) - x) L_b(s_i - s), \end{aligned} \quad (5.2.4)$$

where $K_h(x) = h^{-1}K(x/h)$, K is a kernel function on \mathbb{R}^1 , and $h > 0$ is the bandwidth, and $L_b(s) = b_n^{-2}L(s/b_n)$, L is a kernel function on \mathbb{R}^2 , and $b_n > 0$ is the bandwidth. $\rho_\tau(y) = y(\tau - \mathbf{1}_{y < 0})$ is the traditional check function, $\mathbf{1}_A$ is the indicator function of set A .

Bandwidth Selection

Bandwidth selection is of great importance to weigh between reducing bias and variance in nonparametric quantile regression (c.f., [Yu and Jones, 1998](#); [Cai and Xu, 2008](#)). In this study, we will use the cross-validation method ([Li and Racine, 2004](#)) to choose the optimal bandwidth. The leave-one-out cross-validation estimator under τ -th quantile is

$$CV_\tau(h, b_n) = \frac{1}{TN} \sum_{t=1}^T \sum_{i=1}^N \rho_\tau(Y_t(s_i) - Z_t(s_i)' \check{\beta}(X_{-t}, s_{-i})), \quad (5.2.5)$$

where $\check{\beta}(X_{-t}, s_{-i})$ denotes the local linear estimator obtained from (5.2.4) without $(X_t(s_i), s_i)$ observation. We then select $(h_{\text{opt}}, b_{\text{opt}})$ minimizing $CV_\tau(h, b_n)$ over $h \in [0, h_U]$ and $b_n \in [0, b_U]$, where $0 < h_U$ and $0 < b_U$ are appropriately given. Then, we let

$$h_{\text{opt}}^{b_n} = \arg \min_{1 \leq h \leq h_U} CV_\tau(h, b_n) \quad (5.2.6)$$

and the corresponding minimum for $CV_\tau(h_{\text{opt}}, b_n)$. Thus we can select $b_{n,\text{opt}} = \arg \min_{1 \leq b_n \leq b_U} CV(h_{\text{opt}}^{b_n}, b_n)$, and $h_{\text{opt}} = h_{\text{opt}}^{b_{n,\text{opt}}}$, which are the optimal bandwidth at τ -th quantile.

Discussion

In real applications, one can often specify different spatial weight matrices based on different features of the real data, such as distance based spatial weight matrix W_1 or contiguity based spatial weight matrix W_2 , and so on (c.f., [Anselin, 2013](#)). A significant problem is which spatial weight matrix should be used. Therefore, the model selection procedure, using cross-validation method (5.2.5), can be employed to select an optimal weight matrix among the candidates.

5.3 Asymptotic Theory

In this section, we will provide the asymptotic property for the estimator $\hat{\beta}(x, s)$ in Theorems 5.2 below. First, we use the local Bahadur representation to derive an explicit expression. Let $Y = (Y_1(s_1), \dots, Y_T(s_1), \dots, Y_1(s_N), \dots, Y_T(s_N))'$ denote a NT -dimensional vector. For convenience of the proof, we introduce some notations, and we may simply denote the $\beta(x, s)$ as β . Denote $\theta_0 = a_{NT}^{-1}(\beta - a_0)$

with $a_{NT} = (NThb_n^2)^{-1/2}$, $\theta_1 = ha_{NT}^{-1}(\dot{\beta}^x - a_1)$, $\theta_2 = b_na_{NT}^{-1}(\dot{\beta}^u - a_2)$, $\theta_3 = b_na_{NT}^{-1}(\dot{\beta}^v - a_3)$, then $\beta = a_0 + a_{NT}\theta_0$, $\dot{\beta}^x = a_1 + a_{NT}\theta_1/h$, $\dot{\beta}^u = a_2 + a_{NT}\theta_2/b_n$, and $\dot{\beta}^v = a_3 + a_{NT}\theta_3/b_n$. Define, $Y_{it}^* = Y_t(s_i) - Z_t(s_i)'[\beta(x, s) + \dot{\beta}^x(x, s)(X_t(s_i) - x) + \dot{\beta}^u(x, s)(u_i - u) + \dot{\beta}^v(x, s)(v_i - v)]$, $\mathbf{H} = \text{diag}\{1, h, b_n, b_n\}$,

$$\boldsymbol{\theta} = a_{NT}^{-1}\mathbf{H} \begin{pmatrix} \beta - a_0 \\ \dot{\beta}^x - a_1 \\ \dot{\beta}^u - a_2 \\ \dot{\beta}^v - a_3 \end{pmatrix}, \quad \text{and} \quad G = \begin{pmatrix} 1 \\ X_{ht}(s_i) \\ S_{bi}^u \\ S_{bi}^v \end{pmatrix},$$

where

$$X_{ht}(s_i) = (X_t(s_i) - x)/h, \quad S_{bi}^u = (u_i - u)/b_n, \quad S_{bi}^v = (v_i - v)/b_n. \quad (5.3.1)$$

Thus (5.2.4) can be represented by

$$\Psi(\boldsymbol{\theta}) \equiv \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(Y_{it}^* - a_{NT}\boldsymbol{\theta}^T G Z_t(s_i)) G Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi}). \quad (5.3.2)$$

Then,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(Y_{it}^* - a_{NT}\boldsymbol{\theta}^T G Z_t(s_i)) K(X_{ht}(s_i)) L(S_{bi}) = \underset{\boldsymbol{\theta}}{\text{argmin}} \Psi(\boldsymbol{\theta}). \quad (5.3.3)$$

Now I define the vector function of θ as follows

$$V_{NT}(\boldsymbol{\theta}) = a_{NT} \sum_{i=1}^N \sum_{t=1}^T \psi_\tau(Y_{it}^* - a_{NT}\boldsymbol{\theta}^T G Z_t(s_i)) K(X_{ht}(s_i)) L(S_{bi}), \quad (5.3.4)$$

where $\psi_\tau(y) = \tau - I_{\{y < 0\}}$.

Theorem 5.1. Assume that Assumptions A1–A7 in Section 5.7 are satisfied for some $a \geq 1$, and that the quantile function $\beta(x, s)$ is twice continuously differentiable at x .

$$\begin{aligned} & \sqrt{NThb_n^2}(\hat{\beta}(x, s) - \beta(x, s)) \\ &= \frac{[\Omega^*(x, s)]^{-1}}{\sqrt{NThb_n^2}f_X(x, s)f_S(s)} \sum_{i=1}^N \sum_{t=1}^T \psi_\tau(Y_{it}^*) Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi}) + o_p(1), \end{aligned} \quad (5.3.5)$$

as $T \rightarrow \infty$, where $\psi_\tau(y) = \tau - I_{\{y < 0\}}$, $\Omega^*(x, s) = E[Z_t(s)Z_t(s)'f_{Y|Z,X}(q_\tau(x, Z_t(s)))|X_t(s) = x]$, $f_{Y|Z,X}(\cdot)$ represents the conditional density of Y given Z and X , $f_X(x, s)$ represents the marginal density of X given location s , and $f_S(s)$ represents the spatial sampling density function.

Theorem 5.2. Under Assumptions of Theorem 5.1, we have the following asymptotic normality:

$$\widehat{\beta}(x, s) - \beta(x, s) - B_{TN} = \left\{ (TNhb_n^2)^{-\frac{1}{2}} \Theta_1 + (T)^{-\frac{1}{2}} \Theta_2 \right\} \eta(s)(1 + o_p(1)), \quad (5.3.6)$$

as $T, N \rightarrow \infty$, where

$$\begin{aligned} B_{TN} &= \frac{h^2}{2} \mu_1(x, s) + \frac{b_n^2}{2} \mu_2(x, s), \\ \mu_1(x, s) &= \frac{\partial^2 \beta(x, s)}{\partial x^2} \int u^2 K(u) du, \\ \mu_2(x, s) &= \left(\text{tr} \left\{ \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\}, \dots, \text{tr} \left\{ \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right)', \end{aligned}$$

$\eta(s)$ is a $p + q + 1$ random vector with zero mean and identity variance matrix, and Θ_1 and Θ_2 are two $(p + q + 1) \times (p + q + 1)$ matrices, satisfying

$$\begin{aligned} \Theta_1 \Theta_1' &= \tau(1 - \tau) \frac{[\Omega^*(x, s)]^{-1} \Omega(x, s) [\Omega^*(x, s)]^{-1}}{f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz, \\ \Theta_2 \Theta_2' &= \tau(1 - \tau) \frac{[\Omega^*(x, s)]^{-1} \Omega(x, s, s) [\Omega^*(x, s)]^{-1}}{f_X^2(x, s)} q(x, x; s) \end{aligned}$$

with $\Omega(x, s) = E[Z_t(s)Z_t(s)'|X_t(s) = x]$, $\Omega(x, s_i, s_j) = E[Z_t(s_i)Z_t(s_j)'|X_t(s_i) = x, X_t(s_j) = x]$, $\Omega^*(x, s) = E[Z_t(s)Z_t(s)'f_{Y|Z,X}(q_\tau(x, Z_t(s)))|X_t(s) = x]$, $f_{Y|Z,X}(\cdot)$ represents the conditional density of Y given Z and X , $f_X(x, s)$ represents the marginal density of X given location s , and $f_S(s)$ represents the spatial sampling density function.

Remark 1: Compared with the asymptotic results in Theorem 4.1 in Chapter 4, we can find that the conditional quantile and conditional mean estimators have the same asymptotic bias, but different asymptotic variances.

Remark 2: From Theorem 5.2, the asymptotic mean squared error (AMSE) of $\beta_j(x, s)$ is

$$AMSE_j = \|B_{TN,j}\|^2 + \frac{e_j' \Theta_1 \Theta_1' e_j}{TNhb_n^2} + \frac{e_j' \Theta_2 \Theta_2' e_j}{T}. \quad (5.3.7)$$

Then, the optimal bandwidths are made as follow.

$$\min_{h, b_n} \left[(c_2 h^2 + c_3 b_n^2)^2 + \frac{c_4}{TN h b_n^2} \right]. \quad (5.3.8)$$

This leads to the optimal bandwidths

$$h = \left(\frac{c_3 c_4}{4 c_2^2} \frac{1}{TN} \right)^{\frac{1}{7}},$$

$$b_n = \left(\frac{2^{\frac{3}{2}} c_2^{\frac{3}{2}}}{c_3^{\frac{5}{2}}} \frac{c_4}{TN} \right)^{\frac{1}{7}},$$

where $c_2 = \frac{1}{2} \left\| \frac{\partial^2 \beta_j(x, s)}{\partial x^2} \right\| \int u^2 K(u) du$, $c_3 = \frac{1}{2} \text{tr} \left\{ \frac{\partial^2 \beta_j(x, s)}{\partial s \partial s'} \int z z' \tilde{L}(z) dz \right\}$, and

$$c_4 = \tau(1 - \tau) \frac{e_j' [\Omega^*(x, s)]^{-1} \Omega(x, s) [\Omega^*(x, s)]^{-1} e_j}{f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz$$

Remark 3: Based on the optimal bandwidths, we have $h N b_n^2 = O(N/(TN)^{3/7}) = O((N^4/T^3)^{1/7})$. Therefore, if $h N b_n^2 = O((N^4/T^3)^{1/7}) \rightarrow O(1)$ or $h N b_n^2 = O((N^4/T^3)^{1/7}) \rightarrow \infty$, the rate of the convergence for $\hat{\beta}(x, s)$ is $T^{\frac{1}{2}}$. Otherwise, the rate of the convergence for $\hat{\beta}(x, s)$ is $(Th N b_n^2)^{\frac{1}{2}} = o_p(T^{\frac{1}{2}})$.

