The Hong Kong Polytechnic University

Department of Applied Mathematics

Spatial Regression Models with Spatially Correlated Errors

by

Fan ZHANG

A thesis submitted in partial fulfillment of

the requirements for the Degree of Doctor of Philosophy

December 2010

CERTIFICATION OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Fan ZHANG (Name of student)

Abstract

This thesis studies the spatial regression models with lattice data, with emphasis on models with spatially correlated errors. For the large scale variation of the data, the nonparametric, additive nonparametric and semi-parametric structure are adopted; while for the small scale variation, the errors are assumed to satisfy the torus, separable or unilateral SGAR model. Following Martins-Filho & Yao (2009), we propose to estimate the large scale variation with a two-step fitting procedure, which firstly forms a new process with the same conditional mean as the original one and $i.i.d.$ errors, and secondly applies the estimation to the new process. Such approach takes both nonstationary mean/trend effects and spatial dependencies into account, hence overmatches the traditional estimations.

Asymptotic properties of both first- and second-step estimators are investigated. For the first-step estimators of the unknown regression function, the convergence rate with all three types of errors is considered, and when errors satisfy the separable or unilateral SGAR model, the asymptotic normality is established. For the second-step estimators of the unknown regression function, the asymptotic normality with three types of error structures is established. In the semi-parametric model, we also establish the asymptotic normality of the first- and second-step estimators of the linear parameters.

For all the models, simulations are conducted to assess the performance of our fitting. Under the condition that spatially correlated errors exist, the results show that our estimation works better than the traditional methods. The improvement of our estimation is significant when the volatility of the errors is large. As an illustration of our approach, a case study of the housing price in Hong Kong is given. It is shown that our approach improves the estimation, especially when some key factor is absent in the modelling.

Acknowledgements

I owe my supervisors, Dr. Wai-Cheung Ip, Dr. Heung Wong and Prof. Yuan Li, a debt of gratitude for their enthusiastic guidance and support throughout my study of the Ph.D degree. From the countless insightful discussions with them, I benefit greatly in both aspects of academic research and thesis writing. Their warm encouragements as well as constructive criticisms are precious for me.

I am grateful to Dr. Xianhua Luo and Dr. Quanxi Shao for their valuable suggestions and helps to my work. I am also thankful to my fellow classmate Xingfa Zhang, who always encourages me when I felt frustrated during the study.

I shall express my sincere thanks to all the staff in the Department of Applied Mathematics at the Hong Kong Polytechnic University for their kind assistance, and I greatly appreciate the financial support from The Hong Kong Polytechnic University, whose research studentship supported me in the past three years.

At last, I will thank my parents and my wife, the ones love me and I love. Without the cares, fun, encouragement and understanding that they give me for years, I will not be able to finish my study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Spatial Data and Statistical Treatments

Spatial data arise in various areas, including geology, ecology, environmental science, mining engineering, epidemiology, image analysis, oceanography and econometrics. As spatial data fill in almost every corner of our life, appropriate quantitative analyses of the data always lead to meaningful results. Statistical treatment with such data first appeared in the early 20th century, and have been broadly discussed in recent decades, with great deal of literature found on this topic. For a systematic reading, we refer to Cressie (1993) and Schabenberger & Gotway (2005).

Typically, spatial data can be decomposed into two components, the large-scale and small-scale variation, and modeled generally as

$$Y_{\mathbf{s}} = \mu(\mathbf{X}_{\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{1.1}$$

where $\mathbf{s}$ is the location index on a plain, $Y_{\mathbf{s}}$ represents the spatial observation and is a

scalar, and $\mathbf{X_s}$ is the explanatory random vector, which can be endogenetic or exogenous. $\mu(\mathbf{X_s})$ denotes the large-scale variation, figuring the nonstationary mean(trend) effect. The residual $\epsilon_\mathbf{s}$ denotes the small-scale variation, figuring the stationary spatial dependency of the data. According to past study, the spatial dependency contained in the residuals should not be ignored in some application problems, see Watson (1972), Russo & Bresler (1981) and Burrough (1983). Depending on the spatial distribution of the data point, irregularly spaced or regularly spaced, $\epsilon_\mathbf{s}$ can be modeled with different types of spatial models. We will introduce these in details in Section 1.2.

In the early development of spatial statistics, the large-scale variation $\mu(\mathbf{X_s})$ is often assumed to have some parametric structure, i.e., the linear structure with respect to $\mathbf{X_s}$. Many researchers contributed to such study, for instance, Martin (1982), Cressie & Chan (1989), Basu & Reinsel (1994), among others. The linearity assumption makes the analysis easy to handle and explainable, but it is usually too idealized for application research. It may be more reasonable to model $\mu(\mathbf{X_s})$ with some nonparametric or semi-parametric structures, which have been shown useful to capture the nonlinear trend of the regression function in many applications. Nevertheless, although the nonparametric and semi-parametric models have been greatly studied in the last 20 years, most of the researches focused on the cases with *i.i.d.* and temporally correlated residuals. Only a few scholars have considered the ones with spatially correlated errors, e.g., Liu(2001) and Francisco-Fernandez & Opsomer (2005). Hence, much more problems on this topic deserve further study.

In our thesis, we will focus on the nonlinear estimation of model (1.1) with spatially correlated residuals. Besides the general structure defined in (1.1), the large-scale varia-

tion $\mu(\mathbf{X_s})$ is assumed to take another two special structures, and hence model (1.1) can be separated into three sub-models as below:

$$Y_{\mathbf{s}} = m(\mathbf{X_s}) + \epsilon_{\mathbf{s}}, \tag{1.2}$$

$$Y_{\mathbf{s}} = \mu + \sum_{i=1}^{d} m_i(\mathbf{X}_{i\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{1.3}$$

$$Y_{\mathbf{s}} = \mathbf{X}_{1\mathbf{s}}^T \boldsymbol{\lambda} + m(\mathbf{X}_{2\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{1.4}$$

where μ is some constant, $m(.)$ and $m_i(.)$ are some unknown functions, $\mathbf{X_s}$ and $\mathbf{X}_{i\mathbf{s}}$ are random explanatory vectors, $i = 1, \cdots, d$, and $\boldsymbol{\lambda}$ is a parameter vector. Models (1.2)-(1.4) are well known as the nonparametric, additive and semi-parametric models respectively. The residuals $\epsilon_{\mathbf{s}}$ in each model is assumed to satisfy several types of simultaneous Gaussian auto-regressive models. A two-step estimation introduced in Martins-Filho & Yao (2009) will be applied, and the local linear fitting technique will be used in all models.

This chapter will give a brief overview of the traditional spatial models and local polynomial regression estimation. Two major groups of spatial models, the geostatistic models and lattice models, will be addressed in Section 1.2, with emphasis on the lattice models. In Section 1.3, the local polynomial regression fitting will be introduced. In the last section of this chapter, the organization of the whole thesis will be given.

1.2 A Brief Introduction to Spatial Models

1.2.1 Geostatistics

Geostatistics models is a group of spatial models, which were first introduced in Matheron (1962, 1963a & 1963b). Unlike the case in lattice models that we will review in the next sub-section, the spatial index **s** in a geostatistics model can vary continuously over the support. Such property makes geostatistics models suitable to deal with both irregularly spaced and regularly spaced data.

Before we review the estimation and prediction methods in geostatistics, some backgrounds should be considered. We first introduce two definitions, which are crucial in geostatistics. Let **s** and **t** be spatial indexes, and $\{\epsilon_{\mathbf{s}} : \mathbf{s} \in R^2\}$ be a spatial process. Suppose

$$var\left(\epsilon_{\mathbf{s}+\mathbf{t}} - \epsilon_{\mathbf{s}}\right) = 2\gamma\left(\mathbf{t}\right), \tag{1.5}$$

where $2\gamma(.)$ is a function only of the increment **t**, then $2\gamma(.)$ is called a variogram. Similarly, if we have

$$cov\left(\epsilon_{\mathbf{s}+\mathbf{t}}, \epsilon_{\mathbf{s}}\right) = C\left(\mathbf{t}\right), \tag{1.6}$$

then $C(\mathbf{t})$ is called a covariogram.

The variogram has a vital property called conditional negative-definiteness, which can be shown mathematically as

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\, a_j\, 2\gamma\left(\mathbf{s}_i - \mathbf{s}_j\right) \leq 0, \tag{1.7}$$

for any finite number of spatial locations $\{\mathbf{s}_i : i = 1, \cdots, n\}$ and real number $\{a_i : i =$

$1, \cdots, n\}$ with $\sum_{i=1}^{n} a_i = 0$. The covariogram also has an important property called positive-definiteness, which can be stated as

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \, a_j \, C(\mathbf{s}_i - \mathbf{s}_j) \geq 0, \tag{1.8}$$

where a_i's are real numbers.

Related to variogram and covariogram, two different types of stationarities can be defined. A process $\{\epsilon_{\mathbf{s}}\}$ is called intrinsic stationary if $\epsilon_{\mathbf{s}}$ has constant mean and satisfies (1.5), or second-order stationary if $\epsilon_{\mathbf{s}}$ has constant mean, finite second moment and satisfies (1.6). If $\{\epsilon_{\mathbf{s}}\}$ is second-order stationary, by (1.5) and (1.6) we have

$$2\gamma(\mathbf{t}) = 2\Big(C(\mathbf{0}) - C(\mathbf{t})\Big), \tag{1.9}$$

which implies that $\{\epsilon_{\mathbf{s}}\}$ is also intrinsic stationary. Therefore, there is a relationship between the variogram and the covariogram if the second-order stationarity is given, and on the other hand, the variogram can be defined when the covariogram is not. These make the analysis of variogram preferable to that of covariogram.

Now, we will give a brief review of the estimation and prediction procedure in geostatistics. We only focus on the variogram in the following, while the analyses with covariogram are similar and hence omitted here.

The main idea of the estimation of the variogram is to fit the spatial data to some parametric variogram model $\gamma(\mathbf{t}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector. For example, the simplest variogram model may be the isotropic linear model that may be given by

$$\gamma(\mathbf{t}, \alpha, \beta) = \alpha + \beta \, ||\mathbf{t}||, \quad \text{if } \mathbf{t} \neq \mathbf{0} \text{ and zero elsewhere.} \tag{1.10}$$

So the problem becomes one to estimating the parameters in the semi-variogram model, say α and β in (1.10), by some general method such as the maximum likelihood estimation, the least squared estimation, etc. Note that a valid variogram model must satisfy the conditional negative-definiteness, which will constraint the models to be chosen. For more appropriate variogram models, see Section 2.3.1 and Section 2.5 in Cressie (1993), among other literature. We should also note two points for model (1.10). First, this model is isotropic, because the variogram depends on the distance $\mathbf{t}$ only, but not on the direction. Second, the variogram function is discontinuous at the origin, since the so-called nugget effect α exists. See more details about the nugget effect in Section 2.3.1 of Cressie (1993).

The prediction problem in geostatistics is usually called kriging. We will give a review of ordinary kriging below, which is one of the most commonly used kriging procedure. Assume that the spatial data is modeled as $Y_{\mathbf{s}} = \mu + \epsilon_{\mathbf{s}}$, where μ is some constant and $\epsilon_{\mathbf{s}}$ is intrinsic stationary with variogram $2\gamma(.)$. Then the variogram of $Y_{\mathbf{s}}$ can be given by $2\gamma(\mathbf{t}) = var(Y_{\mathbf{s}+\mathbf{t}} - Y_{\mathbf{s}})$. Given that n observations are available, and we denote them as $Y_{\mathbf{s}_1}, \cdots, Y_{\mathbf{s}_n}$, respectively. We aim to give a predictor of $Y_{\mathbf{s}_0}$ by a linear combination of $\{Y_{\mathbf{s}_1}, \cdots, Y_{\mathbf{s}_n}\}$, where $\mathbf{s}_0$ is some fixed location and n is the data size. The predictor can be given by $\tilde{Y}_{\mathbf{s}_0} = \sum_{i=1}^{n} c_i Y_{\mathbf{s}_i}$ with $\sum_{i=1}^{n} c_i = 1$. Hence, we may minimize

$$E\Big(Y_{\mathbf{s}_0} - \sum_{i=1}^{n} c_i Y_{\mathbf{s}_i}\Big)^2 - 2\lambda\Big(\sum_{i=1}^{n} c_i - 1\Big),$$

over $c_1, \cdots, c_n$ and λ, the Lagrange multiplier. This turns out equivalently by minimizing

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2\sum_{i=1}^{n} c_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - 2\lambda\Big(\sum_{i=1}^{n} c_i - 1\Big),$$

where $\gamma(.)$ can be given by the estimator of the variogram function.

1.2.2 Spatial Gaussian Models on a Lattice

A lattice $\Omega \subseteq Z^2$ is a rectangle grided region in which the nodes are equally spaced. Unlike the geostatistics models, the spatial models concerning $\{\epsilon_{\mathbf{s}} : \mathbf{s} \in \Omega\}$ has district spatial indices. Such models may be considered as a generalization of the time series models. However, some substantial gaps exist. Time series models are some so-called unilateral or unidirectional models, as the random variable at any time point depends only on its former values. Due to this property, the joint probability of the time series can be expressed as a production of the conditional probabilities of the variables given their neighbors. For example, consider a time series $\{\delta_i : i = 1, ..., n\}$ with $N(\delta_i)$ being the neighborhood set of δ_i, $i = 1, \cdots, n$. Then the joint probability of δ_i, $i = 1, \cdots, n$ can be expressed as

$$Pr\Big(\delta_1, \cdots, \delta_n\Big) = \prod_{i=1}^{n} Pr\,(\,\delta_i \,|\, N(\delta_i)\,). \tag{1.11}$$

Such relationship, however, is not naturally satisfied by the spatial models on a lattice, which are multilateral and the random variables depend on neighbors from all four directions. This departure leads to two different types of spatial models. We will review them later in this section.

The first model is called the **simultaneous Gaussian autoregressive model** (SGAR), which is introduced in the fundamental paper of Whittle (1954). Assume that there are n nodes on the lattice Ω, and they are denoted as $\mathbf{s}_1, \cdots, \mathbf{s}_n$ respectively. A simultaneous

autoregressive model with support on Ω can then be expressed as

$$Y_{\mathbf{s}_i} = \mu_i + \sum_{j=1}^{n} b_{ij}\,(Y_{\mathbf{s}_j} - \mu_j) + \tau_{\mathbf{s}_i}, \tag{1.12}$$

where μ_i is some constant and $\tau_{\mathbf{s}_i}$' s are the residuals for $i = 1, \cdots, n$. Moreover, $b_{ii} \equiv 0$ and $b_{ij} = 0$ unless $Y_{\mathbf{s}_j}$ is a neighbor of $Y_{\mathbf{s}_i}$.

If we define $\mathbf{Y} = (Y_{\mathbf{s}_1}, \cdots, Y_{\mathbf{s}_n})^T$, $\mathbf{u} = (\mu_1, \cdots, \mu_n)^T$, $\Upsilon = (\tau_{\mathbf{s}_1}, \cdots, \tau_{\mathbf{s}_n})^T$, $\mathbf{I}$ the unit matrix and $\mathbf{B}$ a matrix with the ij-element being b_{ij}, model (1.12) can be written in an equivalent matrix form:

$$(\mathbf{I} - \mathbf{B})(\mathbf{Y} - \mathbf{u}) = \Upsilon. \tag{1.13}$$

Assume that $\Upsilon \sim \text{Gau}(\mathbf{0}, \mathbf{\Lambda})$, $\mathbf{\Lambda}$ is some diagonal matrix. Then by (1.13), the joint distribution of $\mathbf{Y}$ is given by

$$\mathbf{Y} \sim \text{Gau}\left(\mathbf{u},\, (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Lambda}\,(\mathbf{I} - \mathbf{B}^T)^{-1}\right), \tag{1.14}$$

provided that $(\mathbf{I} - \mathbf{B})^{-1}$ exists. It should be emphasized that $\text{cov}(\Upsilon, \mathbf{Y}) = \mathbf{\Lambda}(\mathbf{I} - \mathbf{B}^T)^{-1}$, which is not diagonal. That means, unlike the time series models, the residuals in the SGAR models are correlated with the autoregressive variables. This leads to the fact that $Y_{\mathbf{s}_i}$ in (1.12) does not only depends on its neighbors, but also depends on other $Y_{\mathbf{s}_j}$ whose coefficient b_{ij} are equal to zero. Hence, it implies that the relationship (1.11) does not hold for SGAR models.

The model introduced above is the original SGAR model. Note that $Y_{\mathbf{s}_i}$ defined in (1.12) has constant mean. If the mean of $Y_{\mathbf{s}_i}$ depends on some large-scale factors, such as experimental treatments, spatial trend or exogenous regressors, model (1.12) may be

modified to

$$Y_{\mathbf{s}_i} = \mu\left(\mathbf{X}_{\mathbf{s}_i}\right) + \sum_{j=1}^{n} b_{ij}\left(Y_{\mathbf{s}_j} - \mu(\mathbf{X}_{\mathbf{s}_j})\right) + \tau_{\mathbf{s}_i}\,, \tag{1.15}$$

where $\mathbf{X}_{\mathbf{s}_i}$ is a vector of the large-scale factors, $i = 1, \cdots, n$, and $\mu(.)$ is some function. As a special case, $\mu\left(\mathbf{X}_{\mathbf{s}_i}\right)$ can be linear with respect to $\mathbf{X}_{\mathbf{s}_i}$, i.e., $\mu\left(\mathbf{X}_{\mathbf{s}_i}\right) = \mathbf{X}_{\mathbf{s}_i}^T\boldsymbol{\beta}$ with $\boldsymbol{\beta}$ being some parameter vector. Such case was considered in many literature, for example, Martin (1982) and Basu & Reinsel (1994).

When $Y_{\mathbf{s}_i}$ has functional mean, (1.13) can also be rewritten as

$$(\mathbf{I} - \mathbf{B})(\mathbf{Y} - \mathbf{U}) = \Upsilon, \tag{1.16}$$

where $\mathbf{U} = \left(\mu\left(\mathbf{X}_{\mathbf{s}_1}\right), \cdots, \mu\left(\mathbf{X}_{\mathbf{s}_n}\right)\right)^T$. If we further define a vector $\mathbf{E}$ as $\mathbf{E} = \mathbf{Y} - \mathbf{U}$, then the model becomes

$$\mathbf{Y} = \mathbf{U} + \mathbf{E}, \qquad (\mathbf{I} - \mathbf{B})\mathbf{E} = \Upsilon. \tag{1.17}$$

Obviously, when $\mu(.)$ is some unknown function, (1.17) is the model that our thesis will focus on.

The second lattice model we will review is usually called the **conditional Gaussian autoregressive model** (CGAR), which is introduced in Besag (1974). Under the assumption of 'pairwise-only dependence', the model can be given as

$$Y_{\mathbf{s}_i} = \mu_i + \sum_{j=1}^{n} c_{ij}\left(Y_{\mathbf{s}_j} - \mu_j\right) + \nu_{\mathbf{s}_i}, \tag{1.18}$$

where $c_{ii} \equiv 0$, $c_{ij} = 0$ unless there is pairwise dependence between $Y_{\mathbf{s}_i}$ and $Y_{\mathbf{s}_j}$, and $\nu_{\mathbf{s}_i}$ is some residual whose conditional mean given all $Y_{\mathbf{s}_j}$, $j \neq i$, is zero and conditional

variance is denoted by r_i^2. We can also rewrite (1.18) in the matrix form

$$(\mathbf{I} - \mathbf{C})(\mathbf{Y} - \mathbf{u}) = \mathbf{V}, \tag{1.19}$$

where $\mathbf{V} = (v_{\mathbf{s}_1}, \cdots, v_{\mathbf{s}_n})^T$, $\mathbf{C}$ is a matrix with the ij-element being c_{ij}, and $\mathbf{I}$, $\mathbf{Y}$ and $\mathbf{u}$ are the same as defined in (1.13). Then, by the factorization theorem of Besag (1974), the joint distribution of $\mathbf{Y}$ can be shown to be

$$\mathbf{Y} \sim \text{Gau}\left(\mathbf{u}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{R}\right), \tag{1.20}$$

where $\mathbf{R} = \text{diag}\,(r_1^2, \cdots, r_n^2)$, provided that $(\mathbf{I} - \mathbf{C})^{-1}$ exists and $(\mathbf{I} - \mathbf{C})^{-1}\mathbf{R}$ is symmetric and positive-definite.