5.4 Simulation Study

In this section, we will study the finite-sample performance of our proposed method in a simulation study. For comparison, we consider the following generative model, which is similar to Chapter 4.

$$Y_t(s_i) = \beta_0(X_t(s_i), s_i) + \sum_{j=1}^2 \beta_j(X_t(s_i), s_i) Y_{t-j}^{SL}(s_i) + \beta_3(X_t(s_i), s_i) Y_{t-1}(s_i) + \varepsilon_t(s_i), \quad (5.4.1)$$

where $s_i = (u_i, v_i)$ is the centroid consisting of the latitude and longitude of the i th country, $i = 1, \dots, 23$, in the EU. $Y_t^{SL}(s_i) = \sum_{k=1}^N W_{ik} Y_t(s_k)$, W_{ik} from a distance-function spatial weight matrix W , which units are considered as $w_{ij} = 1/d_{ij}$, where d_{ij} is the Euclidean distance between the centroids of two countries s_i and s_j , and $w_{jj} = 0$, satisfying $\sum_{k=1}^N W_{ik} = 1$. The covariate process $X_t(s_i)$ follows $X_t(s_i) = \alpha(s_i) X_{t-1} + e_t(s_i)$, where $\alpha(s_i) = 0.9 + 0.05 \times \cos(u_i \times v_i)$ for $s_i = (u_i, v_i)'$

is the latitude and longitude of the i -th country, $e_t(s_i)$ follows i.i.d. $N(0, 1)$. Innovation $\varepsilon_t(s_i) = 0.1 \times \xi_t(s_i)$, where $\xi_t(s_i)$ follows i.i.d. t -distribution with 5 degrees of freedom, which is different from Chapter 4.

$$\begin{aligned}\beta_0(X_t(s_i), s_i) &= 0.2 + 0.05 \times X_t(s_i) + b(s_i) \\ \beta_1(X_t(s_i), s_i) &= 0.2 + 0.1 \times \sin(X_t(s_i) + 1) + b(s_i) \\ \beta_2(X_t(s_i), s_i) &= 0.2 + 0.1 \times \cos(X_t(s_i)) + b(s_i) \\ \beta_3(X_t(s_i), s_i) &= 0.3 + 0.1 \times \cos(X_t(s_i) + 1) + b(s_i)\end{aligned}\tag{5.4.2}$$

where $b(s_i) = 0.05 \sin(u_i \times v_i)$ for $s_i = (u_i, v_i)'$ is the latitude and longitude of the i -th country.

We generate data from model (5.4.1) as follows. At each location s_i for $i = 1, \dots, 23$, the initial values of $Y_0(s_t)$ are set to zero. Then we generate $Y_t(s_i)$ for $t = 1, 2, \dots, T + 50$. The first 50 time points are discarded and the next T time points are saved, denoted as $\{(X_t(s_i), Y_t(s_i))$ for $t = 1, \dots, T$, and $i = 1, \dots, N\}$. We consider three time series lengths: $T = 200$, $T = 400$ and $T = 600$. To assess the median quantile estimators of the unknown functions $\beta(X_t(s_i), s_i)$, we select 50 points of x from $(-2, 2)$ (approximately 10th and 90th percentiles of the covariate $X_t(s_i)$). The temporal bandwidth h is 0.5 for $T = 200$, 0.4 for $T = 400$, 0.35 for $T = 600$, respectively and the spatial bandwidth b_n is 7 at all time length.

The performance of estimation will be assessed by defining a squared estimation error (SEE) as a measure of the accuracy of estimation at a location s , which is similar to Chapter 4. That is, for each location s_i , $i = 1, \dots, 23$, we define

$$\text{SEE}(\hat{\beta}_j(\cdot, s_i)) = \frac{1}{50} \sum_{k=1}^{50} \left\{ \hat{\beta}_j(x_k, s_i) - \beta_j(x_k, s_i) \right\}^2, \quad j = 0, 1, 2, 3 \tag{5.4.3}$$

where x_k for $k = 1, \dots, 50$ are 50 points that equally partition the interval between $(-2, 2)$. Similarly, to measure the accuracy of estimation for $\hat{\beta}_j(\cdot)$, we define

$$\text{SEE}(\hat{\beta}_j(\cdot)) = \frac{1}{50 \times 23} \sum_{i=1}^{23} \sum_{k=1}^{50} \left\{ \hat{\beta}_j(x_k, s_i) - \beta_j(x_k, s_i) \right\}^2, \quad j = 0, 1, 2, 3 \tag{5.4.4}$$

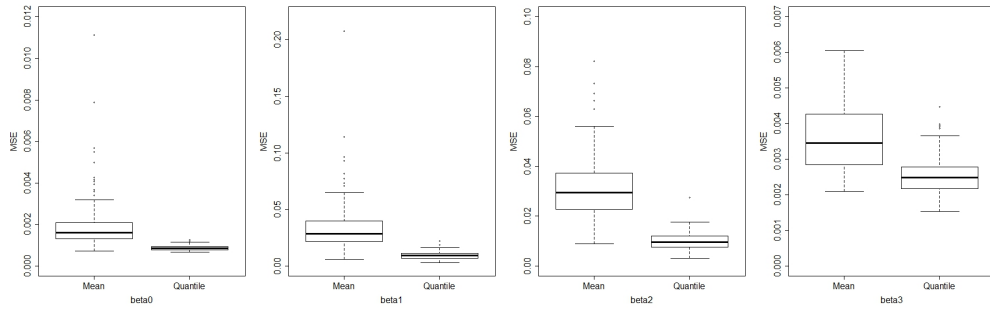


Figure 5.1: Boxplots of 100 times squared estimation error (SEE) for the estimation of $\beta_j(\cdot)$, for $j = 0, 1, 2, 3$, and for conditional mean (left box) and median quantile (right box) with time series length $T = 400$ in each plot.

Firstly, we want to show that conditional median quantile estimation procedure can provide more accurate estimations for $\beta_j(\cdot)$ compared with the conditional mean estimation procedure for non-normal data. We repeat the simulation 100 times with time series length $T = 400$, and thus have 100 values for each SEE of $\beta_j(\cdot)$, $j = 0, 1, 2, 3$, with conditional mean (DyFAST) and median quantile methods, summarized in the boxplots in Figures 5.1. The results indicate that the quantile method can get better estimation results than mean method for non-normal data.

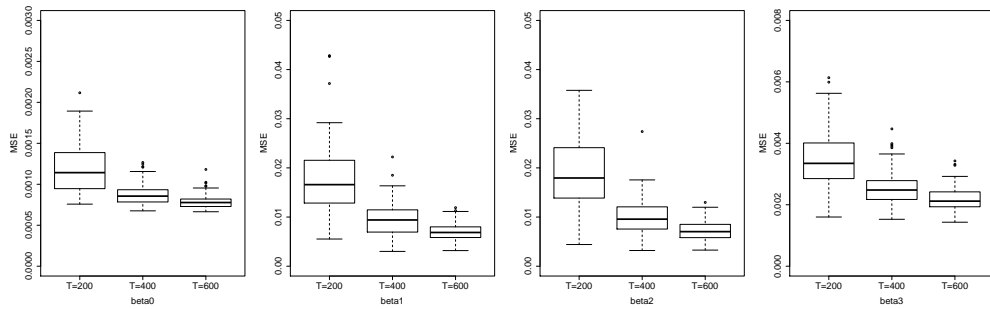


Figure 5.2: Boxplots of 100 times estimation of conditional median quantile estimation for $T = 200$ time points (left box), $T = 400$ time series lengths (middle box) and $T = 600$ time series lengths (right box) in each plot.

Secondly, we consider three time series lengths: $T = 200$, $T = 400$, and $T = 600$. Figure 5.2 shows the squared estimation error (SEE) for the median quantile estimation of $\beta_j(\cdot)$, $j = 0, 1, 2, 3$, in $T = 200$, $T = 400$, and $T = 600$ time points, which appears acceptable. As the sample size increasing, the accuracy of estimate for the $\beta_j(\cdot)$ apparently improve, which indicates the properties of estimators are

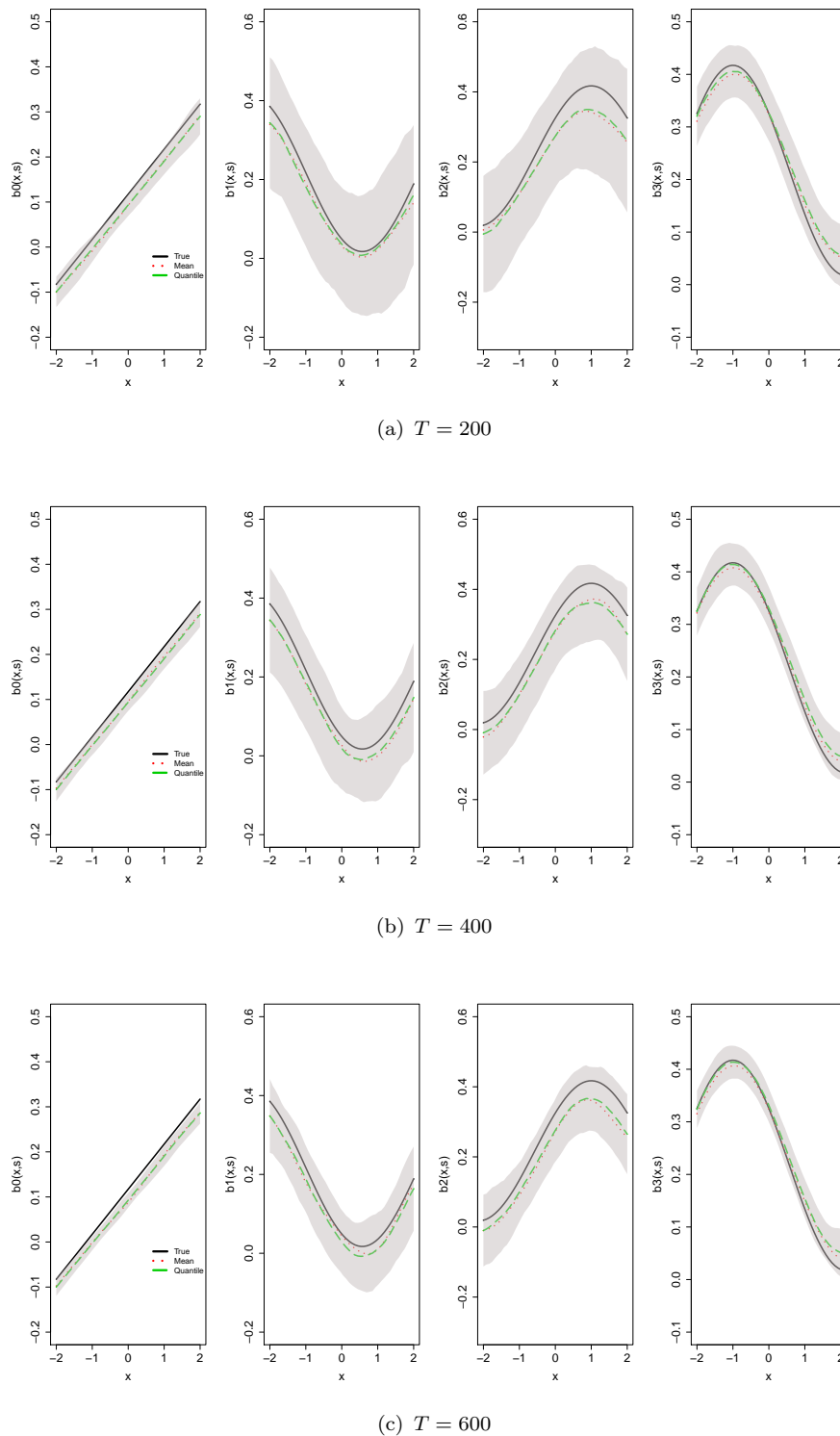


Figure 5.3: Simulation results: the mean of 100 replications at location $i = 1$ of the conditional mean estimation (red lines) and median quantile estimation (green lines) of the coefficient function $\beta_j(\cdot, s_1)$, for $j = 0, 1, 2, 3$ under $T = 200$ time points (top plots), $T = 400$ time series lengths (middle plots) and $T = 600$ time series lengths (bottom plots). In each plot, the solid line shows the true coefficient function. Shaded areas correspond to 95% confidence intervals of median quantile estimation.

asymptotic efficiency. This shows that our method could accurately estimate the coefficient functions for non-normal data.

Thirdly, to derive a more visible conception of the performance of the proposed mean and median quantile estimators, we select its mean performance in 100 replications at location $i = 1$. Figure 5.3 presents the mean of 100 estimated $\beta_j(\cdot)$, $j = 0, 1, 2, 3$ by conditional mean and median methods, compared with the true curves in solid lines. We observe that our estimators $\beta_j(\cdot)$, $j = 0, 1, 2, 3$ mimic the corresponding true curves quite well, which also implies that our quantile method could find the true coefficient functions. Moreover, based on Figure 5.1 and Figure 5.3, we can also find that compared with conditional mean regression, the conditional median quantile regression has the smaller variance for non-normal data. In sum, these findings indicate that our proposed quantile method works very well for non-normal spatial temporal data.

5.5 Real Data Example

In Chapter 4, we have empirically investigated the price transmission of retail gasoline in the EU by our DyFAST. In this section, we want to verify whether the semiparametric varying coefficients spatio-temporal quantile regression model can provide more informations about the retail gasoline markets compared with DyFAST. We first check whether the diesel price returns are normal distribution. The data consist of weekly Brent crude oil price returns (X_t) and diesel price returns (Y_{it}) in 23 European countries from 27/01/2015 to 19/12/2016, which has been considered in Chapter 4. The normality tests for the data are reported in Table 5.1. This table shows that most weekly diesel price returns are negatively skewed and leptokurtic, indicating that it is more likely to observe large negative price changes in the regional diesel markets. The Jarque-Bera test clearly shows that many series depart from normality. These tests indicate the retail gasoline price returns do not comply with the normal distribution. In this case, the classic conditional mean regression will yield less informed results. Therefore, in this section, we will use the semiparametric varying coefficients spatio-temporal quantile regression model to investigate the dynamic linkages between retail gasoline prices in the EU.

To estimate the spatial neighbouring effect for diesel prices, there are different weight matrices. In this study, we employ the distance based spatial weight matrix

Table 5.1: Normality Tests for data.

	Skewness	Kurtosis	JB p-value
Austria	0.1552	-0.2073	0.7956
Belgium	0.0400	0.5886	0.4230
Bulgaria	-2.8406	19.2411	0.0000
Czech.Republic	-0.1349	2.3092	0.0000
Germany	-0.1070	0.1728	0.8168
Denmark	-0.2566	0.5165	0.3050
Estonia	0.2733	1.0790	0.0415
Spain	-0.2092	0.3335	0.5175
Finland	-0.3137	1.2021	0.0191
France	0.3113	-0.0905	0.4672
Greece	-0.6020	5.6492	0.0000
Hungary	-0.3337	-0.0373	0.4171
Italy	-0.4827	2.3879	0.0000
Lithuania	-0.1864	-0.4265	0.5883
Luxembourg	0.2339	-0.2256	0.6190
Latvia	0.2690	4.1546	0.0000
Netherlands	-0.2799	1.2166	0.0208
Poland	-0.0459	-0.0074	0.9771
Portugal	0.2177	0.6072	0.2813
Romania	-0.2918	-0.1013	0.5117
Sweden	-0.1042	-0.2931	0.8289
Slovenia	-0.2010	3.4215	0.0000
Slovakia	-0.4554	1.3192	0.0044

Note: This table displays Skewness and Kurtosis for weekly regional diesel price returns. The sample period is from 27/01/2015 to 19/12/2016. JB p-value is the p-value of empirical statistic of the Jarque-Bera test for normality, which null hypothesis is normal.

W , which has been selected as the optimal spatial weight matrix in Chapter 4. Then, the semiparametric varying coefficients spatio-temporal quantile regression model for price transmission of retail gasoline in the EU is as follow.

$$\begin{aligned}
& Q_{Y_t(s_i)}(\tau | X_t, Y_{t-1}^{sl}(s_i), \dots, Y_{t-p}^{sl}(s_i), Y_{t-1}(s_i), \dots, Y_{t-q}(s_i)) \\
& = a_{0,\tau}(X_t, s_i) + \sum_{j=1}^2 a_{j,\tau}(X_t, s_i) Y_{t-j}^{sl}(s_i) + a_{3,\tau}(X_t, s_i) Y_{t-1}(s_i), \quad (5.5.1) \\
& Y_t^{sl}(s_i) = \sum_{k=1}^N W_{ik} Y_t(s_k), \quad i = 1, \dots, 23, t = 1, \dots, 89
\end{aligned}$$

where X_t denotes the Brent crude oil price return at time t , $Y_t(s_i)$ denotes the return of diesel prices in country i at time t , and $a_{j,\tau}(\cdot), j = 0, 1, 2, 3$ account for the unknown varying coefficients.

By using the cross validation method proposed in Section 5.2, we can obtain the bandwidths for model (5.5.1), which are $h = 3.36$ and $b_n = 0.52$, $h = 2.46$ and $b_n = 0.54$, $h = 3.59$ and $b_n = 0.66$, $h = 2.55$ and $b_n = 1.68$, $h = 9.39$ and $b_n = 3.12$, at 10th, 25th, 50th, 75th and 90th quantiles, respectively. Then, we can get the local linear estimates of the semiparametric varying coefficients spatio-temporal quantile regression at 10th, 25th, 50th, 75th and 90th quantiles, which are plotted in Figure 5.4(a)–5.4(e), respectively.

Compared with the results by DyFAST in Chapter 4 (Figure 4.9), we can find that although the varying trend of effects by the conditional mean and median quantile (Figure 4.9 and 5.4(c)) are approximately similar, the extent of the influences are slightly different. In addition, Figure 5.4(e) clearly shows that these relationships are more than linear at 90th quantile level, which are different from the DyFAST. To be specific, Figure 5.4(e) shows that as the crude oil price returns increase, the temporal lag of spatial neighbouring effects of gasoline price returns decrease at 90th quantile level. These results indicate that the quantile method can provide more robust and efficient estimation results for the non-normal data and reveal the significant distributional (quantile) heterogeneities in the price transmission of retail gasoline in the EU under different crude oil market conditions.

To further evaluate the semiparametric varying coefficients spatio-temporal quantile regression model, we will compare the prediction with the parametric spatio-temporal quantile regression. In addition, due to the fact that nonparametric model may not provide optimal prediction, we also consider the prediction results of β with nonlinear threshold forms. Based on the Figure 5.4, we construct the nonlinear threshold functions of $a_j(x, s)$, $j = 0, 1, 2, 3$ under different quantile levels as follow.

- For $\tau = 0.1$

$$\begin{aligned}
 a_0(x, s) &= a_0^0(s) + a_1^0(s)x \\
 a_1(x, s) &= \{a_{10}^1(s) + a_{11}^1(s)x\} \mathcal{I}(x < -8) + \{a_{20}^1(s) + a_{21}^1(s)x\} \mathcal{I}(-8 \leq x < 0) \\
 &\quad + \{a_{30}^1(s) + a_{31}^1(s)x\} \mathcal{I}(0 \leq x < 5) + \{a_{40}^1(s) + a_{41}^1(s)x\} \mathcal{I}(5 \leq x) \\
 a_2(x, s) &= \{a_{10}^2(s) + a_{11}^2(s)x\} \mathcal{I}(x < -3) + \{a_{20}^2(s) + a_{21}^2(s)x\} \mathcal{I}(-3 \leq x) \\
 a_3(x, s) &= \{a_{10}^3(s) + a_{11}^3(s)x\} \mathcal{I}(x < 6) + \{a_{20}^3(s) + a_{21}^3(s)x\} \mathcal{I}(6 \leq x)
 \end{aligned}$$

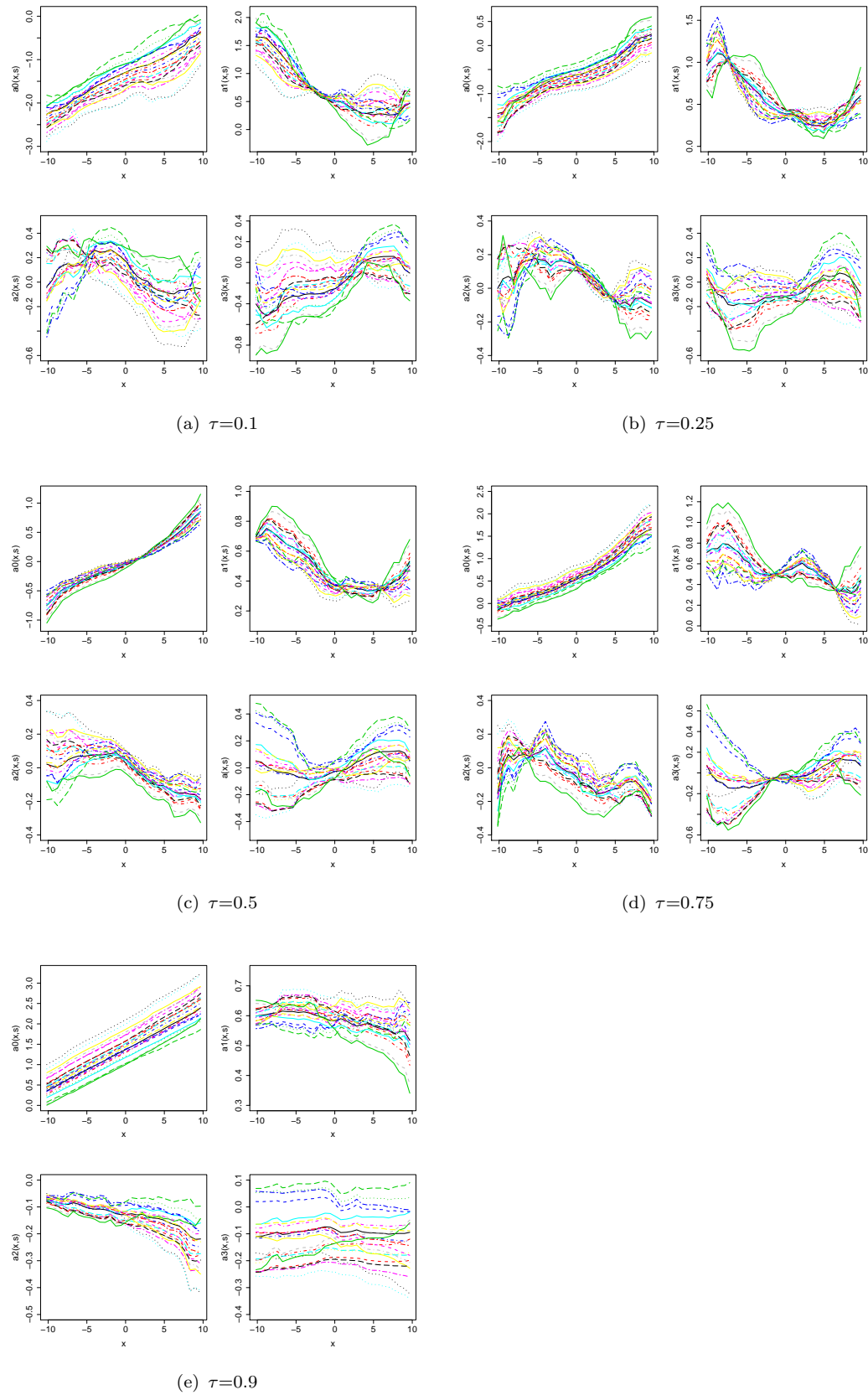


Figure 5.4: The results of semiparametric varying coefficients spatio-temporal quantile regression for the price transmission of retail gasoline in the EU at 10th, 25th, 50th, 75th and 90th quantiles.