From (1.19) and (1.20), we have $E(\mathbf{VY}) = \mathbf{R}$, which means the residuals are not correlated with the autoregressive variables. Therefore, the CGAR models satisfy the Markovian property, and the relationship (1.11) holds. Another desirable property of CGAR models is that the conditional mean of $Y_{\mathbf{s}_i}$ can be expressed as a linear combination of other $Y_{\mathbf{s}_j}$, $j \neq i$. So the prediction of CGAR can be readily made, while the problem is more complicated for SGAR models. However, we should also note that $var(\mathbf{V}) = \mathbf{R}(\mathbf{I} - \mathbf{C}^T)$, which is not diagonal, implying that the residuals v_i, $i = 1, \cdots, n$, are correlated. That is not a desirable property, especially for our study.

Comparing the SGAR models and CGAR models, we can see that they are equivalent when their variance matrix are equal. Via (1.14) and (1.20), that is

$$(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Lambda}\,(\mathbf{I} - \mathbf{B}^T)^{-1} = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{R}. \tag{1.21}$$

From this equation, we can see that any SGAR model can be expressed as a CGAR

model, but the inverse may not be true. A typical example is given in Section 6.3.3 of Cressie (1993), showing that a CGAR model with order less than three generally has no equivalent SGAR model.

1.3 The Local Polynomial Fitting

The local polynomial fitting is a non-parametric method that has been studied widely, such as Stone (1977,1980,1982), Cleveland (1979), Fan (1992, 1993), Fan & Gijbels (1992) and Ruppert and Wand (1994). See also Fan & Gijbels (1996) for a systematic reference. Such method is attractive since it has a lot of advantages. For example, it is suitable for both fixed and random designs; the derivatives of the regression function can be estimated; the boundary effects of the estimation do not occur; and the estimators have nice minimax efficiency properties. The general model in interest can be given as

$$Y_i = m(X_i) + \epsilon_i, \tag{1.22}$$

$i = 1, \cdots, n$, $m(.)$ is some unknown regression function, Y_i is the response variable, X_i is the explanatory variable that may be random or fixed, and ϵ_i is the $i.i.d.$ residual. In this section we mainly review the estimation procedure of model (1.22), while the most recent development of local polynomial regression will be reviewed in each first section of Chapter 2 to Chapter 4.

Assume that the p-th derivative of $m(.)$ at the position x_0 exists. For some x close to x_0, $m(x)$ can be approximated by a polynomial of order p with the Taylor expansion, that is

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p.$$

Letting $\beta_j = m^{(j)}(x_0)/j!$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_n)^T$, then the polynomial above can be fitted by the weighted least square method as

$$\underset{\beta \in R^{p+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \Big[Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \Big]^2 K_h(X_i - x_0), \tag{1.23}$$

where h is a bandwidth and $K_h(.) = K(./h)/h$ with $K(.)$ being some kernel function. If we denote

$$\vec{\mathbf{X}} = \begin{bmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{bmatrix}.$$

and $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$, $\mathbf{W} = \operatorname{diag}\Big\{K_h(X_i - x_0)\Big\}$, then (1.23) can be rewritten in the matrix form as

$$\underset{\beta \in R^{p+1}}{\operatorname{argmin}} \left(\mathbf{Y} - \vec{\mathbf{X}}\boldsymbol{\beta}\right)^T \mathbf{W} \left(\mathbf{Y} - \vec{\mathbf{X}}\boldsymbol{\beta}\right).$$

Therefore, the estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(\vec{\mathbf{X}}^T \mathbf{W} \vec{\mathbf{X}}\right)^{-1} \vec{\mathbf{X}}^T \mathbf{W} \mathbf{Y}. \tag{1.24}$$

There are two special topics about the local linear fitting worth mentioned. The first one is the choice of the bandwidth h. Obviously, h controls the size of the local neighborhood, and so that the smoothness of the estimated function. When the sample size n approaches infinity, h is always assumed to approach infinity, too. However, for a fixed sample size, an appropriate bandwidth h should be chosen. If h is too large, the function will be over-smoothly estimated, resulting in a large bias. If h is too small, the function will be under-smoothly estimated, and that will lead to noisy estimates. Therefore, the choice of h is crucial for local polynomial fitting. The second topic is about the choice

of the order of the local polynomial. Typically, fitting the local polynomial with higher order will reduce the bias of the estimators, but on the other hand will increase the variability. Thoroughly theoretical consideration of this topic can be found in Section 3.3 of Fan & Gijbels (1996). In empirical application, however, multiple explanatory variables are usually introduced into model (1.22). Then fitting the function with a high order of polynomial will lead to an explosion of parameters. Generally the local linear fitting, with only the first order of polynomial, is preferred.

1.4 The Organization of the Thesis

Aside from the Introduction Chapter and the Conclusion Chapter, there are four main sections in our thesis. In Chapter 2 to Chapter 4, we will focus on the analysis of models (1.2)-(1.4), where the errors are assumed to satisfy several types of SGAR models. In Chapter 5, we will examine the house pricing problem in the Hong Kong real estate market, using the models considered in Chapter 2 to Chapter 4. More details about the outline of the thesis will be given below.

In Chapter 2, we will focus on nonparametric model (1.2), where the errors satisfy the torus SGAR model, the separable SGAR model and the unilateral SGAR model respectively. To estimate the unknown regression function $m(.)$, we propose the two-step estimation used in Martins-Filho & Yao (2009), who considered the local polynomial fitting with some temporally correlated errors. For the first-step estimator, the consistency with the three types of errors as well as the asymptotic normality with separable and unilateral SGAR-type errors are investigated. For the second-step estimator, the asymptotic normality with all types of errors are established. Some practical technique about

the selection of bandwidths and the estimation of the SGAR coefficients are discussed. Simulations are conducted to show the performance of our estimation.

In Chapter 3, we will focus on the additive model (1.3), where the errors also satisfy the three SGAR models mentioned above. The two-step estimation procedure is proposed, and the local linear fitting and marginal integration technics are adopted. Similar theoretical analyses as the ones in Chapter 2 are considered. Simulation results are given, accompanied by cross comparisons of our proposed fitting in Chapter 2, Chapter 3 and the traditional estimation of additive model with marginal integration technique.

In Chapter 4, we will further consider the partially linear model (1.4). The errors are still assumed to satisfy the three structures. The two-step estimation and local linear fitting technic are used. For the first-step estimation, the consistency of the estimators of both $\boldsymbol{\lambda}$ and $m(.)$ function is shown, given that the errors satisfy all types of SGAR structure. Moreover, the asymptotic normality of $\boldsymbol{\lambda}$ and $m(.)$ is also established, given that the errors satisfy the separable and unilateral SGAR model. For the second-step estimation, the asymptotic normality of the estimators with all types of errors is considered. Several simulations are conducted to assess our estimation.

In Chapter 5, a case study about the Hong Kong real estate market is shown. Data from a 9×10 grided area in Kowloon is collected. Each data contain the information of the average price per feet^2 and five explanatory factors. It is our aim to identify the factors that determine the price of houses. With consideration of the spatial dependency, we apply all the models considered in our thesis to this empirical problem.

At last, a brief conclusion is given in Chapter 6.

Chapter 2

Local Linear Regression with SGAR-type Error

2.1 Literature Review

The local polynomial regression estimation is a nonparametric method that is widely considered in recent years. The main idea of this kernel smoothing method is to obtain the estimate of the regression function by locally fitting a p-th degree polynomial to the data via weighted least square method. Such method is attractive because it presents many advantages, such as its good boundary behavior and adaptation to estimate the regression derivatives. A great amount of literature has focused on this topic, for example, Stone (1977), Cleveland (1979), Tsybakov (1986), Muller (1988), Fan (1992, 1993), Fan & Gijbels (1992, 1995). See more reviews in Section 1.3 .

In the paper mentioned above, it is assumed that the observations are *i.i.d.*. Such assumption is usually not satisfied in practical research. Masry & Fan (1997) studied the

local polynomial regression estimation with the process $\{x_i, y_i\}$ being strongly mixing or ρ-mixing. They showed that the estimators remain consistent and asymptotically normal when the independent assumption of the observation is relaxed. As a result, the local polynomial regression becomes useful in nonlinear time series modeling. Hallin et al. (2004) further generalized the method to the spatial process modeling. They considered the estimation of the regression function $m(.)$, which is a function of a d-dimensional vector $\mathbf{X_i}$ with $d \geq 2$ and $\mathbf{i} \in \mathbb{Z}^N$, $N > 1$. Under the assumption of the random fields $\{\mathbf{X_i}, Y_\mathbf{i}\}$ satisfying strict stationarity and some spatially mixing condition, they established the asymptotic normality of the estimator. To avoid the number of parameters exploding as the dimension of $\mathbf{X_i}$ increases, they performed the fitting only with the first degree polynomial. Hence their approach belongs to one of the local linear fitting.

Some other scholars also consider the modified local polynomial regression estimation without the *i.i.d.* assumption. For instance, Francisco-Fernandez & Vilar-Fernandez (2001). However, unlike Masry & Fan (1997) and Hallin et al.(2004) who assumed that the explanatory variables are autocorrelated, they assumed the error process $\{\epsilon_i\}$ to be strongly mixing and x_i's fixed. An advantage of their idea is that when the specific structure of the error term is known, one can make use of such additional information to obtain some improvement in estimation. To illustrate this, Vilar-Fernandez & Francisco-Fernandez (2002) assumed the errors to follow the AR(1) model, and obtained a modified estimator by firstly transforming the whole model to get uncorrelated errors and secondly applying the local polynomial fitting. A drawback of their method is that it is only suitable for the case with fixedly designed and equally spaced explanatory variables. When the explanatory variables are random, their method does not work.

Martins-Filho & Yao (2009) considered another modified local polynomial estimation with general parametric error structure. In their approach, the explanatory variables are assumed random and independent, and the residuals follow some general type of temporal process. A new two-step fitting procedure is proposed, which firstly forms a new process with the same conditional mean of the original one but uncorrelated residuals, and secondly applies the local polynomial fitting to the new process. In the area of spatial analysis, the local polynomial fitting with spatially correlated errors is also considered. For example, Francisco-Fernandez & Opsomer (2005) developed a bandwidth selection method taking into account the effect of the spatially correlated errors. See also the work of Liu (2001).

The local polynomial fitting with spatially correlated errors is a good idea to deal with the spatially dependent data. However, this topic has not yet been thoroughly studied. For instance, the asymptotic normality of the local polynomial estimator with spatially correlated errors has not been established, especially when the errors satisfy some multilateral model on a lattice. Moreover, the two-step estimation in Martins-Filho & Yao (2009) with spatially correlated errors has not been considered either. Both problems are charming to us, and it is the initial motivation of our work in this chapter.

The outline of this chapter is as follow. In the second section, the nonparametric model with three distinct SGAR-type errors is presented. The two-step estimation in Martins-Filho & Yao (2009) will be adopted. In the third section, some theoretical results of the estimators with distinct error structures are shown. In the fourth section, a two-step bandwidth selection approach as well as the estimation of the coefficients in the SGAR model are considered. In the fifth section, several simulations are conducted to assess

the performance of our estimation. In the sixth section, the proofs of the mathematical results in this chapter are given. And the last section is a summary conclusion.

2.2 Model and Estimation

We focus on the model

$$Y_{\mathbf{s}} = m(\mathbf{X}_{\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{2.1}$$

where $\mathbf{s} = (s_1, s_2) \in \mathbb{Z}^2$ is the index of location from a $n_1 \times n_2$ rectangular lattice Ω, $m(.)$ is some unknown function, $Y_{\mathbf{s}}$ is the response variable and $X_{\mathbf{s}}$ is a d-dimensional explanatory random vector. Moreover, $\mathbf{X}_{\mathbf{s}}$ for $\mathbf{s} \in \Omega$ are assumed *i.i.d.*, and $\epsilon_{\mathbf{s}}$ satisfies one of the following simultaneous Gaussian autoregressive (SGAR) models,

$$\epsilon_{\mathbf{s}} = \theta_1(\epsilon^*_{(s_1-1,s_2)} + \epsilon^*_{(s_1+1,s_2)}) + \theta_2(\epsilon^*_{(s_1,s_2-1)} + \epsilon^*_{(s_1,s_2+1)}) + \tau_{\mathbf{s}} \ , \tag{2.2}$$

$$\begin{aligned}\epsilon_{\mathbf{s}} = {} & \theta_1(\epsilon_{(s_1-1,s_2)} + \epsilon_{(s_1+1,s_2)}) + \theta_2(\epsilon_{(s_1,s_2-1)} + \epsilon_{(s_1,s_2+1)}) \\ & + \theta_1\theta_2(\epsilon_{(s_1-1,s_2-1)} + \epsilon_{(s_1-1,s_2+1)} + \epsilon_{(s_1+1,s_2-1)} + \epsilon_{(s_1+1,s_2+1)}) + \tau_{\mathbf{s}} \ ,\end{aligned} \tag{2.3}$$

$$\epsilon_{\mathbf{s}} = \theta_1 \, \epsilon_{(s_1-1,s_2)} + \theta_2 \, \epsilon_{(s_1,s_2-1)} + \tau_{\mathbf{s}} \ . \tag{2.4}$$

where $\tau_{\mathbf{s}}$ is normally distributed with zero mean and variance σ^2_τ, $\{\tau_{\mathbf{s}}\}$ is independent of $\{\mathbf{X}_{\mathbf{s}}\}$, $\epsilon^*_{\mathbf{s}} = \epsilon_{\mathbf{u}}$ with $\mathbf{u} = \Big((s_1 - 1 \ mod \ n_1) + 1, (s_2 - 1 \ mod \ n_2) + 1\Big)$ if $\mathbf{s} \notin \Omega$, and $\epsilon^*_{\mathbf{s}} = \epsilon_{\mathbf{s}}$ if $\mathbf{s} \in \Omega$.

(2.2) is a torus SGAR model, which is considered in Kashyap (1979, 1980) and Kashyap & Chellappa (1983). Although the torus assumption is quite artificial, it has an important advantage that all the neighbors of the random variables at the boundary can be well defined. Therefore, such assumption is usually used in spatial analysis. (2.3) is a separable

SGAR model, and same as (2.2), it is bilateral. This model is named after its property that the transfer function can be presented as the product of two one-dimensional polynomial. See more about the separable models in Martin (1979), Basawa et al.(1992) and Basawa (1995). (2.4) is a unilateral SGAR model, which is also known as the causal-type or quadrant-type model. The random variable in this model depends only on the neighbors on a quadrant plain, which may be too idealized for practice. However, such model can be seen as a straight generalization of the AR model in time series, and the Markovian property is naturally held. That release the difficulty in the theoretical analysis, and hence this model is broadly used. For more details, see the work of Tjøstheim (1978, 1983). Note that (2.2)-(2.4) are all the first-order SGAR models, however, our analysis can be generalized to the cases with any other order straightforwardly. We only consider the errors with the first-order SGAR models in this thesis, because these models are the most practical ones in spatial analysis. In time series, 100 observations is usually adequate for the estimation of a multi-order model. Nevertheless, observations from a 10×10 squared area are not quite sufficient for the estimation of a multi-order spatial model, as there are only 10 points on each edge. This kind of reason makes the first-order SGAR model the most feasible choice in practice.

If function $m(.)$ in (2.1) is differentiable at $\mathbf{x}$ with gradient $m'(\mathbf{x})$, we can approximate $m(.)$ in the neighborhood of $\mathbf{x}$ by

$$m(\mathbf{X_s}) \approx m(\mathbf{x}) + (m'(\mathbf{x}))^T(\mathbf{X_s} - \mathbf{x}) \triangleq \beta_0(\mathbf{x}) + \sum_{i=1}^{d} \beta_i(\mathbf{x}) \frac{(\mathbf{X_s} - \mathbf{x})_i}{h_i}, \tag{2.5}$$

where $\beta_i(\mathbf{x}) = h_i \dfrac{\partial\, m(\mathbf{x})}{\partial x_i}$, $(\boldsymbol{\delta})_i$ denotes the i-th element of the vector $\boldsymbol{\delta}$, and h_i is a sequence of bandwidth tending to zero as $\mathbf{n} = (n_1, n_2)$ tends to infinity. Here, we write $\mathbf{n} \to \infty$ as

$min\{n_1, n_2\} \to \infty$. So locally, estimating $m(\mathbf{X_s})$ is approximately equivalent to estimating $(\beta_0(\mathbf{x}), \beta_1(\mathbf{x}), \ldots, \beta_d(\mathbf{x}))$, and the estimates can be obtained from

$$\underset{(\beta_0, \ldots, \beta_d)\in\mathbb{R}^{d+1}}{\operatorname{argmin}} (nh_\pi)^{-1} \sum_{\mathbf{s}\in\Omega} \left(Y_\mathbf{s} - \beta_0(\mathbf{x}) - \sum_{i=1}^{d} \beta_i(\mathbf{x}) \frac{(\mathbf{X_s}-\mathbf{x})_i}{h_i} \right)^2 K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x})\Big), \tag{2.6}$$

where $n = n_1 \times n_2$, $h_\pi = \prod_{i=1}^{d} h_i$, $\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}) = ((\mathbf{X_s}-\mathbf{x})\odot(1/\mathbf{h}))^T$, $\mathbf{h} = (h_1, \cdots, h_d)^T$, $\odot$ denotes the element-wise product of two matrix, $K_\pi(\boldsymbol{\delta}) = \prod_{i=1}^{d} K(\boldsymbol{\delta}_i)$ with $K(.)$ being some kernel. Arrange the points $\mathbf{s} \in \Omega$ in arbitrary order and denote them by $\mathbf{s}_1, \cdots, \mathbf{s}_n$ respectively. Throughout this thesis, we use $\mathbf{s}$ as an arbitrary location index in Ω, while $\mathbf{s}_i, i = 1, \cdots, n$, as the arranged indexes whenever necessary. With the arrangement of the locations, (2.6) can be written equivalently in the matrix form:

$$\underset{\boldsymbol{\beta}\in R^{d+1}}{\operatorname{argmin}} \left(\mathbf{Y} - \vec{\mathbf{X}}(\mathbf{x})\boldsymbol{\beta}(\mathbf{x})\right)^T \mathbf{W}(\mathbf{x}) \left(\mathbf{Y} - \vec{\mathbf{X}}(\mathbf{x})\boldsymbol{\beta}(\mathbf{x})\right), \tag{2.7}$$

where $\mathbf{Y} = (Y_{\mathbf{s}_1}, \cdots, Y_{\mathbf{s}_n})^T$, $\mathbf{W}(\mathbf{x}) = (nh_\pi)^{-1}\text{diag}\Big(K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_1} - \mathbf{x})), \cdots, K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_n} - \mathbf{x}))\Big)$, $\boldsymbol{\beta}(\mathbf{x}) = \left(\beta_0(\mathbf{x}), \beta_1(\mathbf{x}), \cdots, \beta_d(\mathbf{x})\right)^T$ and $\vec{\mathbf{X}}(\mathbf{x})$ is a $n\times(d+1)$ matrix with the i-th row being $\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}_i}-\mathbf{x}) = (1,\ \boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_i}-\mathbf{x}))$. It follows that the local linear estimator of $\boldsymbol{\beta}(\mathbf{x})$ is given by

$$\hat{\boldsymbol{\beta}}(\mathbf{x}) = \left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x})\right)^{-1} \left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\mathbf{Y}\right), \tag{2.8}$$

and the estimator of $m(\mathbf{x})$, denoted as $\hat{m}(\mathbf{x})$, is given by the first element of $\hat{\boldsymbol{\beta}}(\mathbf{x})$.

When observations satisfy some spatially mixing condition, the asymptotic properties of $\hat{\boldsymbol{\beta}}(\mathbf{x})$ have been considered in Hallin et al.(2004). However, with our different setup where the error term $\epsilon_\mathbf{s}$ is spatially correlated, for example, as modeled in (2.2)-(2.4), the problem has not been completely studied. Moreover, $\hat{\boldsymbol{\beta}}(\mathbf{x})$ in (2.8) does not make use of the information of spatial dependency contained in $\epsilon_\mathbf{s}$. Hence some improvement

of the estimation can be pursued.

Rewrite (2.1) into the matrix form as

$$\mathbf{Y} = \mathbf{M} + \mathbf{E}, \tag{2.9}$$

where $\mathbf{M} = \left(m(\mathbf{X}_{\mathbf{s}_1}), \cdots, m(\mathbf{X}_{\mathbf{s}_n})\right)^T$, $\mathbf{E} = (\epsilon_{\mathbf{s}_1}, \cdots, \epsilon_{\mathbf{s}_n})^T$. Similarly, rewrite (2.2)-(2.4) as

$$(\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\mathbf{E} = \Upsilon, \tag{2.10}$$

where $\mathbf{I}$ is a unit matrix, $\Upsilon = (\tau_{\mathbf{s}_1}, \cdots, \tau_{\mathbf{s}_n})^T$, $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\mathbf{B}(\boldsymbol{\theta})$ is a $n \times n$ matrix with the (i, j)th element being the corresponding coefficient defined in (2.2)-(2.4) if $\epsilon_{\mathbf{s}_i}$ depends on $\epsilon_{\mathbf{s}_j}$ and zero otherwise. If we can form a new process $\{P_{\mathbf{s}}\}$ (in the matrix form) as

$$\mathbf{P} = \mathbf{M} + (\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\mathbf{E}, \tag{2.11}$$

it is readily to see that $P_{\mathbf{s}}$ has the same conditional mean as $Y_{\mathbf{s}}$ does, but the new error terms in (2.11) are *i.i.d.*. Therefore, we can apply the local linear fitting to the new process $P_{\mathbf{s}}$ to obtain an improved estimator of the $m(.)$ function.

However, the process $P_{\mathbf{s}}$ cannot be formed directly, as $\mathbf{M}$, $\mathbf{B}(\boldsymbol{\theta})$ and $\mathbf{E}$ in (2.11) are all unknown. Thus, we need to replace the unknown terms by some estimators. Rewrite (2.11) equivalently as

$$\mathbf{P} = (\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\mathbf{Y} + \mathbf{B}(\boldsymbol{\theta})\mathbf{M}. \tag{2.12}$$

If we can obtain some consistent estimator of $\mathbf{M}$, i.e. by the local linear fitting in (2.8), as well as some appropiate estimator of the matrix $\mathbf{B}(\boldsymbol{\theta})$, we then can form $\hat{\mathbf{P}} = (\hat{P}_{\mathbf{s}_1}, \cdots, \hat{P}_{\mathbf{s}_n})^T$ as

$$\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{B}(\hat{\boldsymbol{\theta}}))\mathbf{Y} + \mathbf{B}(\hat{\boldsymbol{\theta}})\hat{\mathbf{M}}. \tag{2.13}$$

As we need some new bandwidths for the improved estimators, let $\mathbf{g}=(g_1,\cdots,g_d)^T$ be a bandwidth vector distinct from $\mathbf{h}$, and consequently $g_\pi=\prod_{i=1}^d g_i$. Substitute $\mathbf{h}$ and h_π by $\mathbf{g}$ and g_π in the expression of $\vec{\mathbf{X}}(\mathbf{x})$ and $\mathbf{W}(\mathbf{x})$ below, then our modified estimator of the $m(.)$ function is given by

$$\tilde{m}(\mathbf{x}) = \eta^T \left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x})\right)^{-1} \left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\hat{\mathbf{P}}\right), \tag{2.14}$$

where η is a $d+1$ vector with the first element being 1 and 0 elsewise. Now let us sum up the full process. Our estimation runs in two steps.

Step 1. Obtain an initial estimator of $\boldsymbol{\beta}(\mathbf{x})$ by (2.8) with bandwidth $\mathbf{h}$. As we will show in the next section, $\hat{\boldsymbol{\beta}}(\mathbf{x})$ in (2.8) is consistent. So $m(.)$ and $\epsilon_\mathbf{s}$ can be estimated consistently.

Step 2. Estimate θ_1 and θ_2 with the estimate of $\epsilon_\mathbf{s}$, so that $\hat{\mathbf{P}}$ in (2.13) can be formed. Then, with bandwidth $\mathbf{g}$, we apply the local linear fitting to $\hat{\mathbf{P}}$ to get the improved estimator $\tilde{m}(\mathbf{x})$.

We denote this estimation procedure by LLR-SCE through this thesis, and call $\hat{m}(\mathbf{x})$ and $\tilde{m}(\mathbf{x})$ as the first- and second-step estimator of $m(\mathbf{x})$.

Remark 2.1 The SGAR model is quite suitable for the two-step estimation shown above. This is because the variance matrix of $\{\epsilon_\mathbf{s}\}$ can be decomposed as the square of $\mathbf{I}-\mathbf{B}(\boldsymbol{\theta})$ naturally. By contrast, the CGAR model is less appropriate. From (1.20), we can see that we need to decompose $(\mathbf{I}-\mathbf{C})^{-1}\mathbf{R}$. However, there is no guaranty that its estimator is positive-definiteness. Hence, the decomposition may not be executed successfully. Moreover, as we have mentioned in Section 1.2.2, the residuals in the

CGAR models are correlated, which does not fit our purpose to form a new process with *i.i.d*. errors. One may also replace the SGAR model by some geostatistics models, see Section 1.2.1. Nonetheless, some spatially mixing condition will be necessary for the establishment of asymptotic normality of the estimators, making the analysis more complicated.

Remark 2.2 One may run the two-step estimation procedure iteratively to earn a more accurate estimate. However, according to the simulation results given in section 2.5, this may not be necessary in practice, as the estimators given by the two-step estimation are usually accurate enough.

2.3 Theoretical Results

We will study the asymptotic properties of both $\hat{\boldsymbol{\beta}}(\mathbf{x})$ and $\tilde{m}(\mathbf{x})$ in this section. A set of assumptions should be given in advance.

(A1) The random vectors $\mathbf{X_s}, \mathbf{s} \in \Omega$, are *i.i.d*. with joint density $f(\mathbf{x})$, $0 < |f(\mathbf{x})| < \infty$. Moreover, $\{\mathbf{X_s}\}$ is independent of $\{\epsilon_\mathbf{s}\}$ and $\{\tau_\mathbf{s}\}$.

(A2) All the second derivatives of $m(.)$ exist and are continuous at all $\mathbf{x}$.

(A3) The random field $\{\epsilon_\mathbf{s}, \mathbf{s} \in \mathbb{Z}^2\}$ is strictly stationary; $\epsilon_\mathbf{s}$ has zero mean and finite variance.

(A3a) The coefficients in (2.2) satisfy $|\theta_1| + |\theta_2| < 1/2$.

(A3b) The coefficients in (2.3) satisfy $|\theta_1| < 1/2$, $|\theta_2| < 1/2$.

(A3c) The coefficients in (2.4) satisfy $|\theta_1| + |\theta_2| < 1$.

(A4) The kernel function $K(.)$ is symmetric, with bounded support, and Lipschitz continuous.

(A5) $h_i > 0, h_i \to 0$ as $\mathbf{n} \to \infty$; moreover, denote h_L the bandwidth with the slowest convergence rate, such as $h_L \in \{h_1, \cdots, h_d\}, h_L = O(h_k^C)$ for $0 < C \leq 1$ and $k = 1, \cdots, d$, then $\mathbf{h}$ converges to zero in the manner that $n = O((h_\pi h_L^4)^{-1})$ and $h_L^2 / h_i \to 0$ for $i = 1, \cdots, d$.

(A6) $g_i > 0$, $g_i \to 0$ as $\mathbf{n} \to \infty$; denote g_L the bandwidth with the slowest convergence rate, such as $g_L \in \{g_1, \cdots, g_d\}, g_L = O(g_k^C)$ for $0 < C \leq 1$ and $k = 1, \cdots, d$; $\mathbf{g}$ converges to zero in the manner that $n = O((g_\pi g_L^4)^{-1})$; moreover, $\lim_{\mathbf{n} \to \infty} h_L / g_L \to 0$.

(A7) There exist some sequences as: $l_1 \to \infty$, $l_2 \to \infty$ and $m \to \infty$ as $\mathbf{n} \to \infty$; $m / l_i \to 0$ and $l_i / n_i \to 0$ for $i = 1, 2$, as $\mathbf{n} \to \infty$.

For the sake of convenience, denote $\mathbf{U_n(x)} = \vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W(x)}\vec{\mathbf{X}}(\mathbf{x})$, $\mathbf{V_n(x)} = \vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W(x)Y}$ and $\mathbf{V_n^*(x)}$ as the centered vector of $\mathbf{V_n(x)}$. Then, $\mathbf{U_n(x)}$ is a $(d+1)\times(d+1)$ matrix with the (j,k)-th element being

$$\Big(\mathbf{U_n(x)}\Big)_{jk} = (nh_\pi)^{-1} \sum_{\mathbf{s} \in \Omega} \Big(\boldsymbol{\zeta}(\mathbf{X_s} - \mathbf{x})\Big)_j \Big(\boldsymbol{\zeta}(\mathbf{X_s} - \mathbf{x})\Big)_k K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s} - \mathbf{x})), \tag{2.15}$$

$j, k = 1, \cdots, d+1$; $\mathbf{V_n(x)}$ and $\mathbf{V_n^*(x)}$ are $d+1$ vectors with the j-th element being

$$\Big(\mathbf{V_n(x)}\Big)_j = (nh_\pi)^{-1} \sum_{\mathbf{s} \in \Omega} Y_\mathbf{s} \Big(\boldsymbol{\zeta}(\mathbf{X_s} - \mathbf{x})\Big)_j K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s} - \mathbf{x})),$$

$$\Big(\mathbf{V_n^*(x)}\Big)_j = (nh_\pi)^{-1} \sum_{\mathbf{s} \in \Omega} (Y_\mathbf{s} - m(\mathbf{X_s})) \Big(\boldsymbol{\zeta}(\mathbf{X_s} - \mathbf{x})\Big)_j K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s} - \mathbf{x})), \tag{2.16}$$

$j = 1, \cdots, d$. The following theorem shows the convergence rate of $\hat{\boldsymbol{\beta}}$, which ensures the first-step estimators are consistent.

Theorem 2.1 *If $\epsilon_{\mathbf{s}}$ follows any of the models in (2.2)-(2.4), with coefficients satisfying assumptions* (A3a)- (A3c) *respectively, and assumptions* (A1)-(A5) *hold, we have*

$$\hat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x}) - \frac{1}{2}\mathbf{U}^{-1}(\mathbf{x})\mathbf{T}(\mathbf{x}) = \mathbf{U}_{\mathbf{n}}^{-1}(\mathbf{x})\mathbf{V}_{\mathbf{n}}^{*}(\mathbf{x}) + o_p(h_L^2), \tag{2.17}$$

where ($\boldsymbol{\delta}$ is some d-dimensional vector)

$$\mathbf{U}(\mathbf{x}) = f(\mathbf{x}) \begin{pmatrix} \int K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} & \int \boldsymbol{\delta}^T K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} \\ \int \boldsymbol{\delta} K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} & \int \boldsymbol{\delta}\boldsymbol{\delta}^T K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} \end{pmatrix},$$

and $\mathbf{T}(\mathbf{x})$ is a $d+1$ dimensional vector with the ith element being ($\delta_0 \equiv 1$)

$$\left(\mathbf{T}(\mathbf{x})\right)_i = f(\mathbf{x}) \operatorname{tr}\left[m''(\mathbf{x}) \int \delta_{i-1}(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T) K_\pi(\boldsymbol{\delta}) d\boldsymbol{\delta}\right].$$

Moreover, we have $\mathbf{U}_{\mathbf{n}}(\mathbf{x}) = \mathbf{U}(\mathbf{x}) + o_p(1)$, $\mathbf{V}_{\mathbf{n}}^{*}(\mathbf{x}) = O_p((nh_\pi)^{-1/2})$, and it is clearly that $\mathbf{U}^{-1}(\mathbf{x})\mathbf{T}(\mathbf{x}) = O(h_L^2)$, therefore

$$\hat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x}) = O_p(h_L^2).$$

When $\epsilon_{\mathbf{s}}$ satisfies model (2.3) or (2.4), we can also establish the asymptotic normality of $\hat{\boldsymbol{\beta}}$. Since the analyses with $\epsilon_{\mathbf{s}}$ satisfying (2.3) are more involved, we omit the other ones in the following. In the literature about the local polynomial fitting with dependent data, some mixing conditions are usually adopted, for example, Masry & Fan (1997) and Hallin, Lu & Tran (2004). In our research, however, as the residuals satisfy some SGAR model, we may make good use of these special structures rather than applying the mixing assumptions. Therefore, we firstly transform (2.3) into an equivalent 2-dimensional moving average model, and then apply the 'large and small block' method with the truncated MA model. The theorem below then follows.

Theorem 2.2 *If $\epsilon_{\mathbf{s}}$ follows model (2.3) with coefficients satisfying* (A3b)*, and assumptions* (A1)-(A5), (A7) *hold, then*

$$(nh_\pi)^{1/2}\Big[\hat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x}) - \frac{1}{2}\mathbf{U}^{-1}(\mathbf{x})\mathbf{T}(\mathbf{x})\Big] \xrightarrow{D} N\Big(0, \mathbf{U}^{-1}(\mathbf{x})\,\Sigma(\mathbf{x})\,\mathbf{U}^{-1}(\mathbf{x})\Big), \tag{2.18}$$

where $\mathbf{U}(\mathbf{x})$ *and* $\mathbf{T}(\mathbf{x})$ *are defined in Theorem 2.1, and (*$\boldsymbol{\delta}$ *is some d-dimensional vector)*

$$\Sigma(\mathbf{x}) = \sigma_\epsilon^2 f(\mathbf{x}) \begin{pmatrix} \int K_\pi^2(\boldsymbol{\delta})d\boldsymbol{\delta} & \int \boldsymbol{\delta}^T K_\pi^2(\boldsymbol{\delta})d\boldsymbol{\delta} \\ \int \boldsymbol{\delta} K_\pi^2(\boldsymbol{\delta})d\boldsymbol{\delta} & \int \boldsymbol{\delta}\boldsymbol{\delta}^T K_\pi^2(\boldsymbol{\delta})d\boldsymbol{\delta} \end{pmatrix}.$$

At last, we consider the asymptotic normality of $\tilde{m}(\mathbf{x})$. Note that $\tilde{m}(\mathbf{x})$ depends on the estimator of $\boldsymbol{\theta}$, which is the parameter vector in the SGAR model. Therefore, the asymptotic properties of $\tilde{m}(\mathbf{x})$ are affected by the convergency of $\hat{\boldsymbol{\theta}}$. Following Martins-Filho & Yao (2009), we assume that some consistent estimator $\hat{\boldsymbol{\theta}}$ is available, say $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + o_p(1)$, then the following theorem can be established.

Theorem 2.3 *If the conditions in Theorem 2.1 as well as assumption* (A6) *are satisfied, and some consistent estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is available, then*

$$(ng_\pi)^{1/2}\Big(\tilde{m}(\mathbf{x}) - m(\mathbf{x}) - \text{Bias}(\mathbf{x})\Big) \xrightarrow{D} N\Big(0,\, \sigma^2\Big), \tag{2.19}$$

where $\text{Bias}(\mathbf{x}) = \text{tr}\Big[m''(\mathbf{x})\int(\boldsymbol{\delta}\boldsymbol{\delta}^T)\odot(\mathbf{g}\mathbf{g}^T)\,K_\pi(\boldsymbol{\delta})\,d\boldsymbol{\delta}\Big]$ *and* $\sigma^2 = f^{-1}(\mathbf{x})\,\sigma_\tau^2\int K_\pi^2(\boldsymbol{\delta})\,d\boldsymbol{\delta}$.

Remark 2.3 We do not consider the asymptotic normality of $\hat{\boldsymbol{\beta}}$ when $\epsilon_{\mathbf{s}}$ satisfies model (2.2). Due to the periodicity property of the torus SGAR model, the covariance of two random variables will not always decrease as their distance increase. Hence, it is not appropriate to use the 'large and small block' method to establish the asymptotic normality of $\hat{\boldsymbol{\beta}}$, making the analysis difficult. In spite of this, we can establish the asymptotic nor-

mality of the second-step estimator $\tilde{m}(\mathbf{x})$. This is because the asymptotic normality of $\tilde{m}(\mathbf{x})$ does not depend on the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, but on its convergence rate only.

2.4 Some Practical Considerations

2.4.1 Bandwidth Selection

Bandwidth selection is always an important problem in kernel estimation. Theoretically, the optimal bandwidth is usually defined to be one that minimizes the squared bias plus variance of the estimator, see Masry & Fan (1997) and Cai & Fan (2000). From the proofs of Theorem 2.1, we have

$$\text{Bias}\left(\hat{m}(\mathbf{x})\right) = \frac{1}{2}\sum_{i=1}^{d}\Big[\, h_i^2\, m^{ii}(\mathbf{x}) \int \delta_i^2 K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta}\,\Big], \tag{2.20}$$

$$\text{Var}\left(\hat{m}(\mathbf{x})\right) = (nh_\pi)^{-1}\sigma_\epsilon^2 f^{-1}(\mathbf{x}) \int K_\pi^2(\boldsymbol{\delta})\, d\boldsymbol{\delta}\,. \tag{2.21}$$

Therefore, define

$$Q = \frac{1}{4}\Big\{\sum_{i=1}^{d}\Big[h_i^2 m^{ii}(\mathbf{x}) \int \delta_i^2 K_\pi(\boldsymbol{\delta}) d\boldsymbol{\delta}\Big]\Big\}^2 + (nh_\pi)^{-1}\sigma_\epsilon^2 f^{-1}(\mathbf{x}) \int K_\pi^2(\boldsymbol{\delta}) d\boldsymbol{\delta}. \tag{2.22}$$

The optimal bandwidth can be obtained by minimizing Q with respect to $\mathbf{h} = (h_1, \cdots, h_d)$. However, the second derivatives of $m(\mathbf{x})$ and $f(\mathbf{x})$ are usually unknown, in practice, we may determine the optimal bandwidth via some data driven methods, i.e., the cross-validation(CV) method.

Note that our method runs in two steps, hence, distinct bandwidth should be used in each step. Recall that $\mathbf{h}$ and $\mathbf{g}$ are the bandwidths in the first and second step respec-

tively. We claim that the optimal **h** and **g** should be obtained 'simultaneously' rather than 'separatively'. What we call 'separatively' here means to obtain the optimal **h** via the CV method first and give an initial estimator of $m(.)$ function so that the new process $\{\hat{P}_{\mathbf{s}}\}$ can be formed, then obtain the optimal **g** via the CV method again with $\{\hat{P}_{\mathbf{s}}\}$. We will explain more about the idea below. Define

$$\mathrm{Cv}_1 = \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\Big(Y_s - \acute{m}(\mathbf{X}_{\mathbf{s}})\Big)^2,$$

where $\acute{m}(\mathbf{X}_{\mathbf{s}})$ is the predictor of $m(\mathbf{X}_{\mathbf{s}})$ in Step one, which is obtained by local linear fitting with observations not including the one at **s**. By the CV method, the optimal **h** can be obtained by minimizing Cv_1. Such method, however, may perform badly in our two step estimation. Actually, what we need in Step one is to get an initial estimator of $m(.)$. So the optimal **h** should be the one minimizing

$$\mathrm{Cv}_2 = \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\Big(m(\mathbf{X}_{\mathbf{s}}) - \acute{m}(\mathbf{X}_{\mathbf{s}})\Big)^2,$$

rather than that minimizing Cv_1. Obviously, the CV method tends to gain bandwidth which will fit $m(\mathbf{X}_{\mathbf{s}})$ closer to the observation $Y_{\mathbf{s}}$, but this is not our purpose.

Now define $\acute{P}_{\mathbf{s}}$ similar to $\hat{P}_{\mathbf{s}}$ but with $\hat{m}(\mathbf{X}_{\mathbf{s}})$ substituted by $\acute{m}(\mathbf{X}_{\mathbf{s}})$, and

$$\mathrm{Cv}_3 = \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\Big(Y_{\mathbf{s}} - \grave{m}(\mathbf{X}_{\mathbf{s}}) - \grave{\epsilon}_{\mathbf{s}}\Big)^2,$$

where $\grave{m}(\mathbf{X}_{\mathbf{s}})$ is the predictor of $m(\mathbf{X}_{\mathbf{s}})$ in Step two, which is obtained by local linear fitting with $\acute{\mathbf{Z}}_{\mathbf{s}}$ but not including the one at **s**, and $\grave{\epsilon}_{\mathbf{s}}$ is the estimate of $\epsilon_{\mathbf{s}}$ based on $Y_{\mathbf{s}} - \grave{m}(\mathbf{X}_{\mathbf{s}})$. Therefore, by minimizing Cv_3, the optimal **h** and **g** can be determined simultaneously.

Obviously, it is a better approach to determine the optimal **h** and **g** simultaneously. Unfortunately, it always leads to large amount of calculations. Therefore, we propose an alternative method, and main steps are described below.

Step I. Obtain an initial **h** via the CV method.

Step II. With **h** given in Step I, determine optimal **g** by minimizing Cv_3.