- For $\tau = 0.25, 0.5, 0.75$

$$\begin{aligned}
 a_0(x, s) &= a_0^0(s) + a_1^0(s)x \\
 a_1(x, s) &= \{a_{10}^1(s) + a_{11}^1(s)x\} \mathcal{I}(x < -8) + \{a_{20}^1(s) + a_{21}^1(s)x\} \mathcal{I}(-8 \leq x < 0) \\
 &\quad + \{a_{30}^1(s) + a_{31}^1(s)x\} \mathcal{I}(0 \leq x < 5) + \{a_{40}^1(s) + a_{41}^1(s)x\} \mathcal{I}(5 \leq x) \\
 a_2(x, s) &= \{a_{10}^2(s) + a_{11}^2(s)x\} \mathcal{I}(x < 0) + \{a_{20}^2(s) + a_{21}^2(s)x\} \mathcal{I}(0 \leq x) \\
 a_3(x, s) &= \{a_{10}^3(s) + a_{11}^3(s)x\} \mathcal{I}(x < -3) + \{a_{20}^3(s) + a_{21}^3(s)x\} \mathcal{I}(-3 \leq x)
 \end{aligned}$$

- For $\tau = 0.9$

$$a_j(x, s) = a_0^j(s) + a_1^j(s)x, \quad j = 0, 1, 2, 3 \quad (5.5.2)$$

where $I(\cdot)$ is an indicator function, $a^l(s)$, $l = 0, 1, 2, 3$, are the spatially-varying linear coefficients. Figure 5.5 represents the results of nonlinear threshold model for the retail gasoline price returns in the EU at 50th quantile. These results indicate that the spatial neighbouring effects of retail gasoline price returns are more complex than linear.

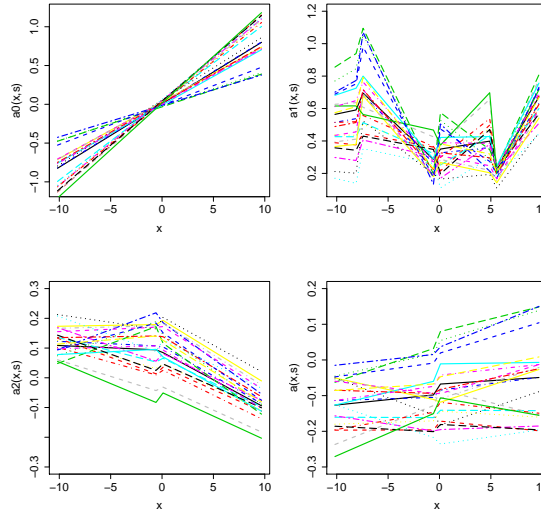


Figure 5.5: The results of nonlinear threshold model for the price transmission of retail gasoline in the EU at 50th quantile.

We set aside the last 18 data for evaluation and use first $T = 71$ data as training set. A quantile prediction error (QPE) of the prediction is computed for the two

Table 5.2: Comparison of prediction results

Methods	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Linear regression	0.231	0.408	0.503	0.402	0.237
Semiparametric regression	0.218	0.388	0.458	0.366	0.209
Nonlinear threshold regression	0.218	0.376	0.450	0.358	0.204

Note: This table shows the quantile prediction error (QPE) for linear model, semiparametric varying coefficients spatio-temporal quantile regression and nonlinear threshold model. The bandwidths for nonlinear threshold model are 0.442, 0.625, 0.472, 0.380 and 0.306 at 10th, 25th, 50th, 75th and 90th quantiles, respectively.

models under ρ th quantile. Here, we define QPE as:

$$QPE(\hat{Y}_\tau(\cdot)) = \frac{1}{N \times T} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(Y_t(s_i) - \hat{Y}_{\tau,t}(s_i)), \quad (5.5.3)$$

where $\rho_\tau(y) = y(\tau - I_{\{y < 0\}})$ and I_A is the indicator function of set A . Then, we can obtain the corresponding quantile prediction error (QPE) of the one-step ahead prediction, which are shown in Table 5.2. Overall, compared with the linear model, the nonlinear threshold model outperforms in prediction. Especially, it offers a 13.9% performance increase compared the linear model at 90th quantile. In sum, these results illustrate that semiparametric varying coefficients spatio-temporal quantile regression can help to uncover the spatial neighbouring effects of retail gasoline price returns, which are more complex than linear and different under different situations of crude oil market.

5.6 Conclusions

This chapter proposes a semiparametric varying coefficients spatio-temporal quantile regression to investigate the nonlinear relationship in non-normal spatio-temporal data. Compared with DyFAST in Chapter 4, this approach is more explicable and robust for the non-normal spatio-temporal data. Both theoretical properties and Monte Carlo simulations are investigated. The proposed method is also applied to empirically investigate the retail gasoline markets in the EU, which further illustrates the usefulness of our models.

5.7 Appendix

Regularity Assumptions

- (A1) (i) For $t = 0, \pm 1, \pm 2, \pm 3, \dots$, $\mathbf{X}_t = (X_t(s_1), \dots, X_t(s_N))$ is α -mixing process
(ii) For some real positive number $\delta > 2$ and $c > 1 - 2/\delta$, the α -mixing coefficient $\alpha(\cdot)$ satisfies

$$\lim_{k \rightarrow \infty} k^c \sum_{k=d_N}^{\infty} \{\alpha(k)\}^{1-2/\delta} = 0$$

(iii) $X_i(s)$ and $X_j(s)$ have a joint probability density function $p(x_i, x_j; s)$, which is continuous and bounded for all $i \neq j$. For each t , the joint probability density function of $X_t(s_1)$ and $X_t(s_2)$ satisfies the following limit: $\lim_{s_1, s_2 \rightarrow s} p(x_1, x_2; s_1, s_2) = q(x_1, x_2; s)$, where $q(x_1, x_2; s)$ is continuous and bounded with respect to both x_1 and x_2 .

(iv) The marginal density of X given location s , $f_X(x, s)$, is continuous, and $0 < f_X(x, s) \leq M < \infty$.

(v) The conditional density of Y given X and Z , $f_{y|z,x}(y)$, is bounded and satisfies the Lipschitz condition.

- (A2) The functions $\beta_j(x, s)$, $j = 1, \dots, p + q + 1$ are continuous at all x and twice differentiable w.r.t. x and $s \in S$.

- (A3) (i) At each location $s \in S$, the innovations $\varepsilon_t(s)$ are *i.i.d.* random variables independent of $\{X_t(s)\}$. Moreover, for each $t > 0$, $\{\varepsilon_t(s)\}_{s \in S}$ are independent of $\{Y_{t-i}^{\text{SL}}(s)\}_{s \in S}$ for $i = 1, \dots, p$, and $\{Y_{t-l}(s)\}_{s \in S}$ for $l = 1, \dots, q$. $E[|Z_t(s)|^{2+\delta}] < \infty$, for some $\delta > 2$.

(ii) For each $t > 0$, the spatial covariance function $\Gamma_t(s_l, s_k) \equiv \text{Cov}[\varepsilon_t(s_l), \varepsilon_t(s_k)]$ is bounded over $S \times S$.

(iii) For each $t \geq 1$ and $s \in S$, $\varepsilon_t(s) = \varepsilon_{1,t}(s) + \varepsilon_{2,t}(s)$, where $\{\varepsilon_{1,t}(s)\}$ and $\{\varepsilon_{2,t}(s)\}$ are two independent processes and both satisfy the condition (A3)(i). Further, $\Gamma_{1t}(s_j, s_k) \equiv \text{Cov}[\varepsilon_{1,t}(s_1), \varepsilon_{1,t}(s_2)]$ is continuous in (s_1, s_2) and $\Gamma_{2t}(s_1, s_2) \equiv \text{Cov}[\varepsilon_{2,t}(s_1), \varepsilon_{2,t}(s_2)] = 0$ if $s_1 \neq s_2$ and $\Gamma_{2t}(s_1, s_2) = \sigma_2^2(s_j) > 0$ is continuous in s_1 .

- (iv) For each t , the $\Omega(x, s_1, s_2) = E[Z_t(s_1)Z_t(s_2)|x] = \Omega_1(x, s_1, s_2) + \Omega_2(x, s_1, s_2)$, where $\Omega_1(x, s_1, s_2)$ is continuous in (s_1, s_2) , and $\Omega_2(x, s_1, s_2) = 0$ if $s_1 \neq s_2$ and $\Omega_2(x, s_1, s_1) = \Omega_2(x, s_1) > 0$. is continuous in s_1 .
- (A4) For any measurable set $A \subset S \subset \mathbb{R}^2$, $N^{-1} \sum_{j=1}^N \mathcal{I}(s_j \in A) \longrightarrow \int_A f_S(s)ds$, $N \rightarrow \infty$, where the density function f satisfies $f > 0$ in a neighbourhood of $s \in S$.
- (A5) (i) The kernel function $K(\cdot)$ is symmetric, uniformly bounded by some constant, and integrable. Further, $\int K(u)du = 1$ and $\int u^2 K(u)du < \infty$.
- (ii) The kernel function $L(\cdot)$ is symmetric and satisfies $\int_{\mathbb{R}^2} L(z)dz = 1$, $\int_{\mathbb{R}^2} zL(z)dz = 0$ and $\int_{\mathbb{R}^2} zz'L(z)dz < \infty$.
- (A6) (i) The temporal bandwidth $h \rightarrow 0$ in such a way that $Th \rightarrow \infty$. The spatial smoothing bandwidth $b_n \rightarrow 0$, $Nb_n^2 \rightarrow \infty$, as $N \rightarrow \infty$.
- (ii) There exists a sequence of positive integers q_T such that $q_T \rightarrow \infty$, $q_T = o\left((Th)^{1/2}\right)$, and $(T/h)^{1/2} \alpha(q_T) \rightarrow 0$, as $T \rightarrow \infty$.
- (iii) $T^{1/2-\delta/4} h^{\delta/\delta^*-1/2-\delta/4} b_n^{2\delta/\delta^*-\delta} = O(1)$ where $\delta^* > \delta > 2$, where δ is given in Assumption A1.
- (A7) $\Omega^*(x, s)$ and $\Omega(x, s)$ are positive definite and continuous in a neighbourhood of x and s .

Assumption (A1)(i)&(ii) assume that the process $X_t(s_i)$ is α -mixing, which is standard in nonlinear time series analysis. Assumption (A1)(iii) assumes the boundedness of the joint probability density functions, which can ensure the consistency w.r.t. the different times and locations. Assumption (A1)(iv) assumes marginal density of $X_t(s_i)$. Assumption (A2) is a standard condition on the estimated function in nonlinear modelling, which ensures the smoothness of the estimated function. Assumption (A3) assumes the independence of innovation, boundedness of spatial covariance function, and the finite moment of $Z_t(s)$. Indeed, for independent data nonparametric quantile estimation, there is no need to impose moment condition for Y in general. However, in this study, both spatio-temporally lagged and autoregressive variables have been contained in $Z_t(s)$, and $Z_t(s)$ should satisfy the moment condition. Therefore, Assumptions (A3)(i) in Chapter 5 imposes moment condition for $Z_t(s)$, which is different from the mean regression (DyFAST) in Chapter 4. Assumption (A4) assumes the spatial sampling density, which is a standard condition in previous literature (e.g., [Lu et al., 2009](#); [Al-Sulami et al., 2017a](#)). This indicates that the asymptotics belong to the category of the fixed-domain or in-fill asymptotics. Assumption (A5) assumes standard conditions for

the kernel functions, which are commonly used in nonparametric kernel estimation. Assumption (A6)(i) assumes the conditions on the bandwidths h_1 and h_2 , which are standard in nonparametric time series or spatial analysis. Assumption (A6)(ii)&(iii) are mainly required to establish the asymptotic normality. Indeed, these assumptions are also imposed for the mean regression (DyFAST) in Chapter 4. Assumption (A7) is the necessary and sufficient condition to ensure that $\beta(x, s)$ is identified. These conditions in (A1)–(A7) are fairly mild and used in the literature (c.f. Zhang et al., 2003; Fan and Yao, 2003; Cai and Xu, 2008; Lu et al., 2009; Al-Sulami et al., 2017a).

Some basic Lemmas

Lemma 5.1. *Let $V_{NT}(\Delta)$ be a vector function that satisfies*

$$(i) -\Delta' V_{NT}(\lambda \Delta) \geq -\Delta' V_{NT}(\Delta), \lambda \geq 1,$$

$$(ii) \sup_{\|\Delta\| \leq M} \|V_{NT}(\Delta) + D\Delta - A_{NT}\| = o_P(1),$$

where $\|A_{NT}\| = O_P(1)$, $0 < M < \infty$, and D is a positive-definite matrix. Suppose that Δ_{NT} such that $\|V_{NT}(\Delta_{NT})\| = o_P(1)$. Then, $\|\Delta_{NT}\| = O_P(1)$ and

$$\Delta_{NT} = D^{-1}A_n + o_P(1).$$

Proof. The proof follows from Koenker and Zhao (1996) and Cai and Xu (2008).

Lemma 5.2. *Under Assumptions (A4) and (A6), for any function $\varphi(s_1, s_2)$ defined on $\mathbb{R}^2 \times \mathbb{R}^2$, which is continuous at (s_0, s_0) , we have*

$$\sum_{i=1}^N \sum_{j=1, j \neq i}^N \varphi(s_i, s_j) L_{b_n}(s_i - s_0) L_{b_n}(s_j - s_0) = N^2 \varphi(s_0, s_0) f^2(s_0) (1 + o(1)) \quad (\text{A.1})$$

Proof: This lemma is borrowed from Lu et al. (2009), where we refer to for a proof.

Lemma 5.3. *Let $\hat{\beta}$ be the minimizer of the function*

$$\sum_{i=1}^N \sum_{t=1}^T w_{NT} \rho_{\tau}(y_{NT} - Z'_{NT} \beta(x, s)),$$

where $w_{NT} > 0$. Then,

$$\left\| \sum_{i=1}^N \sum_{t=1}^T w_{NT} Z_{NT} \psi_{\tau}(y_{NT} - Z'_{NT} \beta(x, s)) \right\| \leq \dim(Z_{NT}) \max_{t \leq T, i \leq N} \|w_{NT} Z_{NT}\|$$

Proof. The proof follows from [Ruppert and Carroll \(1980\)](#).

To get the Bahadur representation for $\hat{\theta}$, we need to verify whether $V_{NT}(\theta)$ satisfies Lemma [5.1](#).

Lemma 5.4. *Under Assumptions in Theorem [5.1](#),*

$$\sup_{\|\theta\| \leq M} \|V_{NT}(\theta) - V_{NT}(0) - E(V_{NT}(\theta) - V_{NT}(0))\| = o_P(1). \quad (\text{A.2})$$

Proof. The proof is divided into two steps. First we prove that for any fixed $\theta : \|\theta\| \leq M$,

$$\|V_{NT}(\theta) - V_{NT}(0) - E(V_{NT}(\theta) - V_{NT}(0))\| = o_P(1). \quad (\text{A.3})$$

Note that

$$\begin{aligned} V_{NT}(\theta) - V_{NT}(0) &= a_{NT} \sum_{i=1}^N \sum_{t=1}^T [\psi_{\tau}(Y_{it}^*(\theta)) - \psi_{\tau}(Y_{it}^*)] Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi}) \\ &= a_{NT} \sum_{t=1}^T \sum_{i=1}^N L(S_{bi}) V_{it}(\theta) \\ &= a_T \sum_{t=1}^T V_t(\theta), \end{aligned} \quad (\text{A.4})$$

where $V_t(\theta) = a_N \sum_{i=1}^N [\psi_{\tau}(Y_{it}^*(\theta)) - \psi_{\tau}(Y_{it}^*)] Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi})$, $a_T = (Th)^{-1/2}$, $a_N = (Nb_2^2)^{-1/2}$ and $Y_{it}^*(\theta) = Y_{it}^* - a_{NT} \theta G Z_t(s_i)$.

Then the left-hand side of [\(A.3\)](#) is bounded by

$$a_T \left| \sum_{t=1}^T (V_t(\theta) - EV_t(\theta)) \right| \triangleq V_T. \quad (\text{A.5})$$

Then considering the variance of V_T , we have

$$\begin{aligned}
 E(V_T)^2 &= (Th)^{-1} \left[\sum_{t=1}^T \text{var}(V_t(\theta)) + 2 \sum_{t=1}^{T-1} \sum_{j=t+1}^T \text{cov}(V_t(\theta), V_j(\theta)) \right] \\
 &\leq h^{-1} \left[\text{var}(V_1(\theta)) + 2 \sum_{j=1}^{d_T} |\text{cov}(V_1(\theta), V_{j+1}(\theta))| \right. \\
 &\quad \left. + 2 \sum_{j=d_T}^{\infty} |\text{cov}(V_1(\theta), V_{j+1}(\theta))| \right] \\
 &= J_1 + J_2 + J_3,
 \end{aligned} \tag{A.6}$$

where d_T is a positive integer depending on T to be specified later. First, for J_3 , with the Davydov's inequality (see, Corollary A.2 of [Hall and Heyde, 1980](#)), one can obtain

$$|\text{cov}(V_1(\theta), V_{j+1}(\theta))| \leq C\alpha^{1-2/\delta}(s) \left[E |V_1(\theta)|^\delta \right]^{2/\delta}. \tag{A.7}$$

$$\begin{aligned}
 E |V_1(\theta)|^\delta &\leq a_N \sum_{i=1}^N E |\psi_\tau(Y_{it}^*(\theta)) - \psi_\tau(Y_{it}^*)|^\delta |Z_t(s_i)|^\delta K^\delta(X_{ht}(s_i)) L^\delta(S_{bi}) \\
 &\leq Ca_N a_{NT} \sum_{i=1}^N E \left[|\theta' G Z_t(s_i)| |Z_t(s_i)|^\delta K^\delta(X_{ht}(s_i)) L^\delta(S_{bi}) \right] \\
 &\leq Ca_T E \left[|\theta' G Z_t(s_i)| |Z_t(s_i)|^\delta K^\delta(X_{ht}(s_i)) f(s) \right] \\
 &\leq Ca_T h.
 \end{aligned} \tag{A.8}$$

Then we have

$$\begin{aligned}
 J_3 &\leq C(a_T)^{2/\delta} h^{2/\delta-1} \sum_{s=d_T}^{\infty} [\alpha(s)]^{1-2/\delta} \\
 &\leq C(a_T)^{2/\delta} h^{2/\delta-1} d_T^{-l} \sum_{s=d_T}^{\infty} s^l [\alpha(s)]^{1-2/\delta}
 \end{aligned} \tag{A.9}$$

$$= o((a_T)^{2/\delta} h^{2/\delta-1} d_T^{-l}) = o(1), \tag{A.10}$$

by choosing d_T such that $d_T h^{1-2/\delta} = C$.

For J_2 , using Lemma 5.2, we have

$$\begin{aligned}
& h^{-1} \sum_{j=1}^{d_T} |\text{cov}(V_1(\theta), V_{j+1}(\theta))| \\
& \leq d_T h^{-1} C a_N^2 \sum_{l=1}^N \sum_{l \neq k=1}^N E |Z_1(s_l) Z_{j+1}(s_k)'| K(X_{h1}(s_i)) K(X_{h,j+1}(s_i)) L(S_{bl}) L(S_{bk}) + C d_T h \\
& \leq C h d_T = o(1).
\end{aligned} \tag{A.11}$$

Last, we have

$$J_1 = h^{-1} \text{var}(V_1(\theta)) \leq h^{-1} E(V_1(\theta))^2 = O((Th)^{-1}) = o(1). \tag{A.12}$$

Then, it follows that for any fixed $\theta : \|\theta\| \leq M$,

$$\|V_{NT}(\theta) - V_{NT}(0) - E(V_{NT}(\theta) - V_{NT}(0))\| = o_P(1). \tag{A.13}$$

The next step is to use [Bickel \(1975\)](#) standard chaining argument to show the above result holds uniformly in $\theta : \|\theta\| \leq M$.

$$\sup_{\|\theta\| \leq M} \|V_{NT}(\theta) - V_{NT}(0) - E(V_{NT}(\theta) - V_{NT}(0))\| = o_p(1). \tag{A.14}$$

Decompose $\{\|\theta\| \leq M\}$ into cubes based on the grid $(j_1 \gamma M, \dots, j_{p+1} \gamma M)$, $j_i = 0, \pm 1, \dots, \pm[1/\gamma] + 1$, where $[1/\gamma]$ denotes the integer part of $1/\gamma$, and γ is a small positive number independent of n . Let $R(\theta)$ be the lower vertex of the cube that contains θ . Clearly, $\|R(\theta) - \theta\| \leq C\gamma$ and the number of the elements of $\{R(\theta) : \|\theta\| \leq M\}$ is finite. Then

$$\sup_{\|\theta\| \leq M} \|V_{NT}(\theta) - V_{NT}(0) - E(V_{NT}(\theta) - V_{NT}(0))\| \leq V_1^* + V_2^* + V_3^*, \tag{A.15}$$

where $V_1^* = \sup_{\|\theta\| \leq M} \|V_{NT}(R(\theta)) - V_{NT}(0) - E(V_{NT}(R(\theta)) - V_{NT}(0))\| = o_P(1)$, $V_2^* = \sup_{\|\theta\| \leq M} \|V_{NT}(\theta) - V_{NT}(R(\theta))\|$, and $V_3^* = \sup_{\|\theta\| \leq M} \|E(V_{NT}(\theta) - V_{NT}(R(\theta)))\|$.

For V_1^* , it follows easily from [\(A.3\)](#) that

$$\sup_{\|\theta\| \leq M} \|V_{NT}(R(\theta)) - V_{NT}(0) - E(V_{NT}(R(\theta)) - V_{NT}(0))\| = o_P(1).$$

For V_2^* , we have

$$\begin{aligned}
 & \sup_{\|\theta\| \leq M} \|V_{NT}(\theta) - V_{NT}(R(\theta))\| \\
 &= a_{NT} \sup_{\|\theta\| \leq M} \left\| \sum_{i=1}^N \sum_{t=1}^T [\psi_\tau(Y_{it}^*(\theta)) - \psi_\tau(Y_{it}^*R(\theta))] Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi}) \right\| \\
 &\leq C\gamma + a_{NT} \sup_{\|\theta\| \leq M} \left\| \sum_{i=1}^N \sum_{t=1}^T [I_{|Y_{it}^*R(\theta)| < a_{NT}C\gamma} - EI_{|Y_{it}^*R(\theta)| < a_{NT}C\gamma}] Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi}) \right\|
 \end{aligned} \tag{A.16}$$

Since the number of the elements in $\{R(\theta) : \|\theta\| \leq M\}$ is finite, we can show that

$$a_{NT} \sup_{\|\theta\| \leq M} \left\| \sum_{i=1}^N \sum_{t=1}^T [I_{|Y_{it}^*R(\theta)| < a_{NT}C\gamma} - EI_{|Y_{it}^*R(\theta)| < a_{NT}C\gamma}] Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi}) \right\| = o_p(1), \tag{A.17}$$

which is similar to the proof of (A.3). Then, let $\gamma \rightarrow 0$, we have

$$\sup_{\|\theta\| \leq M} \|V_{NT}(\theta) - V_{NT}(R(\theta))\| = o_p(1). \tag{A.18}$$

Last, we will consider the V_3^* .

$$\begin{aligned}
 & \sup_{\|\theta\| \leq M} \|E(V_{NT}(\theta) - V_{NT}(R(\theta)))\| \\
 &= NT a_{NT} \sup_{\|\theta\| \leq M} \|EI_{|Y_{it}^*R(\theta)| < a_{NT}C\gamma} Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi})\| \leq C\gamma.
 \end{aligned} \tag{A.19}$$

Let $\gamma \rightarrow 0$, we have

$$\sup_{\|\theta\| \leq M} \|E(V_{NT}(\theta) - V_{NT}(R(\theta)))\| = o_p(1).$$

Finally, we have

$$\sup_{\|\theta\| \leq M} \|V_{NT}(\theta) - V_{NT}(0) - E(V_{NT}(\theta) - V_{NT}(0))\| = o_p(1).$$

The proof of Lemma is complete.

Lemma 5.5. *Under the above Assumptions,*

$$\sup_{\|\theta\| \leq M} \|E(V_{NT}(\theta) - V_{NT}(0)) + D\theta\| = o(1),$$

where $D = \Omega^*(x, s)f_X(x, s)f_S(s)$.

Proof. The proof follows from Xu (2005).