Step III. With **g** given in Step II, determine optimal **h** by minimizing Cv_3, then repeat Step II and III until **h** and **g** become stable.

In this manner the optimal bandwidths are determined 'iteratively'. We can see that the bandwidth obtained in each iteration keep approaching 'optimal' in the sense that Cv_3 decreases monotonically. Although there is no guarantee the bandwidth determined iteratively will converge to the one determined simultaneously, the proposed iterative method is still meaningful, as it works better than the one determining the distinct bandwidth separately, while the amount of calculations involved is practically acceptable.
We shall conduct below a simulation to show the performance of the proposed bandwidth selection method. Consider

$$Y_{(s_1,s_2)} = sin\Big(\pi X_{(s_1,s_2)}\Big) + \epsilon_{(s_1,s_2)}\,,$$

$$\epsilon_{(s_1,s_2)} = 0.39\Big(\epsilon_{(s_1-1,s_2)} + \epsilon_{(s_1+1,s_2)}\Big) - 0.1\Big(\epsilon_{(s_1,s_2-1)} + \epsilon_{(s_1,s_2+1)}\Big) + \tau_{(s_1,s_2)}\,,$$

where $\tau_{(s_1,s_2)} \sim N(0,1)$. Observations are drawn from a 10×10 area, and torus setup is adopted to ensure all neighbors of the variables at the boundary are well defined. We run the simulation 100 times with the optimal bandwidth determined separately and

iteratively. Define

$$\mathrm{MSE}_1 = \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\Big(m(X_\mathbf{s}) - \check{m}(X_\mathbf{s})\Big)^2,$$

$$\mathrm{MSE}_2 = \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\Big(Y_\mathbf{s} - \check{m}(X_\mathbf{s}) - \check{\epsilon}_\mathbf{s}\Big)^2,$$

where $\check{m}(X_\mathbf{s})$ is the estimate of $m(X_\mathbf{s})$ obtained by (2.14), and $\check{\epsilon}_\mathbf{s}$ is the estimate of $\epsilon_\mathbf{s}$ based on $Y_\mathbf{s} - \check{m}(X_\mathbf{s})$. The simulation results are presented in Table 2.1. We can see that the means and variance of $\mathrm{MSE}_\mathbf{1}$ and $\mathrm{MSE}_\mathbf{2}$ of the bandwidth selected iteratively are smaller than those of the bandwidth selected separatively. Moreover, in the 100 simulations, there are 94 times that the bandwidths selected iteratively are coincident with those selected simultaneously. That means the iterative method of selecting bandwidth performs well and with an acceptable calculational speed.

.

Table 2.1: Simulation for bandwidth selection methods

	Separatively	**Iteratively**
mean of MSE_1	0.2245	0.1951
mean of MSE_2	0.8205	0.7620
var of MSE_1	0.0278	0.0213
var of MSE_2	0.0491	0.0344

2.4.2 Estimation of SGAR coefficients

In our method, we need to estimate the SGAR coefficients in (2.2)-(2.4). Note that the model in (2.4) is unilateral, while the others are bilateral. By Tjøstheim (1978), we see that the Yule-Walker type estimators for the unilateral SGAR models are consistent, however, such estimators are not consistent for the multilateral SGAR models, see Whittle (1954). For the maximum likelihood estimation, the case is also complicated. The

main difficulty is how to evaluate the determinant $|\mathbf{I}-\mathbf{B}(\boldsymbol{\theta})|$ in the likelihood function. As it is hard to give a simple expression to this determinant, approximations are used in the estimation, see Whittle (1954), Guyon (1982), Robinson & Vidal Sanz (2006), Kashyap & Chellappa (1983) and Ali (1979). In our thesis, we apply the method in Kashyap & Chellappa (1983) to estimate the coefficients in (2.2), and a modified approach which generalizes the work of Ali (1979) to estimate the coefficients in (2.3). The main procedure will be presented below.

Let $\mathbf{I}_c$ be the $c \times c$ unit matrix, $\boldsymbol{\Gamma}_c$ be a $c \times c$ matrix with the (i, j)th element being 1 if $j = i - 1$ and 0 elsewise. By (2.3) we have

$$\mathbf{I}-\mathbf{B}(\boldsymbol{\theta})=\mathbf{I}_{n_1}\otimes\mathbf{I}_{n_2}-\theta_1(\boldsymbol{\Gamma}_{n_1}+\boldsymbol{\Gamma}'_{n_1})\otimes\mathbf{I}_{n_2}-\theta_2\,\mathbf{I}_{n_1}\otimes(\boldsymbol{\Gamma}_{n_2}+\boldsymbol{\Gamma}'_{n_2})-\theta_1\theta_2(\boldsymbol{\Gamma}_{n_1}+\boldsymbol{\Gamma}'_{n_1})\otimes(\boldsymbol{\Gamma}_{n_2}+\boldsymbol{\Gamma}'_{n_2}).$$

where $\otimes$ denotes the Kronecker product of matrices. As $(\boldsymbol{\Gamma}_{n_1}+\boldsymbol{\Gamma}'_{n_1})\otimes\mathbf{I}_{n_2}$, $\mathbf{I}_{n_1}\otimes(\boldsymbol{\Gamma}_{n_2}+\boldsymbol{\Gamma}'_{n_2})$ and $(\boldsymbol{\Gamma}_{n_1}+\boldsymbol{\Gamma}'_{n_1})\otimes(\boldsymbol{\Gamma}_{n_2}+\boldsymbol{\Gamma}'_{n_2})$ commute mutually, they can be diagonalized simultaneously by an orthogonal transformation. Denote $\mathrm{D}_{1i}, i = 1, \cdots, n_1$, the characteristic roots of $(\boldsymbol{\Gamma}_{n_1}+\boldsymbol{\Gamma}'_{n_1})$ and D_{2j}, $j = 1, \cdots, n_2$, the characteristic roots of $(\boldsymbol{\Gamma}_{n_2}+\boldsymbol{\Gamma}'_{n_2})$. Then

$$|\mathbf{I}-\mathbf{B}(\boldsymbol{\theta})| = \prod_{i,j}\Big(1-\theta_1\mathrm{D}_{1i}-\theta_2\mathrm{D}_{2j}-\theta_1\theta_2\mathrm{D}_{1i}\mathrm{D}_{2j}\Big). \tag{2.23}$$

By Ord (1975), D_{1i} and D_{2j} can be given approximately by

$$\mathrm{D}_{1i} = 2\cos\Big\{\pi i/(n_1+1)\Big\} \quad \textit{and} \quad \mathrm{D}_{2j} = 2\cos\Big\{\pi j/(n_2+1)\Big\}.$$

Recall that the log-likelihood function can be given as

$$\log l = -\frac{1}{2}\,n\log\sigma_\tau^2 + \log|\mathbf{I}-\mathbf{B}(\boldsymbol{\theta})| - \mathbf{E}^T(\mathbf{I}-\mathbf{B}(\boldsymbol{\theta}))^T(\mathbf{I}-\mathbf{B}(\boldsymbol{\theta}))\,\mathbf{E}/(2\sigma_\tau^2). \tag{2.24}$$

It is easy to see that the ML estimators of σ_τ^2 can be given by

$$\hat{\sigma}_\tau^2 = \frac{1}{n}\mathbf{E}^T(\mathbf{I}-\mathbf{B}(\boldsymbol{\theta}))^T(\mathbf{I}-\mathbf{B}(\boldsymbol{\theta}))\,\mathbf{E},$$

and $\hat{\boldsymbol{\theta}}$ can be obtained by minimizing $\hat{\sigma}_\tau^2|\mathbf{I}-\mathbf{B}(\boldsymbol{\theta})|^{-2/n}$. Then, with (2.23), the ML estimators are available.

2.5 Simulations

In this section, we shall conduct several simulations to asses the performance of our method. Observations $\{Y_\mathbf{s}, \mathbf{X}_\mathbf{s}\}$ will be generated with errors $\{\epsilon_\mathbf{s}\}$ satisfying (2.2)-(2.4). We will conduct the simulations with 3 different sample sizes, namely 10×10, 15×15 and 20×20, and will replicate 100 times for each sample size. The results of our estimation will be compared with those given by local linear fitting.

Simulation 2.1 The testing model is

$$Y_\mathbf{s} = \sin(\pi X_\mathbf{s}) + \epsilon_\mathbf{s}, \tag{2.25}$$

where $X_\mathbf{s} \sim U(0,4)$, $\epsilon_\mathbf{s}$ is modeled as (2.2) -(2.4) with $\tau_\mathbf{s} \sim N(0,1)$ and coefficients $\boldsymbol{\theta} = (\theta_1, \theta_2)$ being $(0.38, -0.1)$, $(0.3, -0.2)$ and $(0.4, 0.3)$ respectively.

First, we consider Figure 2.3 and Figure 2.4, which present two typical fittings with errors satisfying (2.2) and (2.3) respectively. The dots refer to the observations and the solid line refers to the curve of the true function $\sin(\pi X_\mathbf{s})$. The estimate of $m(.)$ given by the local linear fitting and our estimate are also shown. It can be seen that the estimates obtained by local linear fitting depart from the true values greatly. Comparatively, the

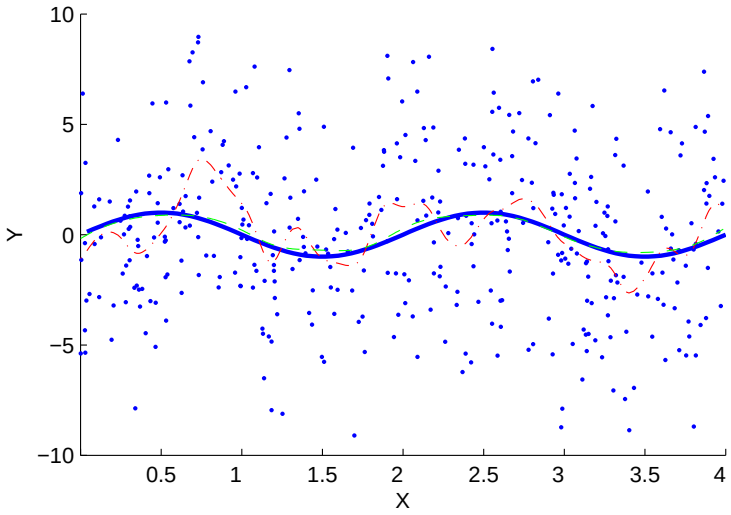


Figure 2.1: Simulation 2.1. Fitting with torus SGAR-type error. Solid line refers to true values of the $m(.)$ function; dashed line refers to estimates given by LLR-SCE; and dash-dot line refers to estimates given by local linear fitting. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

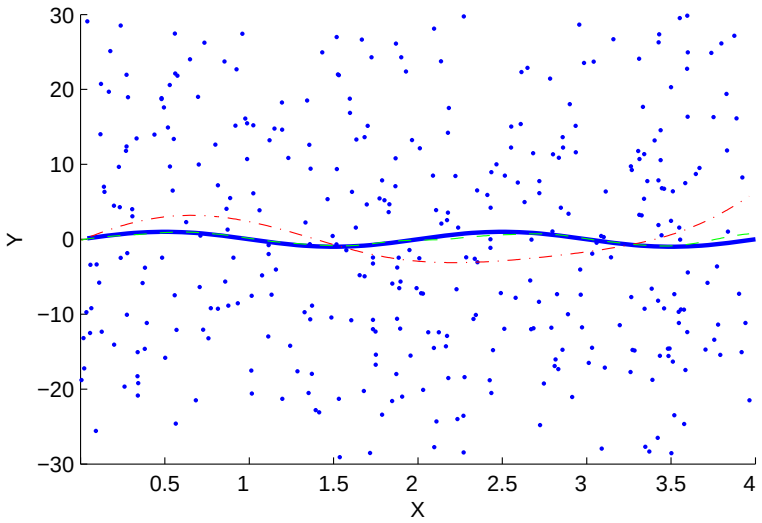


Figure 2.2: Simulation 2.1. Fitting with separable SGAR-type error. Solid line refers to the true values of the $m(.)$ function; dashed line refers to the estimators given by LLR-SCE; and dash-dot line refers to those given by local linear fitting. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

.

Table 2.2: Simulation 2.1. MSE for the estimators of $m(.)$.

Sample Size	10×10	15×15	20×20
Torus$^{(a)}$			
mean of $\text{MSE}_o^{(d)}$	0.6418	0.3692	0.2122
mean of $\text{MSE}_m^{(e)}$	0.1517	0.0613	0.0393
var of MSE_o	0.1941	0.0501	0.0143
var of MSE_m	0.0084	0.0011	0.0003
Separable$^{(b)}$			
mean of MSE_o	3.5894	7.2986	1.2509
mean of MSE_m	0.1270	0.0755	0.0350
var of MSE_o	34.1649	268.7327	1.0319
var of MSE_m	0.0049	0.0045	0.0002
Unilateral$^{(c)}$			
mean of MSE_o	0.1938	0.0953	0.0518
mean of MSE_m	0.1719	0.0835	0.0463
var of MSE_o	0.0160	0.0037	0.0006
var of MSE_m	0.0139	0.0034	0.0006

$^{(a)}$ Torus: $\epsilon_{\mathbf{s}}$ generated by (2.2).
$^{(b)}$ Separable: $\epsilon_{\mathbf{s}}$ generated by (2.3).
$^{(c)}$ Unilateral: $\epsilon_{\mathbf{s}}$ generated by (2.4).
$^{(d)}$ MSE_o refers to the MSE given by local linear fitting.
$^{(e)}$ MSE_m refers to the MSE given by LLR-SCE.

.

Table 2.3: Simulation 2.1. Estimators of $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

Sample Size	10×10	15×15	20×20
Torus$^{(*)}$			
mean of $\hat{\theta}_1$	0.3630	0.3819	0.3906
mean of $\hat{\theta}_2$	-0.1295	-0.1220	-0.1163
Separable$^{(*)}$			
mean of $\hat{\theta}_1$	0.2874	0.2917	0.2889
mean of $\hat{\theta}_2$	-0.2140	-0.2052	-0.2084
Unilateral$^{(*)}$			
mean of $\hat{\theta}_1$	0.3284	0.3506	0.3403
mean of $\hat{\theta}_2$	0.2372	0.2608	0.2640

$^{(*)}$ Torus, Separable and Unilateral are same defined as the ones in Table 2.2.

estimates given by LLR-SCE are much closer to the true. Furthermore, the curve given by LLR-SCE is smoother. More results are given in Table 2.2 and Table 2.3.

From Table 2.2, we can see that almost all the MSE of the estimates given by both the local linear fitting and LLR-SCE decrease as the sample size increases. These certify the consistency of $\hat{\boldsymbol{\beta}}(\mathbf{x})$ and $\tilde{m}(\mathbf{x})$. The only exception is the case given by local linear fitting with $\epsilon_{\mathbf{s}}$ satisfying (2.3) and sample size 15×15. For the Separable case in Table 2.2, we can see that the mean of MSE given by local linear fitting are 3.5894, 7.2986 and 1.2509, for sample sizes 10×10, 15×15 and 20×20 respectively. The reason for this is that the dependence memory of $\epsilon_{\mathbf{s}}$ generated by (2.3) is relatively long. The first sample size 10×10 is not large enough to ensure the variance of $\epsilon_{\mathbf{s}}$ to reach it stable value. As sample size increases from 10×10 to 15×15, the empirical variance of $\epsilon_{\mathbf{s}}$ increases, too. This leads to the increment of the MSE_o. As sample size keeps on increasing, the variance of $\epsilon_{\mathbf{s}}$ becomes stable gradually, so MSE_o turns to decrease. See also (2.21) that the variance of $\hat{m}(\mathbf{x})$ depends on the variance of $\epsilon_{\mathbf{s}}$. We can further find that, as different from MSE_o, MSE_m keeps decreasing all the time. This is because the variance of $\tilde{m}(\mathbf{x})$ depends on the variance of $\tau_{\mathbf{s}}$ instead, see Theorem 2.3.

Table 2.2 also shows that the MSE obtained by LLR-SCE are small than those given by local linear fitting uniformly, indicating that LLR-SCE is preferable. Another phenomenon drawing our attentions is that our method makes large improvements over the local linear fitting, when $\epsilon_{\mathbf{s}}$ satisfies (2.2) and (2.3). However, when $\epsilon_{\mathbf{s}}$ satisfies (2.4), the improvement is less significant. This is due to the special structure of the errors chosen. Model (2.4) is unilateral, while (2.2) and (2.3) are bilateral. The volatility of errors generated from the former is small, in contrast to those from the later. Obviously,

our method is more meaningful when the volatility of errors is large.

Table 2.3 shows the simulation results of $\hat{\boldsymbol{\theta}}$. Note that the estimation of $\boldsymbol{\theta}$ is based on the first-step estimator $\hat{m}(\mathbf{x})$, so the accuracy of $\hat{\boldsymbol{\theta}}$ is affected by the estimation of the $m(.)$ function. In practice, one may obtain a better estimate of $\boldsymbol{\theta}$ basing on the second-step estimator $\tilde{m}(\mathbf{x})$, which has a small variance than $\hat{m}(\mathbf{x})$.

Simulation 2.2 The testing model is

$$Y_{\mathbf{s}} = 2 + \log(X_{\mathbf{s}}) - 2^{X_{\mathbf{s}}/2} + \epsilon_{\mathbf{s}}, \tag{2.26}$$

where $X_{\mathbf{s}} \sim U(0.01, 4)$. Other settings are same as the ones in Simulation 2.1.

In Simulation 2.2, we consider another testing model with single explanatory variable. Fitting results are shown in Figure 2.3, Figure 2.4 and Table 2.4, Table 2.5. According to these results, similar comments as those in Simulation 2.1 can be given.

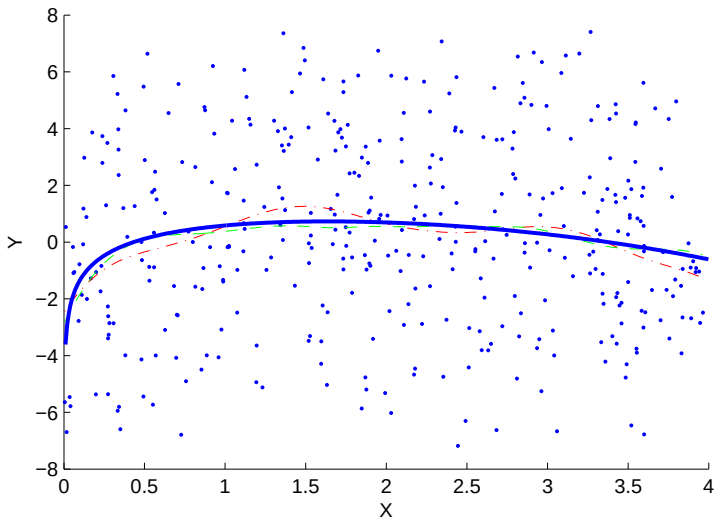


Figure 2.3: Simulation 2.2. Fitting with torus SGAR-type error. Solid line refers to true values of the $m(.)$ function; dashed line refers to estimates given by LLR-SCE; and dash-dot line refers to estimates given by local linear fitting. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

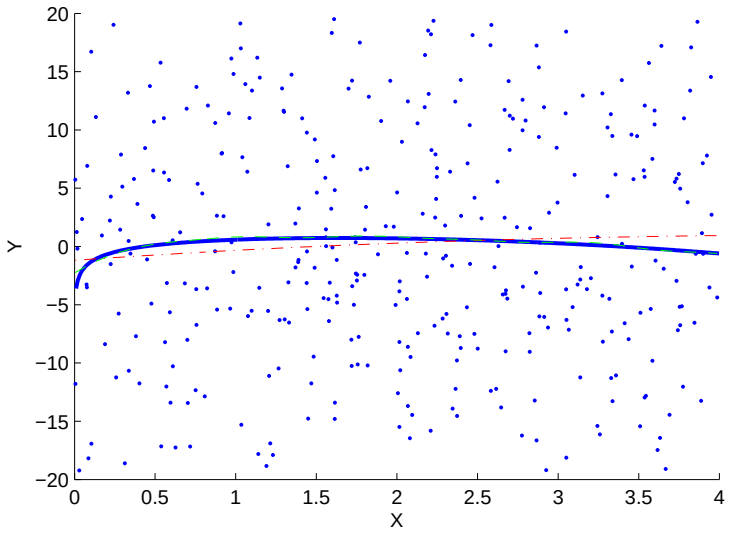


Figure 2.4: Simulation 2.2. Fitting with separable SGAR-type error. Solid line refers to the true values of the $m(.)$ function; dashed line refers to the estimators given by LLR-SCE; and dash-dot line refers to those given by local linear fitting. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

.