Proofs for theorems

Proof of Theorems 5.1 We now check the conditions of Lemma 5.1(ii) $\|V_{NT}(\hat{\theta})\| = o_P(1)$ follows from the Lemmata above together with Assumptions. Take $A_{NT} = V_{NT}(0)$. It can be seen that $A_{NT} = O_P(1)$. Since $\psi_\tau(y)$ is an increasing function of y , the function

$$-\theta' V_{NT}(\lambda\theta) = a_{NT} \sum_{i=1}^N \sum_{t=1}^T \psi_\tau(Y_{it}^* - \lambda a_{NT} \theta' G Z_t(s_i)) (-\theta' G Z_t(s_i)) K(X_{ht}) L(S_{bi})$$

is increasing as a function of λ . Therefore, condition (i) of Lemma 5.1 holds and the result follows.

$$\begin{aligned} \hat{\theta} &= D^{-1} A_{NT} + o_p(1) \\ &= \frac{[\Omega^*(x, s)]^{-1}}{\sqrt{NT} h b_n^2 f_X(x, s) f_S(s)} \sum_{i=1}^N \sum_{t=1}^T \psi_\tau(Y_{it}^*) G Z_t(s_i) K(X_{ht}) L(S_{bi}) + o_p(1), \quad (\text{A.20}) \end{aligned}$$

Proof of Theorems 5.2 Let $\varepsilon_{it} = \psi_\tau(Y_t(s_i) - Z_t(s_i)' a_0)$. Then, $E(\varepsilon_{it}) = 0$, and $\text{Var}(\varepsilon_{it}) = \tau(1 - \tau)$. e_0 denotes a $4(p + q + 1) \times 1$ unit vector with 1 at the 1 to $p + q + 1$ th position.

$$\begin{aligned} a_{NT}^{-1} e_0' \hat{\theta} &\approx \frac{[\Omega^*(x, s)]^{-1}}{NT h b_n^2 f_X(x, s) f_S(s)} \sum_{i=1}^N \sum_{t=1}^T [\psi_\tau(Y_{it}^*) - \varepsilon_{it}] Z_t(s_i) K(X_{ht}) L(S_{bi}) \\ &\quad + \frac{[\Omega^*(x, s)]^{-1}}{NT h b_n^2 f_X(x, s) f_S(s)} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it} Z_t(s_i) K(X_{ht}) L(S_{bi}) \\ &\equiv \mathbf{B} + \boldsymbol{\xi}_{TN}. \end{aligned} \quad (\text{A.21})$$

First, we calculate \mathbf{B} and let $M_{it} = \psi_\tau(Y_{it}^*) Z_t(s_i) K(X_{ht}(s_i)) L(S_{bi})$.

$$\begin{aligned}
 & \sum_{i=1}^N EM_{i1} \\
 &= \sum_{i=1}^N E \psi_\tau(Y_{i1}^*) Z_1(s_i) K(X_{h1}(s_i)) L(S_{bi}) \\
 &= \sum_{i=1}^N E \left(\tau - I_{\{Y_{i1}^* < 0\}} \right) Z_1(s_i) K(X_{h1}(s_i)) L(S_{bi}) \\
 &= \sum_{i=1}^N E \left[\tau - F_{y|z,x}(Z_1(s_i)'(a_0 + a_1(X_1(s_i) - x) + a_2(u_i - u) + a_3(v_i - v)) \mid X_1(s_i), Z_1(s_i)) \right] \\
 & \quad \times Z_1(s_i) K(X_{h1}(s_i)) L(S_{bi}) \\
 &= \sum_{i=1}^N E f_{y|z,x} \left(q_\tau(x, Z_1(s_i)) + Z_1(s_i)'(a_1(X_1(s_i) - x) + a_2(u_i - u) + a_3(v_i - v)) \right. \\
 & \quad \left. + \xi \Lambda \mid X_1(s_i), Z_1(s_i) \right) \Lambda Z_1(s_i) K(X_{h1}(s_i)) L(S_{bi}), \tag{A.22}
 \end{aligned}$$

where with the Taylor expansion of $q_\tau(X_1(s_i), Z_1(s_i))$ at (x, s) ,

$$\begin{aligned}
 \Lambda &= \frac{1}{2} (X_1(s_i) - x)^2 Z_1(s_i)' \frac{\partial^2 \beta(\xi_1, \xi_2)}{\partial \xi_1^2} + (X_1(s_i) - x) Z_1(s_i)' \frac{\partial^2 \beta(\xi_1, \xi_2)}{\partial \xi_1 \partial \xi_2'} (s_i - s) \\
 & \quad + \frac{1}{2} \sum_{j=0}^{p+q} (s_i - s)' \frac{\partial^2 \beta_j(\xi_1, \xi_2)}{\partial \xi_2 \partial \xi_2'} (s_i - s) Z_{1,j}, \tag{A.23}
 \end{aligned}$$

with $\xi_1 = x + \eta_1(X_1(s_i) - x)$ with $|\eta_1| < 1$ and $\xi_2 = s + \eta_2(s_i - s)$ with $|\eta_2| < 1$. Therefore, similar to the proof of Lemma 4.2, one can obtain

$$\begin{aligned}
 & \sum_{i=1}^N EM_{i1} \\
 &= \frac{1}{2} \left[h^3 N b_n^2 f_X(x, s) f_S(s) \Omega^*(x, s) \frac{\partial^2 \beta(x, s)}{\partial x^2} \int u^2 K(u) du \right. \\
 & \quad \left. + h N b_n^4 f_X(x, s) f_S(s) \Omega^*(x, s) \left(\text{tr} \left\{ \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\}, \dots, \right. \right. \\
 & \quad \left. \left. \text{tr} \left\{ \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right) \right] \{1 + o(1)\}. \tag{A.24}
 \end{aligned}$$

$$\begin{aligned}
E[\mathbf{B}] &= \frac{[\Omega^*(x, s)]^{-1}}{Nhb_n^2 f_X(x, s) f_S(s)} \sum_{i=1}^N E[\psi_\tau(Y_{it}^*) Z_t(s_i) K(X_{ht}(s_i)) | X_t(s_i)] L(S_{bi}) \{1 + o(1)\} \\
&= \frac{1}{2} \left[h^2 \frac{\partial^2 \beta(x, s)}{\partial x^2} \int u^2 K(u) du + b_n^2 \left(\text{tr} \left\{ \frac{\partial^2 \beta_0(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\}, \dots, \right. \right. \\
&\quad \left. \left. \text{tr} \left\{ \frac{\partial^2 \beta_{p+q}(x, s)}{\partial s \partial s'} \int z z' L(z) dz \right\} \right)' \right] \{1 + o(1)\}.
\end{aligned} \tag{A.25}$$

Since $\psi_\tau(Y_{it}^*) - \varepsilon_{it} = I(Y_t(s_i) \leq Z_t(s_i)'[a_0 + a_1(X_t(s_i) - x) + a_2(u_i - u) + a_3(v_i - v)]) - I(Y_t(s_i) \leq Z_t(s_i)'a_0)$, then

$$[\psi_\tau(Y_{it}^*) - \varepsilon_{it}]^2 = I(d_{1t} < Y_{it} \leq d_{2t}),$$

where $d_{1t} = \min(c_{1t}, c_{2t})$ and $d_{2t} = \max(c_{1t}, c_{2t})$ with $c_{1t} = Z_t(s_i)'[a_0 + a_1(X_t(s_i) - x) + a_2(u_i - u) + a_3(v_i - v)]$ and $c_{2t} = Z_t(s_i)'a_0$. Further

$$E[\{\psi_\tau(Y_{it}^*) - \varepsilon_{it}\}^2 K^2(X_{ht}(s_i)) L^2(S_{bi}) Z_t(s_i) Z_t(s_i)'] \tag{A.26}$$

$$= E[\{F_{y|z,x}(d_{2t}) - F_{y|z,x}(d_{1t})\} K^2(X_{ht}(s_i)) L^2(S_{bi}) Z_t(s_i) Z_t(s_i)'] \tag{A.27}$$

$$= O(h^3 b_n^6). \tag{A.28}$$

Thus, $\text{Var}(\mathbf{B}) = o(1)$.

Similar to the proof of Theorem 4.1 in Chapter 4, considering the small-block and large-block technique and the Cramér-Wold device, we can show that

$$\boldsymbol{\xi}_{TN} = \left\{ (TNhb_n^2)^{-\frac{1}{2}} \Theta_1 + (T)^{-\frac{1}{2}} \Theta_2 \right\} \eta(s)(1 + o_p(1)), \tag{A.29}$$

where $\eta(s)$ is a $p + q + 1$ random vector with zero mean and identity variance matrix, and Θ_1 and Θ_2 are two $(p + q + 1) \times (p + q + 1)$ matrices, satisfying

$$\Theta_1 \Theta_1' = \tau(1 - \tau) \frac{[\Omega^*(x, s)]^{-1} \Omega(x, s) [\Omega^*(x, s)]^{-1}}{f_S(s) f_X(x, s)} \int K^2(u) du \int L^2(z) dz,$$

$$\Theta_2 \Theta_2' = \tau(1 - \tau) \frac{[\Omega^*(x, s)]^{-1} \Omega(x, s, s) [\Omega^*(x, s)]^{-1}}{f_X^2(x, s)} q(x, x; s).$$

This completes the proof of the theorem.

Chapter 6

Conclusion and Outlook

Spatio-temporal data analysis is important in reality such as in energy market analysis. Most of the literature assume linearity in model structure or normality for variables. However, the majority of real energy data may be non-Gaussian with an obvious peak and fat tails ([Hung et al., 2008](#)), and the relationship between different markets may be more complex than linear ([Aloui et al., 2014](#)). Moreover, the spatio-temporal data in energy market reality are often non-stationary over irregular locations in space. Motivated by these drawbacks in the current literature, this thesis sheds new light on both theoretical investigation into the spatio-temporal model with varying coefficient structure and empirical application to the complex dynamic nature of spatio-temporal data in energy market.

As a preliminary analysis, in Chapter 2, we empirically investigate the dynamic integrations or linkages of the USA's natural gas markets at the state level by proposing a spatio-temporal network modelling from the perspective of quantile analysis. This chapter empirically finds that spatial neighbouring effects significantly exist among the natural gas markets, not only in the eastern and middle states but also in some western and south-west states, and so do the dynamic linkages to the national crude oil market for the natural gas markets in the southern and eastern states, which are heterogeneous at different quantile levels. As this chapter considers parametric model, therefore there may be some shortcomings in uncovering the underlying relationship for some states. This spurs us to further develop the semiparametric time series and spatio-temporal mean and quantile models in the following chapters that will help to verify whether the relationships in the energy market analysis are nonlinear or not.

In Chapter 3, we are mainly concerned with the nonlinear modelling of econometric time series data from the view point of quantile regression. Specifically, we consider quantile regression for a (strictly) stationary time series that is near epoch dependent (NED). It is an extension of the work of [Welsh \(1996\)](#) and [Yu and Jones \(1998\)](#) under *i.i.d.* samples to econometric time series. Our asymptotic normality result generalizes [Lu et al. \(1998\)](#), [Honda \(2000\)](#) and [Hallin et al. \(2009a\)](#) who obtained Bahadur representations with different conditions under alpha-mixing dependence to a more widely applicable family of data generating processes of near epoch dependence. This chapter lays a foundation for establishing asymptotic properties of nonlinear quantile modelling for time series and spatio-temporal data analysis.

In Chapter 4, we develop a semiparametric family of dynamic functional-coefficient autoregressive spatio-temporal models to model the nonlinear structure in spatio-temporal data. We model data that are non-stationary over irregular locations in space, which appears generally in the real spatio-temporal data analysis. Therefore, our proposed model not only well characterises the dynamic regime-switching nature but also overcomes the location-wide non-stationarity existent in spatio-temporal data. It is a useful extension and combination of spatial autoregressive models ([Ord, 1975](#); [Gao et al., 2006](#); [Kelejian and Prucha, 2010](#); [Su and Jin, 2010](#)) and varying coefficient models ([Zhang and Fan, 1999](#); [Wang and Xia, 2009](#); [Zhang et al., 2009](#)) as well as [Al-Sulami et al. \(2017a, 2019\)](#). To model the dynamic spatial neighbouring temporal-lagged effects with the irregular locations, we consider using spatial weight matrix pre-specified either by experts or by the prior information of spatial locations, which is popular in spatial econometrics. Moreover, both one-step and two-step estimation methods are proposed to estimate the functional coefficients in DyFAST models. In practice, such spatial weight matrix can be pre-specified by the features of the data in many different ways. Therefore, the idea of weight matrix selection and combination is also discussed for an optimal weight matrix among many candidates.

In Chapter 5, we propose a semiparametric spatio-temporal quantile model, extending Chapter 4 to quantile regression analysis, to further study the dynamic nature of non-normal spatio-temporal data. The quantile estimators are more robust than the mean estimators considered in Chapter 4 for non-normal spatio-temporal data. Moreover, this approach can provide more information about the dynamic spatial effects at different quantile levels. The empirical results clearly verify that the impacts of oil price return on the retail gasoline price returns in the EU countries are more than linear, especially at middle and upper quantile levels.

Both theoretical asymptotic properties and Monte Carlo simulations have been established for our methodologies. Our empirical applications to energy market data sets further illustrate the usefulness of these models. Application of our methods to real energy market analysis can help relevant industry or governments to better understand the energy markets, enabling to reduce the negative impacts from expected or unexpected price changes in the energy markets and enact appropriate energy policies and investment decisions.

It must be acknowledged that this study does have some limitations. The DyFAST model can not only well characterise the dynamic regime-switching nature with spatio-temporal lag neighbouring effects considered for forecasting but also overcome the location-wide non-stationarity existent in spatio-temporal data. However, this model does not consider contemporaneous spatial neighbouring effect of endogenous nature (but not suitable for forecasting) in spatial econometrics, which needs more investigation. Moreover, there may also exist spatial interaction in errors, namely spatial error autocorrelation if the spatial neighbouring effect is not fully or correctly taken account of.

There are possible directions for future research along this work. A further extension of the theoretical methodologies in Chapters 4 and 5 is to establish a penalised method to identify the zero functional and constant coefficients, which could accurately estimate the structural coefficients of the model covariates. Finally, on the empirical side, our model could be used to analyse the real estate markets and the spatial dependency of environmental data such as carbon emission or PM2.5, among others. These are all left for future investigations.

References

- Aït-Sahalia, Y., Lo, A. W., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance* 53 (2), 499–547.
- Al-Sulami, D., Jiang, Z., Lu, Z., Zhu, J., 2017a. Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data. *Econometrics and Statistics* 2, 22–35.
- Al-Sulami, D., Jiang, Z., Lu, Z., Zhu, J., 2017b. Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data. *Econometrics and Stats* 2, 22–35.
- Al-Sulami, D., Jiang, Z., Lu, Z., Zhu, J., 2019. On a semiparametric data-driven nonlinear model with penalized spatio-temporal lag interactions. *Journal of Time Series Analysis* 40 (3), 327–342.
- Aloui, R., Aïssa, M. S. B., Hammoudeh, S., Nguyen, D. K., 2014. Dependence and extreme dependence of crude oil and natural gas prices with applications to risk management. *Energy Economics* 42, 332–342.
- Alquist, R., Gervais, O., 2013. The role of financial speculation in driving the price of crude oil. *The Energy Journal* 34 (3), 35.
- Andrews, D. W., 1984. Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 930–934.
- Andrews, D. W., 1995. Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11 (03), 560–586.
- Anselin, L., 1993. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. Regional Research Institute, West Virginia University Morgantown, WV.
- Anselin, L., 2013. Spatial econometrics: methods and models. Vol. 4. Springer Science & Business Media.

- Arbia, G., 2014. A primer for spatial econometrics: with applications in R. Springer.
- Atil, A., Lahiani, A., Nguyen, D. K., 2014. Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices. *Energy Policy* 65, 567–573.
- Auffhammer, M., Carson, R. T., 2008. Forecasting the path of China's CO₂ emissions using province-level information. *Journal of Environmental Economics and Management* 55 (3), 229–247.
- Bachmeier, L. J., Griffin, J. M., 2006. Testing for market integration crude oil, coal, and natural gas. *The Energy Journal* 27 (2), 55–71.
- Balaguer, J., Cantavella, M., 2016. Estimating the environmental Kuznets curve for Spain by considering fuel oil prices (1874–2011). *Ecological Indicators* 60, 853–859.
- Banfi, S., Filippini, M., Hunt, L. C., 2005. Fuel tourism in border regions: The case of Switzerland. *Energy Economics* 27 (5), 689–707.
- Benz, E., Trück, S., 2009. Modeling the price dynamics of CO₂ emission allowances. *Energy Economics* 31 (1), 4–15.
- Bera, A. K., Galvao, A. F., Montes-Rojas, G. V., Park, S. Y., 2016. Asymmetric laplace regression: Maximum likelihood, maximum entropy and quantile regression. *Journal of Econometric Methods* 5 (1), 79–101.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.
- Bettendorf, L., van der Geest, S. A., Kuper, G. H., 2009. Do daily retail gasoline prices adjust asymmetrically? *Journal of Applied Statistics* 36 (4), 385–397.
- Bickel, P. J., 1975. One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70 (350), 428–434.
- Blair, B. F., Campbell, R. C., Mixon, P. A., 2017. Price pass-through in US gasoline markets. *Energy Economics* 65, 42–49.
- Bollerslev, T., Chou, R. Y., Kroner, K. F., 1992. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics* 52 (1-2), 5–59.

- Bosch, R. J., Ye, Y., Woodworth, G. G., 1995. A convergent algorithm for quantile regression with smoothing splines. *Computational statistics & data analysis* 19 (6), 613–630.
- Brown, S. P., Yücel, M. K., 2008. What drives natural gas prices? *The Energy Journal* 29 (2), 45–60.
- Cai, Z., 2002. Regression quantiles for time series. *Econometric theory* 18 (01), 169–192.
- Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for non-linear time series. *Journal of the American Statistical Association* 95 (451), 941–956.
- Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics* 167 (2), 413–425.
- Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association* 103 (484), 1595–1608.
- Cárdenas, J., Gutiérrez, L. H., Otero, J., 2017. Investigating diesel market integration in France: Evidence from micro data. *Energy Economics* 63, 314–321.
- Chang, C.-L., McAleer, M., Tansuchat, R., 2010. Analyzing and forecasting volatility spillovers, asymmetries and hedging in major oil markets. *Energy Economics* 32 (6), 1445–1455.
- Chen, R., Tsay, R. S., 1993. Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88 (421), 298–308.
- Cheng, M.-Y., Zhang, W., Chen, L.-H., 2009. Statistical estimation in generalized multiparameter likelihood models. *Journal of the American Statistical Association* 104 (487), 1179–1191.
- Chevallier, J., 2011. A model of carbon price interactions with macroeconomic and energy dynamics. *Energy Economics* 33 (6), 1295–1312.
- Chuang, C.-C., Kuan, C.-M., Lin, H.-Y., 2009. Causality in quantiles and dynamic stock return–volume relations. *Journal of Banking & Finance* 33 (7), 1351–1360.

- Clemenzen, G., Gugler, K., 2006. Locational choice and price competition: some empirical results for the austrian retail gasoline market. *Empirical Economics* 31 (2), 291–312.
- Clerides, S., et al., 2010. Retail fuel price response to oil price shocks in eu countries. *Cyprus Economic Policy Review* 4 (1), 25–45.
- Cliff, A., Ord, K., 1972. Testing for spatial autocorrelation among regression residuals. *Geographical analysis* 4 (3), 267–284.
- Cong, R.-G., Wei, Y.-M., Jiao, J.-L., Fan, Y., 2008. Relationships between oil price shocks and stock market: An empirical analysis from china. *Energy Policy* 36 (9), 3544–3553.
- Cressie, N., Wikle, C. K., 2015. *Statistics for spatio-temporal data*. John Wiley & Sons.
- Cuddington, J. T., Wang, Z., 2006. Assessing the degree of spot market integration for US natural gas: evidence from daily price data. *Journal of Regulatory Economics* 29 (2), 195–210.
- Davidson, J., 1994. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.
- De Vany, A., Walls, W. D., 1993. Pipeline access and market integration in the natural gas industry: Evidence from cointegration tests. *The Energy Journal* 14 (4), 1–19.
- Deo, C. M., 1973. A note on empirical processes of strong-mixing sequences. *The Annals of Probability*, 870–875.
- Devroye, L., Györfi, L., 1985. *Local polynomial modelling and its applications: monographs on statistics and applied probability* 66. New York: John Wiley & Sons.
- Ding, H., Kim, H.-G., Park, S. Y., 2016. Crude oil and stock markets: Causal relationships in tails? *Energy Economics* 59, 58–69.
- Efron, B., 1991. Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 93–125.
- Fan, J., Gijbels, I., 1996. *Local polynomial modelling and its applications: monographs on statistics and applied probability*. Vol. 66. CRC Press.

- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
- Fan, J., Zhang, W., 2000. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics* 27 (4), 715–731.
- Fan, Y., Zhang, Y.-J., Tsai, H.-T., Wei, Y.-M., 2008. Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics* 30 (6), 3156–3171.
- Fingleton, B., 2008. A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. *Empirical Economics* 34 (1), 35–57.
- Gallant, A. R., 2009. *Nonlinear statistical models*. Vol. 310. John Wiley & Sons.
- Gao, J., Lu, Z., Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. *The Annals of Statistics* 34 (3), 1395–1435.
- Ghoddusi, H., 2016. Integration of physical and futures prices in the US natural gas market. *Energy Economics* 56, 229–238.
- Gorodetskii, V., 1978. On the strong mixing property for linear sequences. *Theory of Probability & Its Applications* 22 (2), 411–413.
- Granger, C. W. J., White, H., Kamstra, M., 1989. Interval forecasting: an analysis based upon ARCH-quantile estimators. *Journal of Econometrics* 40 (1), 87–96.
- Gregory, A. W., Hansen, B. E., 1996. Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70 (1), 99–126.
- Hall, P., Heyde, C., 1980. *Martingale limit theory and its applications*, acad. Press, New York.
- Hall, P., Heyde, C. C., 2014. *Martingale limit theory and its application*. Academic press.
- Hall, P., Wolff, R. C., Yao, Q., 1999. Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* 94 (445), 154–163.
- Hallin, M., Lu, Z., Tran, L. T., 2004. Local linear spatial regression. *The Annals of Statistics* 32 (6), 2469–2500.

- Hallin, M., Lu, Z., Yu, K., 2009a. Local linear spatial quantile regression. *Bernoulli* 15 (3), 659–686.
- Hallin, M., Lu, Z., Yu, K., 2009b. Local linear spatial quantile regression. *Bernoulli* 15 (3), 659–686.
- Hamilton, J. D., 1983. Oil and the macroeconomy since World War II. *Journal of political economy* 91 (2), 228–248.
- Hamilton, J. D., 2009. Causes and consequences of the oil shock of 2007-08. Tech. rep., National Bureau of Economic Research.
- Härdle, W., Lütkepohl, H., Chen, R., 1997. A review of nonparametric time series analysis. *International Statistical Review* 65 (1), 49–72.
- Hartley, P. R., Medlock III, K. B., Rosthal, J. E., 2008. The relationship of natural gas to oil prices. *The Energy Journal* 29 (3).
- He, X., Hu, F., 2002. Markov chain marginal bootstrap. *Journal of the American Statistical Association* 97 (459), 783–795.
- He, X., Shao, Q.-M., et al., 1996. A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics* 24 (6), 2608–2630.
- Holly, S., Pesaran, M. H., Yamagata, T., 2010. A spatio-temporal model of house prices in the USA. *Journal of Econometrics* 158 (1), 160–173.
- Holmes, M. J., Otero, J., Panagiotidis, T., 2013. On the dynamics of gasoline market integration in the united states: Evidence from a pair-wise approach. *Energy Economics* 36, 503–510.
- Honda, T., 2000. Nonparametric estimation of a conditional quantile for α -mixing processes. *Annals of the Institute of Statistical Mathematics* 52 (3), 459–470.
- Honda, T., 2004. Quantile regression in varying coefficient models. *Journal of statistical planning and inference* 121 (1), 113–125.
- Hooten, M. B., Wikle, C. K., 2008. A hierarchical bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics* 15 (1), 59–70.
- Hu, T., Xia, Y., 2012. Adaptive semi-varying coefficient model selection. *Statistica Sinica* 22, 575–599.