Table 2.4: Simulation 2.2. MSE for the estimators of $m(.)$.

Sample Size	10×10	15×15	20×20
Torus[(a)]			
mean of $\text{MSE}_o^{(d)}$	0.5522	0.2769	0.1765
mean of $\text{MSE}_m^{(e)}$	0.1379	0.0783	0.0564
var of MSE_o	0.2914	0.0353	0.0075
var of MSE_m	0.0109	0.0023	0.0010
Separable[(b)]			
mean of MSE_o	4.5963	7.4921	1.3215
mean of MSE_m	0.1205	0.0788	0.0436
var of MSE_o	93.9917	160.4989	3.4335
var of MSE_m	0.0067	0.0027	0.0006
Unilateral[(c)]			
mean of MSE_o	0.1689	0.0912	0.0736
mean of MSE_m	0.1597	0.0871	0.0716
var of MSE_o	0.0144	0.0037	0.0026
var of MSE_m	0.0151	0.0039	0.0036

[(a)] Torus: ϵ_s generated by (2.2).
[(b)] Separable: ϵ_s generated by (2.3).
[(c)] Unilateral: ϵ_s generated by (2.4).
[(d)] MSE_o refers to the MSE given by local linear fitting.
[(e)] MSE_m refers to the MSE given by LLR-SCE.

.

Table 2.5: Simulation 2.2. Estimators of $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

Sample Size	10×10	15×15	20×20
Torus(*)			
mean of $\hat{\theta}_1$	0.3707	0.3940	0.4004
mean of $\hat{\theta}_2$	-0.1281	-0.1134	-0.1087
Separable(*)			
mean of $\hat{\theta}_1$	0.2910	0.2888	0.2914
mean of $\hat{\theta}_2$	-0.2084	-0.2082	-0.2066
Unilateral(*)			
mean of $\hat{\theta}_1$	0.3207	0.3498	0.3660
mean of $\hat{\theta}_2$	0.2510	0.2608	0.2685

(*) Torus, Separable and Unilateral are same defined as the ones in Table 2.2.

Simulation 2.3 The testing model is

$$Y_{\mathbf{s}} = \sin(\pi X_{1,\mathbf{s}}) + \cos(\pi X_{2,\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{2.27}$$

where $X_{i,\mathbf{s}} \sim U(0, 4)$, $i = 1, 2$. Other settings are same as the ones in Simulation 2.1.

In Simulation 2.3, we consider a model with two explanatory variables. The results are shown in Table 2.6 and Table 2.7. We can see that our estimation still performs much better than the local linear fitting. The improvement is especially significant in the Torus and Separable cases. However, comparing with results of Simulation 1, the ones of Simulation 2 are obviously worse. This is because there are more explanatory variables in Simulation 2, and therefore the fitting suffers from the so-called curse of dimensionality.

.

Table 2.6: Simulation 2.3. MSE for the estimators of $m(.)$.

Sample Size	10×10	15×15	20×20
Torus$^{(a)}$			
mean of $\text{MSE}_o^{(d)}$	1.6022	1.0910	0.7169
mean of $\text{MSE}_m^{(e)}$	0.6055	0.2985	0.1968
var of MSE_o	1.5741	0.1970	0.0243
var of MSE_m	0.0298	0.0065	0.0019
Separable$^{(b)}$			
mean of MSE_o	10.3038	24.4096	2.7245
mean of MSE_m	0.7344	0.5233	0.1944
var of MSE_o	234.7356	2.80×10^3	4.2181
var of MSE_m	0.2645	0.1793	0.0016
Unilateral$^{(c)}$			
mean of MSE_o	0.5180	0.2729	0.2081
mean of MSE_m	0.4885	0.2485	0.1816
var of MSE_o	0.0172	0.0044	0.0020
var of MSE_m	0.0188	0.0045	0.0017

$^{(a)}$ Torus: ϵ_s generated by (2.2).
$^{(b)}$ Separable: with ϵ_s generated by (2.3).
$^{(c)}$ Unilateral: with ϵ_s generated by (2.4).
$^{(d)}$ MSE_o refers to the MSE given by local linear fitting.
$^{(e)}$ MSE_m refers to the MSE given by LLR-SCE.

.

Table 2.7: Simulation 2.3. Estimators of $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

Sample Size	10×10	15×15	20×20
Torus$^{(*)}$			
mean of $\hat{\theta}_1$	0.3142	0.3400	0.3573
mean of $\hat{\theta}_2$	-0.1440	-0.1428	-0.1318
Separable$^{(*)}$			
mean of $\hat{\theta}_1$	0.2814	0.2823	0.2829
mean of $\hat{\theta}_2$	-0.2187	-0.2117	-0.2127
Unilateral$^{(*)}$			
mean of $\hat{\theta}_1$	0.2455	0.3100	0.3067
mean of $\hat{\theta}_2$	0.1903	0.2324	0.2540

$^{(*)}$ Torus, Separable and Unilateral are same defined as the ones in Table 2.6.

2.6 Proofs

In the following, we will omit the notation $'(\mathbf{x})'$ in all the expressions wherever no confusion occurs. We first present several lemmas needed in the proof of Theorem 2.1. The first lemma is quite standard in the literature of local polynomial fitting, and the proofs can be given similarly as those of Lemma 2.2, hence we present it directly with proofs skipped.

Lemma 2.1 *Under assumptions* (A1)*,* (A4) *and* (A5)*, we have*

$$\mathbf{U_n} - \mathbf{U} = o_p(1), \tag{2.28}$$

where

$$\mathbf{U} = f(\mathbf{x}) \begin{pmatrix} \int K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} & \int \boldsymbol{\delta}^T K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} \\ \int \boldsymbol{\delta} K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} & \int \boldsymbol{\delta}\boldsymbol{\delta}^T K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta} \end{pmatrix}, \tag{2.29}$$

and $\boldsymbol{\delta}$ is some d dimensional vector.

Let $R(\mathbf{k}) \hat{=} \mathrm{Cov}(\epsilon_{\mathbf{s}}, \epsilon_{\mathbf{s+k}})$, where $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$. The lemma below is crucial for Theorem 2.1.

Lemma 2.2 *Given the condition that*

$$\sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} |R(k_1, k_2)| < \infty \tag{2.30}$$

and assumptions (A1), (A3)-(A5) *hold, then*

$$nh_\pi \mathrm{Var}(\mathbf{V}_\mathbf{n}^*) = \Sigma + o(1), \tag{2.31}$$

where Σ is defined in Theorem 2.2.

Proof of Lemma 2.2: For $j,k=1,\ldots,d+1$, we have

$$
\begin{aligned}
&\mathrm{Cov}\Big((\mathbf{V}_{\mathbf{n}}^*)_j,(\mathbf{V}_{\mathbf{n}}^*)_k\Big)\\
=\;&(nh_\pi)^{-2}\sum_{\mathbf{s}\in\Omega}\mathrm{Cov}\Big[\Big(Y_{\mathbf{s}}-m(\mathbf{X}_{\mathbf{s}})\Big)\Big(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}}-\mathbf{x})\Big)_j K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x})\Big),\Big(Y_{\mathbf{s}}-m(\mathbf{X}_{\mathbf{s}})\Big)\Big(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}}-\mathbf{x})\Big)_k K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x})\Big)\Big]\\
&+(nh_\pi)^{-2}\sum_{\substack{\mathbf{s}'\neq\mathbf{s}''\\ \mathbf{s}',\mathbf{s}''\in\Omega}}\mathrm{Cov}\Big[\Big(Y_{\mathbf{s}'}-m(\mathbf{X}_{\mathbf{s}'})\Big)\Big(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}'}-\mathbf{x})\Big)_j K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}'}-\mathbf{x})\Big),\Big(Y_{\mathbf{s}''}-m(\mathbf{X}_{\mathbf{s}''})\Big)\Big(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}''}-\mathbf{x})\Big)_k K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}''}-\mathbf{x})\Big)\Big]\\
\hat{=}\;&Q_1+Q_2\ .
\end{aligned}
$$

Let $\boldsymbol{\delta}=(\delta_1,\cdots,\delta_d)^T$ and $\delta_0\equiv 1$, then

$$
\begin{aligned}
Q_1 &= (nh_\pi)^{-2}\sum_{\mathbf{s}\in\Omega}E\left[(Y_{\mathbf{s}}-m(\mathbf{X}_{\mathbf{s}}))^2\,(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))_j\,(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))_k\,K_\pi^2(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\right]\\
&= (nh_\pi^2)^{-1}\sigma_\epsilon^2\int(\boldsymbol{\zeta}(\mathbf{r}-\mathbf{x}))_j\,(\boldsymbol{\zeta}(\mathbf{r}-\mathbf{x}))_k\,K_\pi^2(\boldsymbol{\zeta}_1(\mathbf{r}_{\mathbf{s}}-\mathbf{x}))\,f(\mathbf{r})\,d\mathbf{r}\\
&= (nh_\pi)^{-1}\sigma_\epsilon^2\int\delta_{j-1}\delta_{k-1}K_\pi^2(\boldsymbol{\delta})f(\boldsymbol{\delta}\odot\mathbf{h}+\mathbf{x})\,d\boldsymbol{\delta}\\
&= (nh_\pi)^{-1}\sigma_\epsilon^2 f(\mathbf{x})\int\delta_{j-1}\delta_{k-1}K_\pi^2(\boldsymbol{\delta})\,d\boldsymbol{\delta}+o((nh_\pi)^{-1}).
\end{aligned}
$$

$$
\begin{aligned}
|Q_2| &\le (nh_\pi)^{-2}\sum_{\substack{\mathbf{s}'\neq\mathbf{s}''\\ \mathbf{s}',\mathbf{s}''\in\Omega}}\left|\mathrm{Cov}(\epsilon_{\mathbf{s}'},\epsilon_{\mathbf{s}''})E\Big[\Big(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}'}-\mathbf{x})\Big)_j K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}'}-\mathbf{x})\Big)\Big]E\Big[\Big(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}''}-\mathbf{x})\Big)_k K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}''}-\mathbf{x})\Big)\Big]\right|\\
&= n^{-2}\sum_{\substack{\mathbf{s}'\neq\mathbf{s}''\\ \mathbf{s}',\mathbf{s}''\in\Omega}}\left|\mathrm{Cov}(\epsilon_{\mathbf{s}'},\epsilon_{\mathbf{s}''})\int\delta_{j-1}K_\pi(\boldsymbol{\delta})f(\boldsymbol{\delta}\odot\mathbf{h}+\mathbf{x})\,d\boldsymbol{\delta}\int\delta_{k-1}K_\pi(\boldsymbol{\delta})f(\boldsymbol{\delta}\odot\mathbf{h}+\mathbf{x})\,d\boldsymbol{\delta}\right|\\
&\le Cn^{-2}\Big[\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}(n_1-|k_1|)(n_2-|k_2|)\Big|R(k_1,k_2)\Big|-nR(0,0)\Big]\\
&\le Cn^{-1}\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}|R(k_1,k_2)|,
\end{aligned}
$$

which is equal to $O(n^{-1})$. Thus we have

$$
Cov\left((\mathbf{V}_{\mathbf{n}}^*)_j,(\mathbf{V}_{\mathbf{n}}^*)_k\right)=(nh_\pi)^{-1}\left[\sigma_\epsilon^2 f(\mathbf{x})\int\delta_{j-1}\delta_{k-1}K_\pi^2(\boldsymbol{\delta})d\boldsymbol{\delta}+o(1)\right].
$$

The proof is completed. □

To ensure Lemma 2.2 is valid for our analysis, we need to verify that the SGAR models in (2.2)-(2.4) satisfy condition (2.30). Consider model (2.2), we need the following lemma.

Lemma 2.3 *If* $\epsilon_{\mathbf{s}}$ *satisfies model (2.2) and assumptions* (A3), (A3a) *hold, then condition (2.30) is valid.*

Proof of Lemma 2.3: We first consider model (2.2), the torus SGAR model, on a finite lattice Ω. The structure of the covariance function will be given. Then, by letting n_1 and n_2 tend to infinity, the covariance function for the torus SGAR model on infinity lattice will be available. The covariance matrix of $\mathbf{E}$ can be given by

$$\mathbf{\Lambda} = \sigma_\tau^2\Big((\mathbf{I}-\mathbf{B}(\boldsymbol{\theta}))^T(\mathbf{I}-\mathbf{B}(\boldsymbol{\theta}))\Big)^{-1}.$$

Define the fourier vectors f_{pq}, $p = 1,\ldots,n_1$, $q = 1,\ldots,n_2$,

$$f_{pq} = (\, t_q^T,\, \phi_p t_q^T, \ldots, \phi_p^{n_1-1} t_q^T\,)^T,$$

$$t_q = (\,1,\, \varphi_q,\, \varphi_q^2, \ldots,\, \varphi_q^{n_2-1})^T,$$

$$\xi_p = exp(\,\sqrt{-1}\,2\pi p/n_1), \quad \varphi_q = exp(\,\sqrt{-1}\,2\pi q/n_2).$$

By Kashyap(1980), we can see that the eigenvectors of $\mathbf{\Lambda}$ are f_{pq}, $p = 1,\cdots,n_1$, $q = 1,\cdots,n_2$, and the eigenvalue associated with f_{pq} is $\rho^2|u_{pq}|^{-2}$,

$$u_{pq} = 1 - \sum_{\mathbf{t}\in\omega} \theta_{\mathbf{t}} \xi_p^{t_1} \varphi_q^{t_2},$$

where w is the neighborhood set defined in (2.2), say $w = \{(-1,0),(1,0),(0,-1),(0,1)\}$, and $\theta_{\mathbf{t}}$ is the coefficient for $\epsilon_{\mathbf{s}+\mathbf{t}}$ in (2.2). Therefore, $\mathbf{\Lambda}$ can be expressed as

$$\mathbf{\Lambda} = \frac{1}{n_1 n_2} \sum_{p=1}^{n_1} \sum_{q=1}^{n_2} f_{pq} (f_{pq}^*)^T \sigma_\tau^2 |u_{pq}|^{-2},$$

and the covariance between $\epsilon_{\mathbf{s}}$ and $\epsilon_{\mathbf{s}+(k_1,k_2)}$ can be given as

$$R(k_1, k_2) = \frac{1}{n_1 n_2} \sum_{p=1}^{n_1} \sum_{q=1}^{n_2} \xi_p^{k_1} \varphi_q^{k_2} \sigma_\tau^2 |u_{pq}|^{-2}. \tag{2.32}$$

Now let n_1 and n_2 tend to infinity, and write $z_1 = \xi_p$ and $z_2 = \varphi_q$, then

$$R(k_1, k_2) = \frac{-1}{4\pi^2} \int_{|z_1|=1} \int_{|z_2|=1} z_1^{k_1 - 1} z_2^{k_2 - 1} \sigma_\tau^2 \left| 1 - \sum_{\mathbf{t} \in \omega} \theta_{\mathbf{t}} z_1^{t_1} z_2^{t_2} \right|^{-2} dz_1 dz_2. \tag{2.33}$$

The integration above converges with (A3a), and $R(k, l)$ depends only on k and l. By the Laurent series, we have

$$\sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} z_1^{k_1} z_2^{k_2} R(k_1, k_2) = \sigma_\tau^2 \left| 1 - \sum_{\mathbf{t} \in \omega} \theta_{\mathbf{t}} z_1^{t_1} z_2^{t_2} \right|^{-2}. \tag{2.34}$$

By (A3a), the series converges for any $1 - C \leq |z_1|, |z_2| \leq 1 + C$, C is some positive constant. Simply let $z_1 = z_2 = 1$ in the last equation above, and note that $R(k_1, k_2)$ is always positive if both θ_1 and θ_2 are positive, the absolutely convergence of the series is obviously satisfied. So for any $\boldsymbol{\theta}$ satisfying (A3a), condition (2.30) hold. □

To show that $\epsilon_{\mathbf{s}}$ in model (2.3) and (2.4) satisfy condition (2.30), we first need to transfer them into the equivalent 2-dimensional moving average models.

Lemma 2.4 *Under assumption* (A3b)*, model (2.3) can be transferred as*

$$\epsilon_{(s_1, s_2)} = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \psi_{(k_1, k_2)} \tau_{(s_1 - k_1, s_2 - k_2)}, \tag{2.35}$$

where $\psi_{(k_1,k_2)} = \phi^{(1)}_{k_1}\phi^{(2)}_{k_2}$ *with* $\phi^{(i)}_u = \sum_{v=0}^{\infty} \theta_i^{|u|+2v}\binom{|u|+2v}{v}$, *i=1, 2.*

Remark 2.4 By some simple calculation, we can see that the coefficient $\psi_{\mathbf{k}}$ in (2.35) satisfies

$$|\psi_{\mathbf{k}}| \leq C\|\mathbf{k}\|^{-\alpha}, \tag{2.36}$$

for $\alpha = 3 + \delta$, δ is some positive value. Moreover, as mentioned in Hallin, Lu & Tran (2001), model (2.4) under assumption (A3c) can also be transformed to a moving average model with the same structure as (2.35) but distinct coefficient $\psi_{\mathbf{k}} = \binom{k_1+k_2}{k_1}\theta_1^{k_1}\theta_2^{k_2}$; $k_1 > 0, k_2 > 0$. Clearly, such coefficient $\psi_{\mathbf{k}}$ also satisfies condition (2.36).

Proof of Lemma 2.4: Let B_1 and B_2 be the transformation factors for $\epsilon_{\mathbf{s}}$, such as $B_1\epsilon_{(s_1,s_2)} = \epsilon_{(s_1-1,s_2)}$ and $B_2\epsilon_{(s_1,s_2)} = \epsilon_{(s_1,s_2-1)}$. Then (2.3) can be rewritten as

$$\epsilon_{(s_1,s_2)}\left(1-(B_1+B_1^{-1})\,\theta_1\right)\left(1-(B_2+B_2^{-1})\,\theta_2\right) = \tau_{(s_1,s_2)}\,. \tag{2.37}$$

To obtain the corresponding MA model, we may apply the power series expansion to $\left(1-(B_i+B_i^{-1})\,\theta_1\right)^{-1}$, $i = 1, 2$, separatively. If $|\theta_i| < 1/2$, when $1 - C \leq |B_i| \leq 1 + C$ for some $C > 0$, we have

$$\begin{aligned}
&\left(1-(B_i+B_i^{-1})\,\theta_1\right)^{-1} \\
&= \sum_{k=0}^{\infty} \theta_i^k \left(B_i + B_i^{-1}\right)^k \\
&= \sum_{k=0}^{\infty}\sum_{j=0}^{k} \theta_i^k\binom{k}{j} B_i^j B_i^{-(k-j)} \qquad (\textit{letting } p=j\,,\ q=k-j\,) \\
&= \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \theta_i^{p+q}\binom{p+q}{q} B_i^{p-q} I\{p \leq q\} \ + \ \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \theta_i^{p+q}\binom{p+q}{q} B_i^{p-q} I\{p > q\} \\
&\hat{=}\ I_1 + I_2.
\end{aligned} \tag{2.38}$$

Let $u = p - q$ and $v = q$, we then have

$$I_1 = \sum_{u=-\infty}^{0} \sum_{v=-u}^{\infty} \theta_i^{u+2v} \binom{u+2v}{v} B_i^u,$$

while letting $u = q - p$ and $v = p$, we have

$$I_2 = \sum_{u=-\infty}^{-1} \sum_{v=-u}^{\infty} \theta_i^{u+2v} \binom{u+2v}{v} B_i^{-u}.$$

Thus, (2.38) can be written as

$$\left(1-(B_i+B_i^{-1})\,\theta_i\right)^{-1} = \sum_{u=-\infty}^{\infty} \phi_u^{(i)} B_i^u, \tag{2.39}$$

where $\phi_u^{(i)} = \phi_{-u}^{(i)}$ with $\phi_u^{(i)} = \sum_{v=|u|}^{\infty} \theta_i^{-|u|+2v} \binom{-|u|+2v}{v}$, or, which can be further rewritten as

$$\phi_u^{(i)} = \sum_{v=0}^{\infty} \theta_i^{|u|+2v} \binom{|u|+2v}{v}. \tag{2.40}$$

By (2.37), (2.39) and (2.40), Lemma 2.4 follows. □

From Lemma 2.4, we can see that $\epsilon_{\mathbf{s}}$ has infinite dependency. For some m that $m \to \infty$ as $\mathbf{n} \to \infty$, we define a truncated MA model as

$$\epsilon_{\mathbf{s}}^m = \sum_{\substack{|k_i|<m \\ i=1,2}} \psi_{\mathbf{k}} \tau_{\mathbf{s}-\mathbf{k}} \quad \& \quad \tilde{\epsilon}_{\mathbf{s}} = \epsilon_{\mathbf{s}}^m + \eta_{\mathbf{s}},$$

where $\eta_{\mathbf{s}}$ is *i.i.d.* with zero mean and finite variance. $\{\eta_{\mathbf{s}}\}$ is independent of $\{\tau_{\mathbf{s}}\}$ and $\{\mathbf{X}_{\mathbf{s}}\}$. Moreover, $\eta_{\mathbf{s}} \backsim \epsilon_{\mathbf{s}} - \epsilon_{\mathbf{s}}^m$ where $\backsim$ denote equality in distribution. Hence, $\epsilon_{\mathbf{s}}$ and $\tilde{\epsilon}_{\mathbf{s}}$ have same distribution, and each pair of $\tilde{\epsilon}_{\mathbf{s}}$, say $\tilde{\epsilon}_{\mathbf{s}'}$ and $\tilde{\epsilon}_{\mathbf{s}''}$, are independent if $|s_i' - s_i''| \geq m$ for $i = 1, 2$. Now with Remark 2.4 and the process $\{\tilde{\epsilon}\}$, we can establish the following lemma, which ensures models (2.3) and (2.4) satisfy condition (2.30).

Lemma 2.5 *If $\epsilon_{\mathbf{s}}$ satisfies the 2-dimensional moving average model (2.35), and the coefficients $\psi_{\mathbf{k}}$ satisfying (2.36), then condition (2.30) is valid.*

Proof of Lemma 2.5: Define $\tilde{R}(\mathbf{k}) = Cov(\tilde{\epsilon}_{\mathbf{s}}, \tilde{\epsilon}_{\mathbf{s+k}})$, where $\mathbf{k} = (k_1, k_2)$ and $k_1 \geq 0, k_2 \geq 0$. We can obtain

$$\left|\tilde{R}(\mathbf{k})\right| \leq C\sigma_\tau^2 \sum_{t_1=k_1-m}^{m} \sum_{t_2=k_2-m}^{m} \left|\psi_{\mathbf{t}}\right| \left|\psi_{\mathbf{t-k}}\right| = C\sigma_\tau^2 \left(\sum_{t_1=k_1-m}^{m} \left|\phi_{t_1}^{(1)}\right| \left|\phi_{t_1-k_1}^{(1)}\right| \right) \left(\sum_{t_2=k_2-m}^{m} \left|\phi_{t_2}^{(2)}\right| \left|\phi_{t_2-k_2}^{(2)}\right| \right).$$

Note that $\phi_u^{(i)}$ is an even function. When $k_i \leq m$, by Remark 2.4 we have

$$\begin{aligned}
\sum_{t_i=k_i-m}^{m} \left|\phi_{t_i}^{(i)}\right| \left|\phi_{t_i-k_i}^{(i)}\right| &\leq C \sum_{t_1=0}^{m} \left|\phi_{t_i}^{(i)}\right| \left|\phi_{t_i-k_i}^{(i)}\right| \\
&\leq C \left(\sum_{t_i=k_i}^{m} \left|\phi_{t_i}^{(i)}\right| \left|\phi_{t_i-k_i}^{(i)}\right| + 2 \sum_{t_i=[k_i/2]+1}^{k_i} \left|\phi_{t_i}^{(i)}\right| \left|\phi_{t_i-k_i}^{(i)}\right| \right) \\
&\leq C \left(k_i^{-\alpha} \sum_{t_i=k_i}^{m} \left|\phi_{t_i-k_i}^{(i)}\right| + (k_i/2)^{-\alpha} \sum_{t_i=[k_i/2]+1}^{k_i} \left|\phi_{t_i-k_i}^{(i)}\right| \right),
\end{aligned}$$

which is bounded by $Ck_i^{-\alpha}$, $i = 1, 2$. When $m < k_1 \leq 2m$, we also have

$$\sum_{t_i=k_i-m}^{m} \left|\phi_{t_i}^{(i)}\right| \left|\phi_{t_i-k_i}^{(i)}\right| \leq C \sum_{t_i=[k_i/2]+1}^{m} \left|\phi_{t_i}^{(i)}\right| \left|\phi_{t_i-k_i}^{(i)}\right| \leq Ck_i^{-\alpha}. \tag{2.41}$$

$i = 1, 2$. Hence, $\left|\tilde{R}(\mathbf{k})\right|$ is bounded by $C(k_1k_2)^{-\alpha}$, an amount not dependent on m. Let m tend to infinity, we can then conclude that

$$\sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \left|R(\mathbf{k})\right| \leq C \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} (k_1k_2)^{-\alpha} < \infty. \qquad \square$$

At last we present a lemma relative to the asymptotic bias of $\hat{\boldsymbol{\beta}}$. Denote $m''(\mathbf{x})$ as the matrix of second derivatives at $\mathbf{x}$, and the (j,k)-th element by $\partial^2 m(\mathbf{x})/\partial x_j \partial x_k$. Let $\mathbf{H}(\mathbf{x})$ being a n dimensional vector with the i-th element being $H_i(\mathbf{x}) = tr[m''(\mathbf{x})(\mathbf{X}_{\mathbf{s}_i} - \mathbf{x})(\mathbf{X}_{\mathbf{s}_i} -$

$\mathbf{x})^T]$. Moreover, define $\mathbf{T_n} = \vec{\mathbf{X}}^T\mathbf{WH}$ (omitting $'(\mathbf{x})'$), and then we have

$$(\mathbf{T_n})_j = (nh_\pi)^{-1} \sum_{i=1}^{n} H_i(\mathbf{x}) \left(\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}_i} - \mathbf{x})\right)_j K_\pi \left(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_i} - \mathbf{x})\right), \tag{2.42}$$

for $i = 1, \cdots, d+1$.

Lemma 2.6 *Assume that* (A1)*,* (A2) *and* (A5) *hold, then*

$$E\left((\mathbf{T_n})_j\right) = f(\mathbf{x}) tr\left[m''(\mathbf{x}) \int \delta_{j-1}(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T) K_\pi(\boldsymbol{\delta}) d\boldsymbol{\delta}\right] + o(h_L^2), \tag{2.43}$$

$$Var\left((\mathbf{T_n})_j\right) = (nh_\pi)^{-1}\left\{f(\mathbf{x}) \int tr^2\left[m''(\mathbf{x})(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T)\right] \delta_{j-1}^2 K_\pi^2(\boldsymbol{\delta}) d\boldsymbol{\delta} + o(h_L^4)\right\}, \tag{2.44}$$

for $j = 1, \cdots, d+1$, $\delta_0 \equiv 1$.

Proof of Lemma 2.6: For $j = 1, \ldots, d+1$, we have

$$\begin{aligned}
E\left((\mathbf{T_n})_j\right) &= h_\pi^{-1} \int \mathrm{tr}\left[m''(\mathbf{x})(\mathbf{r}-\mathbf{x})(\mathbf{r}-\mathbf{x})^T\right] \left(\boldsymbol{\zeta}(\mathbf{r}-\mathbf{x})\right)_j K_\pi\left(\boldsymbol{\zeta}_1(\mathbf{r}-\mathbf{x})\right) f(\mathbf{r})\, d\mathbf{r} \\
&= \int \mathrm{tr}\left[m''(\mathbf{x})(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T)\right] \delta_{j-1} K_\pi(\boldsymbol{\delta})\, f(\boldsymbol{\delta} \odot \mathbf{h} + \mathbf{x})\, d\boldsymbol{\delta} \\
&= f(\mathbf{x})\, \mathrm{tr}\left[m''(\mathbf{x}) \int \delta_{i-1}(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T) K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta}\right] + o(h_L^2),
\end{aligned}$$

$$\begin{aligned}
Var\left((\mathbf{T_n})_j\right) &= (nh_\pi)^{-2} \sum_{i=1}^{n} Var\left[H_i(\mathbf{x}) \left(\boldsymbol{\zeta}(\mathbf{r}-\mathbf{x})\right)_j K_\pi\left(\boldsymbol{\zeta}_1(\mathbf{r}-\mathbf{x})\right)\right] \\
&\le (nh_\pi^2)^{-1} \int \mathrm{tr}^2\left[m''(\mathbf{x})(\mathbf{r}-\mathbf{x})(\mathbf{r}-\mathbf{x})^T\right] \left(\boldsymbol{\zeta}(\mathbf{r}-\mathbf{x})\right)_j^2 K_\pi^2(\boldsymbol{\zeta}_1(\mathbf{r}-\mathbf{x})) f(\mathbf{r})\, d\mathbf{r} \\
&= (nh_\pi)^{-1} \int \mathrm{tr}^2\left[m''(\mathbf{x})(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T)\right] \delta_{j-1}^2 K_\pi^2(\boldsymbol{\delta})\, f(\boldsymbol{\delta} \odot \mathbf{h} + \mathbf{x})\, d\boldsymbol{\delta} \\
&= (nh_\pi)^{-1} f(\mathbf{x}) \int \mathrm{tr}^2\left[m''(\mathbf{x})(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T)\right] \delta_{j-1}^2 K_\pi^2(\boldsymbol{\delta})\, d\boldsymbol{\delta} + o\left((nh_\pi h_L^{-4})^{-1}\right). \quad \square
\end{aligned}$$

Proof of Theorem 2.1: As the local linear fitting is always conducted in the neighborhood $\left\{\mathbf{X_s} : |(\mathbf{X_s} - \mathbf{x})_i| \le c_i h_i,\ i = 1, \ldots, d\right\}$, where $c_i's$ are some constants, we can expand

$m(\mathbf{X_s})$ as

$$m(\mathbf{X_s}) = m(\mathbf{x}) + (m'(\mathbf{x}))^T(\mathbf{X_s}-\mathbf{x}) + \frac{1}{2}(\mathbf{X_s}-\mathbf{x})^T m''(\mathbf{x})(\mathbf{X_s}-\mathbf{x}) + o_p(h_L^2)\ . \tag{2.45}$$

Rewrite (2.45) into the matrix form, we get

$$\mathbf{M} = \vec{\mathbf{X}}\boldsymbol{\beta} + \frac{1}{2}\mathbf{H} + o_p(h_L^2). \tag{2.46}$$

Substitute this expression into the following equation, then

$$\hat{\boldsymbol{\beta}} = \mathbf{U_n^{-1}}\left(\mathbf{V_n^*} + \vec{\mathbf{X}}^T\mathbf{WM}\right) = \boldsymbol{\beta} + \mathbf{U_n^{-1}V_n^*} + \frac{1}{2}\mathbf{U_n^{-1}}\vec{\mathbf{X}}^T\mathbf{W}\left(\mathbf{H} + o_p(h_L^2)\right). \tag{2.47}$$

By Lemma 2.1 and Lemma 2.2, we have $\mathbf{U_n^{-1}V_n^*} = O_p((nh_\pi)^{-1/2})$. By Lemma 2.6, we can see that

$$(\mathbf{T_n})_j = f(\mathbf{x})\,\mathrm{tr}\left[m''(\mathbf{x})\int \delta_{j-1}(\boldsymbol{\delta}\boldsymbol{\delta}^T)\odot(\mathbf{h}\mathbf{h}^T)K_\pi(\boldsymbol{\delta})d\boldsymbol{\delta}\right] + o_p(h_L^2), \tag{2.48}$$

where the first term on the right hand side of the equation is equal to $O(h_L^2)$. Hence, we have $\mathbf{U_n^{-1}}\vec{\mathbf{X}}^T\mathbf{WH} = O_p(h_L^2)$. It is also easy to show that $\vec{\mathbf{X}}^T\mathbf{W1} = O_p(1)$, where $\mathbf{1}$ is a vector with all elements being 1. Now let us turn back to (2.47),

$$h_L^{-2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = h_L^{-2}\left(O_p((nh_\pi)^{-1/2}) + O_p(h_L^2) + o_p(h_L^2)\right), \tag{2.49}$$

which is equal to $O_p(1)$ via assumption (A5). This concludes the proof. □

Next, we shall present the proof of Theorem 2.2. Recall (2.17) and the fact that $\mathbf{U_n} = \mathbf{U} + o_p(1)$, then Theorem 2.2 can be established by showing $(nh_\pi)^{1/2}\mathbf{V_n^*}$ is asymptotically

normal. By the Cramer-Wold device, this is equivalent for us to prove

$$A_{\mathbf{n}} = (nh_\pi)^{1/2}\mathbf{a}^T\mathbf{V}_{\mathbf{n}}^* = (nh_\pi)^{-1/2}\sum_{\mathbf{s}\in\Omega}\epsilon_{\mathbf{s}}K_{\mathbf{a}}\{\zeta_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x})\},$$

is asymptotically normal, where $\mathbf{a} = (a_0, \mathbf{a}_1^T)^T \in \mathbb{R}^{d+1}$ is an arbitrary constant vector, and $K_{\mathbf{a}}(\boldsymbol{\delta}) = (a_0 + \mathbf{a}_1^T\boldsymbol{\delta})K_\pi(\boldsymbol{\delta})$. Define

$$\tilde{A}_{\mathbf{n}} = (nh_\pi)^{-1/2}\sum_{\mathbf{s}\in\Omega}\tilde{\epsilon}_{\mathbf{s}}K_{\mathbf{a}}\{\zeta_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x})\}.$$

Then Theorem 2.2 can be established directly with (2.17) and the two lemmas below.

Lemma 2.7 *If $\epsilon_{\mathbf{s}}$ can be expressed in the form of (2.35) with coefficient $\psi_{\mathbf{k}}$ satisfying (2.36), and assumptions* (A1), (A3) *and* (A5) *hold, then*

$$|A_{\mathbf{n}} - \tilde{A}_{\mathbf{n}}| = o_p(1).$$

Proof of Lemma 2.7:

$$\begin{aligned}
\Pr\left[\,|A_{\mathbf{n}}-\tilde{A}_{\mathbf{n}}|>\gamma\,\right] &= \Pr\left[\,\left|\sum_{\mathbf{s}\in\Omega}(\epsilon_{\mathbf{s}}-\tilde{\epsilon}_{\mathbf{s}})\,K_{\mathbf{a}}(\zeta_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\right| > (nh_\pi)^{1/2}\gamma\right] \\
&\le \sum_{\mathbf{s}\in\Omega}\Pr\left[\,\left|(\epsilon_{\mathbf{s}}-\tilde{\epsilon}_{\mathbf{s}})\,K_{\mathbf{a}}(\zeta_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\right| > (n^{-1}h_\pi)^{1/2}\gamma\right] \\
&\le n^2h_\pi^{-1}\gamma^2\,\mathrm{Var}\Big[(\epsilon_{\mathbf{s}}-\tilde{\epsilon}_{\mathbf{s}})K_{\mathbf{a}}(\zeta_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\,\Big] \qquad (2.50)
\end{aligned}$$

Note that $\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}} = \sum_{\mathbf{k}\in\Gamma}\psi_{\mathbf{k}}\tau_{\mathbf{s}-\mathbf{k}} + \eta_{\mathbf{s}}$, where $\Gamma = \left\{\mathbf{k} = (k_1,k_2) : k_i \ge m \text{ for any } i = 1,2\right\}$, therefore $\{\mathbf{X}_{\mathbf{s}}\}$ is independent of $\{\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}}\}$. Then, (2.50) is bounded by $Cn^2h_\pi^{-1}E[(\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}})^2]E[K_{\mathbf{a}}^2(\zeta_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))]$. It is easy to show that

$$E\Big[\,K_{\mathbf{a}}^2(\zeta_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\,\Big] = (1+o(1))\,h_\pi f(\mathbf{x})\int K_{\mathbf{a}}^2(\boldsymbol{\delta})d\boldsymbol{\delta}.$$

For any $\mathbf{k} \in \Gamma$ we have $\|\mathbf{k}\| \geq m$, hence by (2.36)

$$E\Big[(\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}})^2\Big] \leq C\sigma_\tau^2 \sum_{\mathbf{k}\in\Omega} \psi_{\mathbf{k}}^2 \leq Cnm^{-2\alpha},$$

$\alpha = 3 + \delta$, $\delta > 0$. Now (2.50) is bounded by $Cn^3m^{-2\alpha}$. Let $m = [n(h_\pi h_L^4)^{c_1}]^{1/2}$ for some constant c_1, $0 < c_1 < \delta/(3+\delta)$. By assumption (A5), we have $m \to \infty$ and $n^3m^{-2\alpha} \to 0$ as $\mathbf{n} \to \infty$. This conclude the proof. □

Lemma 2.8 *If $\epsilon_{\mathbf{s}}$ can be expressed in the form of (2.35) with coefficient $\psi_{\mathbf{k}}$ satisfying (2.36), and assumptions* (A1), (A3), (A5) *and* (A7) *hold, then we have*

$$\tilde{A}_{\mathbf{n}} \xrightarrow{D} N\Big(0,\, \mathbf{a}^T\Sigma\mathbf{a}\Big),$$

where Σ is defined in Theorem 2.2.

Proof of Lemma 2.8: We shall establish the asymptotic normality of $\tilde{A}_{\mathbf{n}}$, then the proof can be completed with Lemma 2.7. Let (l_1, l_2) be the length of the large blocks, satisfying $n_i = r_i(2m + l_i)$, $l_i/m \to \infty$ as $\mathbf{n} \to \infty$, and r_i is some integer also tending to infinity. The consideration with r_i taking non-integer value is similar to that presented in the following, and hence is omitted. As we let $m = [n(h_\pi h_L^4)^{c_1}]^{1/2}$ in the proof of Lemma 2.7, we may further let $l_i = [n(h_\pi h_L^4)^{c_2}]^{1/2}$, where $0 < c_2 < c_1$ and c_2 can take different values with distinct $l_i, i = 1, 2$. Now denote $\mathbf{j} = (j_1, j_2)$ and

$$U(1, \mathbf{n}, \mathbf{j}) = \sum_{\substack{s_i = j_i(l_i+2m)+1 \\ i=1,2}}^{j_i(l_i+2m)+l_i} (nh_\pi)^{-1/2}\Delta_{\mathbf{s}},$$

where $\Delta_{\mathbf{s}} = \tilde{\epsilon}_{\mathbf{s}} K_{\mathbf{a}}(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}} - \mathbf{x}))$. Then $U(1, \mathbf{n}, \mathbf{j})$ is the sum of random variables over the $\mathbf{j}$-th large block. To sum up the random variables over all large blocks, we may define

$$T(\mathbf{n},1)=\sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} U(1,\mathbf{n},\mathbf{j}).$$

Also define that

$$U(2,\mathbf{n},\mathbf{j})=\sum_{s_1=j_1(l_1+2m)+1}^{j_1(l_1+2m)+l_1}\ \sum_{s_2=j_2(l_2+2m)+l_2+1}^{(j_2+1)(l_2+2m)}(nh_\pi)^{-1/2}\Delta_{\mathbf{s}},$$
$$U(3,\mathbf{n},\mathbf{j})=\sum_{s_1=j_1(l_1+2m)+l_1+1}^{(j_1+1)(l_1+2m)}\ \sum_{s_2=j_2(l_2+2m)+1}^{j_2(l_2+2m)+l_2}(nh_\pi)^{-1/2}\Delta_{\mathbf{s}},$$
$$U(4,\mathbf{n},\mathbf{j})=\sum_{\substack{s_i=j_i(l_i+2m)+l_i+1\\ i=1,2}}^{(j_i+1)(l_i+2m)}(nh_\pi)^{-1/2}\Delta_{\mathbf{s}},$$

which are the sum of random variables over different types of small blocks. Moreover, $T(\mathbf{n},k)$ for $2\le k\le 4$ can be defined similarly as $T(\mathbf{n},1)$. Hence we have $\tilde{A}_{\mathbf{n}}=\sum_{k=1}^{4}T(\mathbf{n},k)$. Similar to the proof of Lemma 2.2, we can show that $\mathrm{Var}(\tilde{A}_{\mathbf{n}})=\mathbf{a}^T\Sigma\mathbf{a}$ and $\sum_{\substack{\mathbf{s}',\mathbf{s}''\in\Omega\\ \mathbf{s}'\neq\mathbf{s}''}}\mathrm{Cov}(\Delta_{\mathbf{s}'},\Delta_{\mathbf{s}''})=O(n)$. To show Lemma 2.8, we may process to prove

(a1) $T(\mathbf{n},k)=o_p(1),$ for $2\le k\le 4,$

(a2) $\mathrm{Var}\,[T(\mathbf{n},1)]\doteq S_{\mathbf{n}}^2\to\mathbf{a}^T\Sigma\mathbf{a},$ as $\mathbf{n}\to\infty,$

(a3) $T(\mathbf{n},1)/S_{\mathbf{n}}\xrightarrow{D}N(0,\mathbf{1}).$

Proof of (a1): Without loss of generality, we only prove $T(\mathbf{n},2)=o_p(1)$, while the proofs for the rest are similar. It is sufficient for us to show the variance of $T(\mathbf{n},2)$ tends to zero. Define

$$S(2,\mathbf{j})=\{\,\mathbf{s}:\ j_1(l_1+2m)+1\le s_1\le j_1(l_1+2m)+l_1,\ j_2(l_2+2m)+l_2+1\le s_2\le(j_2+1)(l_2+2m)\,\},$$

which is the location set of the random variables in $U(2,\mathbf{n},\mathbf{j})$, and it contains $l_1 m$ elements . Notice that all $U(2,\mathbf{n},\mathbf{j})$ are independent, thus we have

$$\begin{aligned} Var[T(\mathbf{n},2)] &= \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} Var[U(2,\mathbf{n},\mathbf{j})] \\ &= (nh_\pi)^{-1}\sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}}\left(\sum_{\mathbf{s}\in S(2,\mathbf{j})} Var(\Delta_{\mathbf{s}}) + \sum_{\mathbf{s}',\mathbf{s}''\in S(2,\mathbf{j})} Cov(\Delta_{\mathbf{s}'},\Delta_{\mathbf{s}''})\right) \\ &\le (nh_\pi)^{-1}(\,2l_1 m\times r_1\times r_2)Var(\Delta_{\mathbf{s}}) + (nh_\pi)^{-1}\sum_{\substack{\mathbf{s}',\mathbf{s}''\in\Omega\\ \mathbf{s}'\neq\mathbf{s}''}} Cov(\Delta_{\mathbf{s}'},\Delta_{\mathbf{s}''}). \end{aligned}$$

The second term on the right hand side of the last inequality tends to zero, while the first term also tends to zero, due to the fact that $h_\pi^{-1}Var(\Delta_{\mathbf{s}})$ is bounded and $m/l_2\to 0$. (a1) then follows.