- Hung, J.-C., Lee, M.-C., Liu, H.-C., 2008. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. *Energy Economics* 30 (3), 1173–1191.
- Hurvich, C. M., Simonoff, J. S., Tsai, C.-L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60 (2), 271–293.
- Ibragimov, I. A., 1962. Some limit theorems for stationary processes. *Theory of Probability & Its Applications* 7 (4), 349–382.
- Jenish, N., 2012. Nonparametric spatial regression under near-epoch dependence. *Journal of Econometrics* 167 (1), 224–239.
- Jones, C. M., Kaul, G., 1996. Oil and the stock markets. *The Journal of Finance* 51 (2), 463–491.
- Jorion, P., 1997. *Value at risk*. McGraw-Hill, New York.
- Kelejian, H. H., Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157 (1), 53–67.
- Kim, M.-O., 2007. Quantile regression with varying coefficients. *The Annals of Statistics*, 92–108.
- Koenker, R., 2005. *Quantile regression*. No. 38. Cambridge university press.
- Koenker, R., Bassett Jr, G., 1978. Regression quantiles. *Econometrica: Journal of the Econometric Society* 46 (1), 33–50.
- Koenker, R., Bassett Jr, G., 1982. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, 43–61.
- Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. *Biometrika* 81 (4), 673–680.
- Koenker, R., Portnoy, S., Ng, P. T., Zeileis, A., Grosjean, P., Ripley, B. D., Dec. 2018. Package 'quantreg'.
URL <https://cran.r-project.org/web/packages/quantreg/quantreg.pdf>
- Koenker, R., Zhao, Q., 1996. Conditional quantile estimation and inference for arch models. *Econometric Theory* 12 (05), 793–813.

- Kuester, K., Mittnik, S., Paolella, M. S., 2006. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* 4 (1), 53–89.
- Kuper, G. H., Mulder, M., 2016. Cross-border constraints, institutional changes and integration of the Dutch–German gas market. *Energy Economics* 53, 182–192.
- Lahiani, A., Miloudi, A., Benkraiem, R., Shahbaz, M., 2017. Another look on the relationships between oil prices and energy prices. *Energy Policy* 102, 318–331.
- Lee, B.-J., Yang, C. W., Huang, B.-N., 2012. Oil price movements and stock markets revisited: A case of sector stock price indexes in the G-7 countries. *Energy Economics* 34 (5), 1284–1300.
- Lee, S., 2003. Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric theory* 19 (01), 1–31.
- Lee, Y.-H., Chiou, J.-S., 2011. Oil sensitivity and its asymmetric impact on the stock market. *Energy* 36 (1), 168–174.
- LeSage, J. P., Pace, R. K., 2009. *Introduction to Spatial Econometrics* (Statistics, textbooks and monographs). CRC Press.
- Li, D., Lu, Z., Linton, O., 2012a. Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory* 28 (5), 935–958.
- Li, J., Zhang, W., 2011. A semiparametric threshold model for censored longitudinal data analysis. *Journal of the American Statistical Association* 106 (494), 685–696.
- Li, Q., Racine, J., 2004. Cross-validated local linear nonparametric regression. *Statistica Sinica* 14, 485–512.
- Li, S.-F., Zhu, H.-M., Yu, K., 2012b. Oil prices and stock market in China: A sector analysis using panel cointegration with multiple breaks. *Energy Economics* 34 (6), 1951–1958.
- Loader, C. R., 1999. Bandwidth selection: classical or plug-in? *The Annals of Statistics* 27 (2), 415–438.
- Lu, Z., 2001. Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics* 53 (3), 447–468.

- Lu, Z., Gijbels, I., 2001. Asymptotics for partly linear regression with dependent samples and ARCH errors: consistency with rates. *Science in China Series A: Mathematics* 44 (2), 168–183.
- Lu, Z., Hui, Y., Zhao, Q., 1998. Local linear quantile regression under dependence: Bahadur representation and application. Tech. rep., Discussion paper, Dept. Management Sciences, City Univ. Hong Kong.
- Lu, Z., Linton, O., 2007. Local linear fitting under near epoch dependence. *Econometric Theory* 23 (01), 37–70.
- Lu, Z., Lundervold, A., Tjøstheim, D., Yao, Q., 2007. Exploring spatial nonlinearity using additive approximation. *Bernoulli* 13 (2), 447–472.
- Lu, Z., Steinskog, D. J., Tjøstheim, D., Yao, Q., 2009. Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71 (4), 859–880.
- Lu, Z., Tjøstheim, D., 2014. Nonparametric estimation of probability density functions for irregularly observed spatial data. *Journal of the American Statistical Association* 109 (508), 1546–1564.
- Mansanet-Bataller, M., Pardo, A., Valor, E., 2007. CO2 prices, energy and weather. *The Energy Journal* 28 (3).
- Masry, E., Fan, J., 1997. Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24 (2), 165–179.
- Moran, P. A., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37 (1/2), 17–23.
- Neumann, A., Siliverstovs, B., von Hirschhausen, C., 2006. Convergence of European spot market prices for natural gas? A real-time analysis of market integration using the Kalman Filter. *Applied Economics Letters* 13 (11), 727–732.
- Oberhofer, W., Haupt, H., 2016. Asymptotic theory for nonlinear quantile regression under weak dependence. *Econometric Theory* 32 (3), 686–713.
- Ord, K., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70 (349), 120–126.
- Panagiotidis, T., Rutledge, E., 2007. Oil and gas markets in the UK: Evidence from a cointegrating approach. *Energy economics* 29 (2), 329–347.

- Papapetrou, E., 2001. Oil price shocks, stock market, economic activity and employment in greece. *Energy economics* 23 (5), 511–532.
- Park, H., Mjelde, J. W., Bessler, D. A., 2008. Price interactions and discovery among natural gas spot markets in north america. *Energy Policy* 36 (1), 290–302.
- Pesaran, M. H., may 2007. A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics* 138 (1), 312–355.
- Ramberg, D. J., Parsons, J. E., 2012. The weak tie between natural gas and oil prices. *The Energy Journal*, 13–35.
- Reboredo, J. C., Ugolini, A., 2016. Quantile dependence of oil price movements and stock returns. *Energy economics* 54, 33–49.
- Renou-Maissant, P., 2012. Toward the integration of European natural gas markets: A time-varying approach. *Energy Policy* 51, 779–790.
- Robinson, P. M., 1983. Nonparametric estimators for time series. *Journal of Time Series Analysis* 4 (3), 185–207.
- Ruppert, D., Carroll, R. J., 1980. Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75 (372), 828–838.
- Serletis, A., Rangel-Ruiz, R., 2004. Testing for common features in North American energy markets. *Energy Economics* 26 (3), 401–414.
- Shahbaz, M., Lahiani, A., Abosedra, S., Hammoudeh, S., 2018. The role of globalization in energy consumption: A quantile cointegrating regression approach. *Energy Economics* 71, 161–170.
- Shao, Q.-M., Yu, H., 1996. Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* 24 (4), 2098–2127.
- Sherwood, B., Wang, L., 2016. Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics* 44 (1), 288–317.
- Sim, N., Zhou, H., 2015. Oil prices, US stock return, and the dependence between their quantiles. *Journal of Banking & Finance* 55, 1–8.
- Stone, C. J., 1980. Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* 8 (6), 1348–1360.

- Stone, C. J., 1984. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics* 12 (4), 1285–1297.
- Su, L., Jin, S., 2010. Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics* 157 (1), 18–33.
- Subba Rao, S., 2008. Statistical analysis of a spatio-temporal model with location-dependent parameters and a test for spatial stationarity. *Journal of Time Series Analysis* 29 (4), 673–694.
- Sun, Y., Zhang, W., Tong, H., 2007. Estimation of the covariance matrix of random effects in longitudinal studies. *The Annals of Statistics* 35 (6), 2795–2814.
- Thompson, M., Davison, M., Rasmussen, H., 2009. Natural gas storage valuation and optimization: A real options application. *Naval Research Logistics (NRL)* 56 (3), 226–238.
- Tjøstheim, D., Auestad, B. H., 1994. Nonparametric identification of nonlinear time series: selecting significant lags. *Journal of the American Statistical Association* 89 (428), 1410–1419.
- Tong, H., 1990. *Non-linear time series: a dynamical system approach*. Oxford University Press.
- Villar, J. A., Joutz, F. L., 2006. The relationship between crude oil and natural gas prices. *Energy Information Administration, Office of Oil and Gas*, 1–43.
- Volkonskii, V., Rozanov, Y. A., 1959. Some limit theorems for random functions. i. *Theory of Probability & Its Applications* 4 (2), 178–197.
- Vücel, M. K., Guo, S., 1994. Fuel taxes and cointegration of energy prices. *Contemporary Economic Policy* 12 (3), 33–41.
- Wang, H., Xia, Y., 2009. Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association* 104 (486), 747–757.
- Wang, H. J., Zhu, Z., Zhou, J., 2009. Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, 3841–3866.
- Welsh, A., 1996. Robust estimation of smooth regression and spread functions and their derivatives. *Statistica Sinica*, 347–366.
- Wikle, C. K., Holan, S. H., 2011. Polynomial nonlinear spatio-temporal integro-difference equation models. *Journal of Time Series Analysis* 32 (4), 339–350.

- Wlazlowski, S., Giuliatti, M., Binner, J., Milas, C., 2009. Price dynamics in European petroleum markets. *Energy Economics* 31 (1), 99–108.
- Wolfe, M. H., Rosenman, R., 2014. Bidirectional causality in oil and gas markets. *Energy Economics* 42, 325–331.
- Wu, W., Zhou, Z., 2017. Nonparametric inference for time-varying coefficient quantile regression. *Journal of Business & Economic Statistics* 35 (1), 98–109.
- Xu, B., Lin, B., 2016. A quantile regression analysis of China's provincial CO₂ emissions: Where does the difference lie? *Energy Policy* 98, 328–342.
- Xu, X., 2005. Semiparametric and nonparametric estimation for dynamic quantile regression models. The University of North Carolina at Charlotte.
- Yilmazkuday, D., Yilmazkuday, H., 2016. Understanding gasoline price dispersion. *The Annals of Regional Science* 57 (1), 223–252.
- Yu, H., 2012. The influential factors of China's regional energy intensity and its spatial linkages: 1988–2007. *Energy Policy* 45, 583–593.
- Yu, K., Jones, M., 1998. Local linear quantile regression. *Journal of the American statistical Association* 93 (441), 228–237.
- Yu, K., Lu, Z., 2004. Local linear additive quantile regression. *Scandinavian Journal of Statistics* 31 (3), 333–346.
- Yu, K., Lu, Z., Stander, J., 2003. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)* 52 (3), 331–350.
- Zhang, C., Chen, X., 2011. The impact of global oil price shocks on China's stock returns: Evidence from the ARJI (-ht)-EGARCH model. *Energy* 36 (11), 6627–6633.
- Zhang, C., Chen, X., 2014. The impact of global oil price shocks on China's bulk commodity markets and fundamental industries. *Energy policy* 66, 32–41.
- Zhang, W., Fan, J., 1999. Statistical estimation in varying coefficient models. *The Annals of Statistics* 27 (5), 1491–1518.
- Zhang, W., Fan, J., Sun, Y., 2009. A semiparametric model for cluster data. *The Annals of Statistics* 37 (5A), 2377–2408.

-
- Zhang, W., Lee, S.-Y., Song, X., 2002. Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis* 82 (1), 166–188.
- Zhang, W., Yao, Q., Tong, H., Stenseth, N. C., 2003. Smoothing for spatiotemporal models and its application to modeling muskrat-mink interaction. *Biometrics* 59 (4), 813–821.
- Zhang, X., Yu, J., 2018. Spatial weights matrix selection and model averaging for spatial autoregressive models. *Journal of Econometrics* 203 (1), 1–18.
- Zhou, Z., Wu, W. B., et al., 2009. Local linear quantile estimation for nonstationary time series. *The Annals of Statistics* 37 (5B), 2696–2729.