Proof of (a2): By the proof of (a1), we can see that the variance of $T(\mathbf{n},i), i=2,3,4$, tends to zero, so (a2) follows straightforwardly.

Proof of (a3): Notice that all $U(1,\mathbf{n},\mathbf{j})$ are independent. By the Linderberg central limit theorem, (a3) is held if we can show

$$\sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} E\left[(U(1,\mathbf{n},\mathbf{j}))^2 I\left\{|U(1,\mathbf{n},\mathbf{j})|>\epsilon S_{\mathbf{n}}\right\}\right]\longrightarrow 0, \tag{2.51}$$

for any $\epsilon>0$. As $\tilde{\epsilon}_{\mathbf{s}}$ may not be bounded, we define the truncated variables $\tilde{\epsilon}_{\mathbf{s}}^L=\tilde{\epsilon}_{\mathbf{s}}I\{\,|\tilde{\epsilon}_{\mathbf{s}}|\le L\}$, L is some positive constant. Similarly, define $\Delta_{\mathbf{s}}^L=\tilde{\epsilon}_{\mathbf{s}}^L K_{\mathbf{a}}(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))$, $U^L(1,\mathbf{n},\mathbf{j}) = \sum_{\mathbf{s}\in S(1,\mathbf{j})}(nh_\pi)^{-1/2}\Delta_{\mathbf{s}}^L$ and $T^L(\mathbf{n},1)=\sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} U^L(1,\mathbf{n},\mathbf{j})$, where $S(1,\mathbf{j})=\{\,\mathbf{s}: j_i(\,l_i+2m)+1\le s_i\le j_i(\,l_i+2m)+l_i, i=1,2\,\}$. Denote $S_{\mathbf{n}}^L$ as the asymptotic standard deviation of $T^L(\mathbf{n},1)$. Now we first show $T^L(\mathbf{n},1)\Big/S_{\mathbf{n}}^L\xrightarrow{D}N(\,0,1)$, or equivalently

$$J\triangleq\sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} E\left[(U^L(1,\mathbf{n},\mathbf{j}))^2 I\left\{|U^L(1,\mathbf{n},\mathbf{j})|>\epsilon S_{\mathbf{n}}^L\right\}\right]\longrightarrow 0\,, \tag{2.52}$$

and then generalize the result to (a3). As $|\tilde{\epsilon}^L_{\mathbf{s}}| < L$, we have

$$\begin{aligned} J &\le CL^2(nh_\pi)^{-1}l_1^2l_2^2 \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} Pr\Big[|U^L(1,\mathbf{n},\mathbf{j})| > \epsilon\, S^L_{\mathbf{n}}\Big] \\ &\le CL^2(nh_\pi)^{-1}l_1^2l_2^2 \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} (\epsilon\, S^L_{\mathbf{n}})^{-2} E\left[(U^L(1,\mathbf{n},\mathbf{j}))^2\right]. \end{aligned} \tag{2.53}$$

Examine the last expectation, we get

$$\begin{aligned} E\left[(U^L(1,\mathbf{n},\mathbf{j}))^2\right] &\le C(nh_\pi)^{-1}E\left\{\Big[\sum_{\mathbf{s}\in S(1,\mathbf{j})} K_{\mathbf{a}}(\,\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\,\Big]^2\right\} \\ &\le C(nh_\pi)^{-1}\Big\{\, l_1\, l_2\, E[K^2_{\mathbf{a}}(\,\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\,] + l_1^2l_2^2\, E^2[K_{\mathbf{a}}(\,\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}}-\mathbf{x}))\,]\Big\} \\ &\le C\Big(\, n^{-1}l_1l_2\,(1+h_\pi l_1l_2\,)\Big). \end{aligned}$$

Substitute this into J and rewrite n, l_1 and l_2 in terms of h_π and h_L, we then have

$$J \le C\Big((nh_\pi)^{-1}l_1^2l_2^2 + n^{-1}l_1^3l_2^3\Big) = C\Big((h_\pi^{c_2-1}h_L^{4c_2-2})^2 + (h_\pi h_L^4)^{3c_2-2}\Big). \tag{2.54}$$

Hence, for some c_2 which is close enough to 1, we can obtain $J \to 0$. Now we come to show (a3). Define $T^{L*}(\mathbf{n},1) = T(\mathbf{n},1) - T^L(\mathbf{n},1)$, then

$$\begin{aligned} &\Big\|E\left[\,\exp\left(i\lambda T(\mathbf{n},1)\right) - \exp\left(-\lambda^2 S^2_{\mathbf{n}}/2\right)\right]\Big\| \\ \le\ &\Big\|E\left[\exp\left(i\lambda T^L(\mathbf{n},1)\right) - \exp\left(-\lambda^2 (S^L_{\mathbf{n}})^2/2\right)\right]\Big\| + E\left\|\exp(i\lambda T^{L*}(\mathbf{n},1)) - 1\right\| \\ &+ \left|\exp(\,-\lambda^2(S^L_{\mathbf{n}})^2/2\,) - \exp(\,-\lambda^2 S^2_{\mathbf{n}}/2\,)\right| \\ \hat{=}\ & E_1 + E_2 + E_3. \end{aligned}$$

E_1 tends to zero by (2.54), while E_3 also tends to zero by dominated convergence theorem. It remains for us to show E_2 converges, and this is held if $\mathrm{Var}\,[T^{L*}(\mathbf{n},1)] \to 0$. We

can show that

$$\lim_{\mathbf{n}\to\infty} \mathrm{Var}\,[T^{L*}(\mathbf{n},1)] = \mathrm{Var}\Big(\tilde{\epsilon}_{\mathbf{s}} I\{|\tilde{\epsilon}_{\mathbf{s}}| > L\}\Big) f(\mathbf{x}) \int K_{\mathbf{a}}^2(\boldsymbol{\delta})\, d\boldsymbol{\delta}\,,$$

which converges to zero as $L \to \infty$ by dominated convergence theorem. The proof is completed. □

Proof of Theorem 2.3: Note that $\hat{\mathbf{P}}$ in (2.13) can also be expressed as $\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{B}(\hat{\boldsymbol{\theta}}))\hat{\mathbf{E}} + \hat{\mathbf{M}}$. Substitute $\hat{\boldsymbol{\theta}}$ by $\boldsymbol{\theta}$, we form another process as $\check{\mathbf{P}} = (\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\hat{\mathbf{E}} + \hat{\mathbf{M}}$. Furthermore, define $\check{m}(\mathbf{x}) = \eta^T \mathbf{U}_{\mathbf{n}}^{-1} \vec{\mathbf{X}}^T \mathbf{W} \check{\mathbf{P}}$, then we have $\tilde{m}(\mathbf{x})$–$m(\mathbf{x}) = \Big(\tilde{m}(\mathbf{x})$–$\check{m}(\mathbf{x})\Big) + \Big(\check{m}(\mathbf{x})$–$m(\mathbf{x})\Big)$. To establish the theorem, we should show that $\tilde{m}(\mathbf{x})$–$\check{m}(\mathbf{x}) = o_p((ng_\pi)^{-1/2})$ and $(ng_\pi)^{1/2}\Big(\check{m}(\mathbf{x})$–$m(\mathbf{x})\Big)$ is asymptotically normal.

We consider the former term first. By Theorem 2.1, we can see that $\hat{\mathbf{E}} = \mathbf{E} + O_p(h_L^2)$. By the symmetry property of $K(.)$ and Lemma 2.1, we have

$$\mathbf{U}_{\mathbf{n}} = f(\mathbf{x}) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \int \boldsymbol{\delta}\boldsymbol{\delta}^T K_\pi(\boldsymbol{\delta}) d\boldsymbol{\delta} \end{pmatrix} + o_p(1). \tag{2.55}$$

Hence, $\eta^T \mathbf{U}_{\mathbf{n}}^{-1} = \eta^T (f^{-1}(\mathbf{x}) + o_p(1))$. Define $\mathbf{G}$ as a vector with the i-th element being $G_i = \sum_{j \in N(\mathbf{s}_i)} \epsilon_{\mathbf{s}_j}$, where $N(\mathbf{s}_i)$ is the neighborhood set of $\epsilon_{\mathbf{s}_i}$. Also note that there are finite nonzero elements in each row of $B(\boldsymbol{\theta})$ and $B(\hat{\boldsymbol{\theta}})$. Therefore, we have

$$\begin{aligned}
\tilde{m}(\mathbf{x}) - \check{m}(\mathbf{x}) &= \eta^T \mathbf{U}_{\mathbf{n}}^{-1} \vec{\mathbf{X}}^T \mathbf{W}\Big(\mathbf{B}(\boldsymbol{\theta}) - \mathbf{B}(\hat{\boldsymbol{\theta}})\Big)\hat{\mathbf{E}} \\
&= \eta^T \Big(f^{-1}(\mathbf{x}) + o_p(1)\Big) \vec{\mathbf{X}} \mathbf{W} \Big(G + O_p(h_L^2)\Big) \cdot o_p(1) \\
&= \eta^T f^{-1}(\mathbf{x})\, \vec{\mathbf{X}}^T \mathbf{W} \mathbf{G} \cdot o_p(1) + \eta^T f^{-1}(\mathbf{x})\, \vec{\mathbf{X}}^T \mathbf{W} \mathbf{1} \cdot o_p(h_L^2) \\
&\hat{=} I_{11} \cdot o_p(1) + I_{12} \cdot o_p(h_L^2).
\end{aligned} \tag{2.56}$$

It is easy to show that $I_{12} = O_p(1)$. For I_{11}, we have

$$I_{11} = (ng_\pi)^{-1} f^{-1}(\mathbf{x}) \sum_{\mathbf{s}\in\Omega} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))G_\mathbf{s} = (ng_\pi)^{-1} f^{-1}(\mathbf{x}) \sum_{\mathbf{s}\in\Omega} \sum_{\mathbf{t}\in N(\mathbf{s})} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))\,\epsilon_\mathbf{t}\,. \quad (2.57)$$

It is clear that I_{11} has zero mean, and its variance can be given as

$$\begin{aligned}
\mathrm{Var}\,(I_{11}) \;=\; & (ng_\pi)^{-2} f^{-2}(\mathbf{x}) \sum_{\mathbf{s}\in\Omega} \sum_{\mathbf{t}'\in N(\mathbf{s})} \sum_{\mathbf{t}''\in N(\mathbf{s})} \mathrm{Cov}\Big[\, K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))\epsilon_{\mathbf{t}'},\; K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))\epsilon_{\mathbf{t}''}\Big] \\
& + (ng_\pi)^{-2} f^{-2}(\mathbf{x}) \sum_{\substack{\mathbf{s}'\neq\mathbf{s}'' \\ \mathbf{s}',\mathbf{s}''\in\Omega}} \sum_{\mathbf{t}'\in N(\mathbf{s}')} \sum_{\mathbf{t}''\in N(\mathbf{s}'')} \mathrm{Cov}\Big[\, K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_{s'}}-\mathbf{x}))\epsilon_{\mathbf{t}'},\; K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_{s''}}-\mathbf{x}))\epsilon_{\mathbf{t}''}\Big] \\
\hat{=}\; & I\!V_1 + I\!V_2. \qquad (2.58)
\end{aligned}$$

As there are finite elements in each $N(\mathbf{s})$ set, we can obtain that

$$I\!V_1 \le C(ng_\pi)^{-1} f^{-1}(\mathbf{x})\, \sigma_\epsilon^2 \int K_\pi^2(\boldsymbol{\delta})\, d\boldsymbol{\delta}, \qquad (2.59)$$

$$I\!V_2 \le Cn^{-1}\sigma_\epsilon^2 \;+\; Cn^{-1} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \Big| R(k_1,k_2) \Big|. \qquad (2.60)$$

Therefore, via (2.58)-(2.60), we have $I_{11} = O_p\Big((ng_\pi)^{-1/2}\Big)$. Turn back to (2.56), it then can be seen that $\tilde{m}(\mathbf{x}) - \check{m}(\mathbf{x}) = o_p\Big((ng_\pi)^{-1/2}\Big) + o_p(h_L^2)$, which is equal to $o_p((ng_\pi)^{-1/2})$ by (A6). Next, we will establish the asymptotic normality of $\check{m}(\mathbf{x})$. Define $\mathbf{F} = \Big(m(\mathbf{x}), (m'(\mathbf{x}) \odot \mathbf{g})^T\Big)^T$. Since $\eta^T \mathbf{U}_\mathbf{n}^{-1} \vec{\mathbf{X}}^T \mathbf{W} \vec{\mathbf{X}} = (1, \mathbf{0})$, we can obtain that

$$\begin{aligned}
& \check{m}(\mathbf{x}) - m(\mathbf{x}) \\
=\; & \eta^T \mathbf{U}_\mathbf{n}^{-1} \vec{\mathbf{X}} \mathbf{W} (\check{P} - \vec{\mathbf{X}}\mathbf{F}) \\
=\; & (ng_\pi)^{-1} f^{-1}(\mathbf{x})(1+o_p(1)) \sum_{\mathbf{s}\in\Omega} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))\Big[\, \frac{1}{2}(\mathbf{X_s}-\mathbf{x})^T m''(\mathbf{x})(\mathbf{X_s}-\mathbf{x}) \\
& + \tau_\mathbf{s} + O_p(h_L^2) + o_p(g_L^2)\Big] \\
=\; & \Big(I_{21} + I_{22} + I_{23}\,(O_p(h_L^2) + o_p(g_L^2))\Big)(1 + o_p(1)), \qquad (2.61)
\end{aligned}$$

where

$$I_{21} = (ng_\pi)^{-1} f^{-1}(\mathbf{x}) \sum_{\mathbf{s}\in\Omega} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))\tau_\mathbf{s}, \tag{2.62}$$

$$I_{22} = (ng_\pi)^{-1} f^{-1}(\mathbf{x}) \sum_{\mathbf{s}\in\Omega} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))H_\mathbf{s}(\mathbf{x}), \tag{2.63}$$

$$I_{23} = (ng_\pi)^{-1} f^{-1}(\mathbf{x}) \sum_{\mathbf{s}\in\Omega} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x})). \tag{2.64}$$

It is clear that $I_{23} = O_p(1)$, and similar to Lemma 2.6, we can show

$$I_{22} = \mathrm{tr}\left[m''(\mathbf{x})\int(\boldsymbol{\delta}\boldsymbol{\delta}^T)\odot(\mathbf{g}\mathbf{g}^T)\,K_\pi(\boldsymbol{\delta})\,d\boldsymbol{\delta}\right] + o_p((ng_\pi)^{-1/2}). \tag{2.65}$$

Hence, via (A6) and (2.61), we have

$$((ng_\pi)^{1/2})\Big(\check{m}(\mathbf{x})-m(\mathbf{x})-\mathrm{Bias}(\mathbf{x})\Big) = (ng_\pi)^{-1/2} f^{-1}(\mathbf{x})\sum_{\mathbf{s}\in\Omega} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X_s}-\mathbf{x}))\tau_\mathbf{s}(1+o_p(1))+o_p(1). \tag{2.66}$$

It remains for us to show the first term on the right hand side of the equation above is asymptotically normal with zero mean and variance σ^2. Note that both $\{\mathbf{X_s}\}$ and $\{\tau_\mathbf{s}\}$ are *i.i.d.*, and mutually independent, therefore, the asymptotic normality is obviously satisfied. The proof is completed.

2.7 Summary and Remark

In this chapter, we focus on the nonparametric model with SGAR-type errors. For our study, we assume that residual $\epsilon_\mathbf{s}$ satisfies the torus, separable and unilateral SGAR model. The two-step estimation in Martins-Filho & Yao (2009) is adopted, which firstly forms a new process with the same conditional mean as the original one but *i.i.d.* errors, and then apply the local linear fitting to the new process to achieve an improved estima-

tor of the regression function. For the first-step estimator, we show its convergence rate with the three types of error structures, and establish the asymptotic normality when ϵ_s satisfies the separable and unilateral SGAR model. For the second-step estimator, the asymptotic normality with all three error structures is established. For practice, we propose an iterative method to determine the bandwidth for the first- and second-step estimation. Some technique for the estimation of the SGAR coefficients is also considered. Simulations are conducted, showing that our estimation performs better than the local linear fitting when spatially correlated errors exist.

Chapter 3

Additive Model with SGAR-type Error

3.1 Literature Review

The nonparametric additive models were first introduced in the statistical literature in the early 1980's, and have drawn a great deal of attentions since then. Comparing with the general nonparametric models, the additive models are charming since it naturally generalize the linear regression models and allow interpretation of marginal changes, such as the effect of single explanatory variable on the conditional mean of the regression function. Moreover, it was shown in Stone (1985, 1986) that each additive component in the model can be estimated with the one-dimensional rate of convergence. Therefore, the additive models avoid the so-called 'curse of dimensionality' that affects the general nonparametric models seriously. In the early development of additive models, the estimation of the additive components are based on the back-fitting technique, whose main idea is to project the data onto the space of additive functions via the least squares method. Some iterative algorithms must be used in this procedure, and although calculations usually converge quickly, no closed form of the estimators are available. This

makes the analysis of asymptotic properties very difficult. No breakthrough on this topic had been made until late 1990's, see the work of Opsomer & Ruppert (1997) and Mammen et al. (1999). Lu et al. (2007) further generalized the analysis to the spatial problems. However, many problems are still unclear, such as how to estimate the derivatives of the function, how to identify the significant regressors, among others.

To get round the difficulty of the iterative procedure, Linton & Nielsen (1994) propose another method to estimate the additive component based on marginal integration of the regression function. The key point of the marginal integration approach is to estimate the whole surface of the regression function by local polynomial fitting first, and then take marginal average of the estimators of the regression function to obtain the estimators of the additive components. As this method provides the closed form of the estimators, the asymptotic properties can be investigated easily, and hence it attracts many researches. Some important literature will be mentioned here. Chen & Härdle (1995) considered the same method with the Nadaraya-watson estimator. Severance-Lossin & Sperlich (1999) extended to estimate both the regression function and its derivatives simultaneously with the local polynomial fitting. Cai & Fan (2000) considered the problem with dependent data, different from the *i.i.d*. assumption of observations used in the above paper. Some weights were also used when averaging the regression surface in their research. Fan & Li (2003) considered an additive partially linear model, which was further generalized to the spatial case in Gao et al.(2006).

In this chapter, we will consider the additive model with spatially correlated errors. Few attention has ever been paid on this problem before. The marginal integration procedure and the two-step estimation in Martins-Filho & Yao (2009) will be adopted. Note that

each additive function can contain one or multiple variables, and in both cases the additive model considered here can be seem as a special case of the general nonparametric model considered in Chapter 2. Therefore, some results obtained in Chapter 2 can be referred in the following analysis.

The outline of this chapter is as follow. In section 2, we will present the model as well as the estimation procedure. In section 3, some theoretical results of the estimators are given. Simulations will be shown in section 4 to assess the performance of our estimation, together with a cross comparison between the fitting with the methods considered in Chapter 2 and this chapter under the situation that some additive structure of the regression function really exists. Proofs will be presented in section 5, and a brief conclusion is given in the last section.

3.2 Model and Estimation

Consider the model

$$Y_{\mathbf{s}} = m(\mathbf{X}_{\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{3.1}$$

where $m(\mathbf{X}_{\mathbf{s}})$ has some additive structure as

$$m(\mathbf{X}_{\mathbf{s}}) = \mu + \sum_{i=1}^{d} m_i(\mathbf{X}_{i\mathbf{s}}), \tag{3.2}$$

and $\epsilon_{\mathbf{s}}$ satisfying the SGAR models defined in (2.2)-(2.4). Here, $\mathbf{s} \in Z^2$ is the index of position from a $n_1 \times n_2$ rectangular lattice Ω, μ is some constant, $Y_{\mathbf{s}}$ is the response variable, $\mathbf{X}_{\mathbf{s}} = (\mathbf{X}_{1\mathbf{s}}^T, \ldots, \mathbf{X}_{d\mathbf{s}}^T)^T$ is the explanatory vector with $\mathbf{X}_{i\mathbf{s}}$ being b_i- dimensional, and $m_i(.)$, $i = 1, \ldots, d$, are some unknown functions. We denote the joint density of $\mathbf{X}_{\mathbf{s}}$

by $f(\mathbf{x})$, $\mathbf{x} = (\mathbf{x}_1^T, \cdots, \mathbf{x}_d^T)^T$, and the marginal density of the i-th random vector by $f_i(\mathbf{x}_i)$. Moreover, we assumee $m_i(\mathbf{X}_{i\mathbf{s}})$ has zero expectation, say $\int m_i(\mathbf{x}_i)\, f_i(\mathbf{x}_i)\, d\mathbf{x}_i = 0$.

Without taking the spatial dependency of $\epsilon_{\mathbf{s}}$ into account, the estimation of $m_k(.)$ can be conducted with the marginal integration technique and local linear fitting as follows. For some fixed vector $\mathbf{x}_k$, define $\mathbf{X}_{\mathbf{s}}^{(k)} = (\mathbf{X}_{1\mathbf{s}}^T, \cdots, \mathbf{X}_{(k-1)\mathbf{s}}^T, \mathbf{x}_k^T, \mathbf{X}_{(k+1)\mathbf{s}}^T, \cdots, \mathbf{X}_{d\mathbf{s}}^T)^T$. Obviously, $\mathbf{X}_{\mathbf{s}}^{(k)}$ is a vector specified by $\mathbf{x}_{\mathbf{s}}$, but we use the notation $\mathbf{X}_{\mathbf{s}}^{(k)}$ only when no confusion occurs. Furthermore, throughout this chapter, we always use k for the fixed index of the random vector while other indices for the arbitrary ones. We also defined $\mathbf{X}_{\mathbf{s}}^{(-k)} = (\mathbf{X}_{1\mathbf{s}}^T, \cdots, \mathbf{X}_{(k-1)\mathbf{s}}^T, \mathbf{X}_{(k+1)\mathbf{s}}^T, \cdots, \mathbf{X}_{d\mathbf{s}}^T)^T$ with density denoted by $f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})$. From (3.2) we can see that $E\Big(m(\mathbf{X}_{\mathbf{s}}^{(k)})\Big) = \mu + m_k(\mathbf{x}_k)$. Therefore, the estimator of $m_k(\mathbf{x}_k)$ is given by

$$\hat{m}_k(\mathbf{x}_k) = \frac{1}{n} \sum_{\mathbf{s}\in\Omega} \hat{m}(\mathbf{X}_{\mathbf{s}}^{(k)}) - \overline{Y}, \tag{3.3}$$

where $n = n_1 \times n_2$, $\overline{Y}$ is the empirical mean of $Y_{\mathbf{s}}$, and $\hat{m}(\mathbf{X}_{\mathbf{s}}^{(k)})$ is the local linear estimator of $m(\mathbf{X}_{\mathbf{s}}^{(k)})$.

To show the expression of $\hat{m}(\mathbf{X}_{\mathbf{s}}^{(k)})$, more notations should be introduced first. Arrange the position $\mathbf{s} \in \Omega$ in arbitrary order, and denote them by $\mathbf{s}_1, \cdots, \mathbf{s}_n$ respectively. Let $D = \sum_{i=1}^{d} b_i$, $\mathbf{h}_i = (h_{i1}, \cdots, h_{ib_i})^T$ be some bandwidth vector and $\mathbf{h} = (\mathbf{h}_1^T, \cdots, \mathbf{h}_d^T)^T$. Define $h_{\pi i} = \prod_{j=1}^{b_i} h_{ij}$, $h_\pi = \prod_{i=1}^{d} h_{\pi i}$ and $K_\pi(\boldsymbol{\delta}) = \prod K_i(\delta_i)$, where $\boldsymbol{\delta}$ is some vector with the i-th element being δ_i and $K_i(.)$ is some kernel function. Moreover, $\boldsymbol{\zeta}_1(\boldsymbol{\delta}) = \Big(\boldsymbol{\delta} \odot (1/\mathbf{h}_\delta)\Big)^T$, where $\odot$ denotes the element-wise product of two matrix or vectors with same size. Especially, we will use $\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}} - \mathbf{x}_{\mathbf{s}}) = \Big((\mathbf{X}_{\mathbf{s}} - \mathbf{x}_{\mathbf{s}}) \odot (1/\mathbf{h})\Big)^T$ frequently below.

Now the local linear estimators $\hat{m}_k(\mathbf{X}_{\mathbf{s}}^{(k)})$ can be expressed as

$$\hat{m}(\mathbf{X}_{\mathbf{s}}^{(k)}) = \eta^T \Big[\vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)})\, \vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\Big]^{-1} \vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)})\, \mathbf{Y}, \tag{3.4}$$

where η is a vector with the first element valued 1 and 0 elsewhere, $\vec{\mathbf{X}}(\mathbf{X}_{\mathbf{s}}^{(k)})$ is a $n\times(D+1)$ matrix function concerning the vector $\mathbf{X}_{\mathbf{s}}^{(k)}$, with the i-th row being $\Big(1, \boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_i} - \mathbf{X}_{\mathbf{s}}^{(k)})\Big)$, $W(\mathbf{X}_{\mathbf{s}}^{(k)}) = (nh_\pi)^{-1} diag\Big(K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_1} - \mathbf{X}_{\mathbf{s}}^{(k)})), \ldots, K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_n} - \mathbf{X}_{\mathbf{s}}^{(k)})) \Big)$ and $\mathbf{Y} = (Y_{\mathbf{s}_1}, \cdots, Y_{\mathbf{s}_n})^T$. Some researches have studied estimators similar to $\hat{m}_k(\mathbf{x}_k)$ given in (3.3), for example, Severance-Lossin & Sperlich (1999) and Cai & Fan (2000). However, few literature considered the case where the errors are spatially correlated. In the following, we first rewrite models (3.1)-(3.2) into the matrix form as

$$\mathbf{Y} = \mu + \sum_{i=1}^{d} \mathbf{M}_i + \mathbf{E}, \tag{3.5}$$

where $\mathbf{M}_i = (m_i(\mathbf{X}_{i\mathbf{s}_1}), \cdots, m_i(\mathbf{X}_{i\mathbf{s}_n}))^T$, $i = 1, \cdots, d$, and $\mathbf{E} = (\epsilon_{\mathbf{s}_1}, \cdots, \epsilon_{\mathbf{s}_n})^T$. We consider a new process similar to $\{Y_{\mathbf{s}}\}$ but with $\mathbf{E}$ in (3.5) substituted by $(\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\mathbf{E}$, where $\mathbf{B}(\boldsymbol{\theta})$ is defined in Chapter 2. That is

$$\mathbf{P} = \Big(\mathbf{I} - \mathbf{B}(\boldsymbol{\theta})\Big)\mathbf{Y} + \mathbf{B}(\boldsymbol{\theta}) \Big(\mu + \sum_{i=1}^{d} \mathbf{M}_i\Big), \tag{3.6}$$

$\mathbf{P} = (P_{\mathbf{s}_1}, \cdots, P_{\mathbf{s}_n})^T$. Then $\mathbf{P}$ has the same conditional mean as $\mathbf{Y}$ does, but with *i.i.d.* error term. Hence, by substitute $\mathbf{Y}$ and $Y_{\mathbf{s}}$ by $\mathbf{P}$ and $P_{\mathbf{s}}$ in (3.3)-(3.4), we can obtain an improved estimator of $m_k(\mathbf{x}_k)$. However, $\mathbf{B}(\boldsymbol{\theta})$ and $\mathbf{M}_i$, $i = 1, \cdots, d$, in (3.6) are all unknown, so some estimators of them should be used instead. Then, we may found the new process as

$$\hat{\mathbf{P}} = \Big(\mathbf{I} - \mathbf{B}(\hat{\boldsymbol{\theta}})\Big)\mathbf{Y} + \mathbf{B}(\hat{\boldsymbol{\theta}}) \Big(\hat{\mu} + \sum_{i=1}^{d} \hat{\mathbf{M}}_i\Big), \tag{3.7}$$

where $\hat{\mu} = \overline{Y}$, $\hat{\mathbf{M}}_i$, $i = 1, \cdots, d$, are obtained by (3.3) and $\hat{\mathbf{B}}(\boldsymbol{\theta})$ are obtained by some general SGAR estimation methods. Let $\mathbf{g}$ be a bandwidth vector distinct from $\mathbf{h}$ as $\mathbf{g} = (\mathbf{g}_1^T, \cdots, \mathbf{g}_d^T)^T$, $\mathbf{g}_i = (g_{i1}, \cdots, g_{ib_i})^T$, and consequently $g_{\pi i} = \prod_{j=1}^{b_i} g_{ij}$ and $g_\pi = \prod_{i=1}^{d} g_{\pi i}$.

Substitute $\mathbf{h}$ and h_π by $\mathbf{g}$ and g_π in the expression of $\vec{\mathbf{X}}(\mathbf{X}_\mathbf{s}^{(k)})$ and $\mathbf{W}(\mathbf{X}_\mathbf{s}^{(k)})$ below, then an improved estimator of $m_k(\mathbf{x}_k)$ can be given as

$$\tilde{m}_k(\mathbf{x}_k) = \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\tilde{m}(\mathbf{X}_\mathbf{s}^{(k)}) - \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\hat{P}_\mathbf{s}\,, \tag{3.8}$$

$$\tilde{m}(\mathbf{X}_\mathbf{s}^{(k)}) = \eta\Big[\vec{\mathbf{X}}^T(\mathbf{X}_\mathbf{s}^{(k)})\,W(\mathbf{X}_\mathbf{s}^{(k)})\,\vec{\mathbf{X}}(\mathbf{X}_\mathbf{s}^{(k)})\Big]^{-1}\vec{\mathbf{X}}^T(\mathbf{X}_\mathbf{s}^{(k)})\,W(\mathbf{X}_\mathbf{s}^{(k)})\,\hat{\mathbf{P}}\,. \tag{3.9}$$

We sum up the full estimation procedure below, which we denote as ADD-SCE through the thesis.

Step 1. Obtain an initial estimator of $m_k(.), k = 1, \ldots, d$, by (3.3) with bandwidth $\mathbf{h}$. Consequently, the estimates of $\epsilon_\mathbf{s}$ are available.

Step 2. Estimate θ_1 and θ_2 with the estimates of $\epsilon_\mathbf{s}$, so that $\hat{\mathbf{P}}$ in (3.7) can be formed. Then, with bandwidth $\mathbf{g}$, we can obtain the improved estimator of $m_k(.)$ by (3.8).

3.3 Theoretical Results

In this section, we will examine the asymptotical property of $\hat{m}_k(\mathbf{x}_k)$ and $\tilde{m}_k(\mathbf{x}_k)$ given in (3.3) and (3.8). The following assumptions are needed in the analysis.

(B1) The D-dimensional random vectors $\mathbf{X}_\mathbf{s}, \mathbf{s} \in \Omega$, are *i.i.d.* with joint density $f(\mathbf{x})$, $0 < |f(\mathbf{x})| < \infty$. Moreover, $\{\mathbf{X}_\mathbf{s}\}$ is independent of $\{\epsilon_\mathbf{s}\}$ and $\{\tau_\mathbf{s}\}$, where $\tau_\mathbf{s}$ is defined in (2.2)-(2.4).

(B2) The marginal density $f_{-k}(\mathbf{x})$ is uniformly bounded.

(B3) All the second derivatives of $m(.)$ exist and are continuous at any $\mathbf{x}$, $\mathbf{x}$ is some D-dimensional vector; for $j = 1, \cdots, d$, $m_j(\mathbf{x}_j)$ has zero mean and bounded second moment.

(B4) The random field $\{\epsilon_{\mathbf{s}}, \mathbf{s} \in \mathbb{Z}^2\}$ is strictly stationary; $\epsilon_{\mathbf{s}}$ has zero mean and finite variance.

(B4a) The coefficients in (2.2) satisfy $|\theta_1| + |\theta_2| < 1/2$.

(B4b) The coefficients in (2.3) satisfy $|\theta_1| < 1/2$, $|\theta_2| < 1/2$.

(B4c) The coefficients in (2.4) satisfy $|\theta_1| + |\theta_2| < 1$.

(B5) The kernel function $K(.)$ is symmetric, with bounded support, and Lipschitz continuous.

(B6) There exist some sequences such that: $l_1 \to \infty$, $l_2 \to \infty$ and $m \to \infty$ as $\mathbf{n} \to \infty$, $\mathbf{n} = (n_1, n_2)$; $m/l_i \to 0$ and $l_i/n_i \to 0$ for $i = 1, 2$, as $\mathbf{n} \to \infty$.

For the bandwidth $\mathbf{h}$, some assumptions should be made. Note that we need to estimate d unknown functions $m_i(.)$, $i = 1, \cdots, d$, and the correlated bandwidths are denoted by $\mathbf{h}_i$ respectively. When estimating $m_k(.)$, we will assume that some element in $\mathbf{h}_k$, denoted by h_{kl}, tends to zero slower than any other h_{ij}, $i = 1, \cdots, d$, $j = 1, \cdots, b_i$. Therefore, for each k=1, $\cdots, d$, we need a separate set of bandwidths $\mathbf{h}^{(k)} = (\mathbf{h}_1^{(k)}, \cdots, \mathbf{h}_d^{(k)})$ that satisfy $h_{kl}^{(k)}$ has the slowest convergent rate in $\mathbf{h}^{(k)}$. Such idea also appears in Gao et al.(2006), who considered a partially linear model with additive nonparametric components and with non *i.i.d.* explanatory variables. For the sake of simplifying notations, we omit the superscript (k) in the following analysis whenever no confusion may occur.

(B7) All $h_{ij} > 0$, $h_{ij} \to 0$ and $nh_\pi \to \infty$ as $\mathbf{n} \to \infty$; moreover, when estimating $m_k(.)$, $\mathbf{h}_k$ tends to zero in the manner as $n = O\left((h_{\pi k} h_{k\iota}^4)^{-1}\right)$, where $h_{k\iota}$ is an element in $\mathbf{h}_k$ satisfying that $h_{ij}/h_{k\iota} \to 0$ as $\mathbf{n} \to \infty$ for all $h_{ij} \neq h_{k\iota}$.

We first consider the asymptotic properties of the first-step estimator $\hat{m}_k(\mathbf{x}_k)$ in (3.3). The following theorem gives the convergent rate of $\hat{m}_k(\mathbf{x}_k)$.

Theorem 3.1 *When assumptions* (B1)-(B5), (B7) *hold, and* $\epsilon_{\mathbf{s}}$ *satisfies one of the model defined in (2.2)-(2.4) and the corresponding assumptions* (B4a)-(B4c)*, then we have*

$$\hat{m}_k(\mathbf{x}_k) - m_k(\mathbf{x}_k) - \text{Bias}(\mathbf{x}_k, \mathbf{h}_k) = (nh_{\pi k})^{-1} \sum_{\mathbf{s}\in\Omega} f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)}) f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\epsilon_{\mathbf{s}} + o_p(h_{k\iota}^2)\,, \tag{3.10}$$

where $\text{Bias}(\mathbf{x}_k, \mathbf{h}_k) = \frac{1}{2}\, tr[m_k''(\mathbf{x}_k) \int K_\pi(\boldsymbol{\delta})(\delta_k \delta_k^T) \odot (\mathbf{h}_k \mathbf{h}_k^T)\, d\boldsymbol{\delta}]$, $\boldsymbol{\delta} = (\delta_1^T, \cdots, \delta_d^T)^T$, δ_i *is some* b_i*-dimensional random vector*. *Obviously,* $\text{Bias}(\mathbf{x}_k) = O(h_{k\iota}^2)$*, and it can be seen that the first term on the right hand side of (3.10) is equal to* $O_p\left((nh_{\pi k})^{-1/2}\right)$. *Therefore, we have*

$$\hat{m}_k(\mathbf{x}_k) - m_k(\mathbf{x}_k) = O_p(h_{k\iota}^2). \tag{3.11}$$

Remark 3.1 From Theorem 3.1 and assumption (B7), we can see that $\hat{m}_k(\mathbf{x}_k)$ converge to its true value at the speed of $(nh_{\pi k})^{-1/2}$. However, in Theorem 2.1 where the general nonparametric regression model is proposed, the convergence speed of the regression function is shown to be $(nh_\pi)^{-1/2}$. It is clear that the former converges in a higher speed than the later, and hence the curse of dimensionality is partly avoided. Moreover, when there is only one explanatory variable in the $m_k(.)$ function, this additive component can be estimated with the one-dimensional rate of convergence. Hence, our result is

coincident with the prior research.

When $\epsilon_{\mathbf{s}}$ satisfies the non-torus SGAR models defined in (2.3) or (2.4), we can further establish the asymptotic normality of $\hat{m}_k(\mathbf{x}_k)$. As the analysis with $\epsilon_{\mathbf{s}}$ modeled by (2.3) is more involved, we focus on that below. Same as in Chapter 2, we first transform (2.3) into an equivalent 2-dimensional moving average model, and then adopt the ′large and small block′ method with the truncated MA model to settle the problem. The theorem below can then be established.

Theorem 3.2 *If $\epsilon_{\mathbf{s}}$ satisfies model (2.3) and the assumptions* (B1)-(B7), (B4b) *hold, then we have*

$$(nh_{\pi k})^{1/2}\Big(\hat{m}_k(\mathbf{x}_k)-m_k(\mathbf{x}_k)-\text{Bias}(\mathbf{x}_k,\mathbf{h}_k)\Big)\xrightarrow{D} N\Big(0,\,\Sigma_\epsilon(\mathbf{x}_k)\Big), \tag{3.12}$$

where $\Sigma_\epsilon(\mathbf{x}_k) = \sigma_\epsilon^2 f_k(\mathbf{x}_k)\int K_\pi^2(\boldsymbol{\delta})d\boldsymbol{\delta}\, E\left[f_{-k}^2(\mathbf{X}_{\mathbf{s}}^{(-k)})f^{-2}(\mathbf{X}_{\mathbf{s}})\,|\,\mathbf{X}_{k\mathbf{s}}=\mathbf{x}_k\right]$, $\text{Bias}(\mathbf{x}_k,\mathbf{h}_k)$ *is defined in Theorem 3.1,* $\boldsymbol{\delta}$ *is a* b_k*-dimensional random vector, and* $\text{Bias}(\mathbf{x}_k)$ *is defined in Theorem 3.1.*

Finally, we will consider the asymptotic normality of the second-step estimator $\tilde{m}_k(\mathbf{x}_k)$ in (3.8). More statement about h_{kL} defined in (B7) should be given. Recall that, for $j = 1,\cdots,d$, h_{jL} is the bandwidth with the slowest convergence rate in the estimation of the single component $m_j(.)$. Now, we need to define a bandwidth with the slowest convergence rate in the estimation of the whole regression function $m(.)$. Among all h_{jL}, $j = 1,\cdots,d$, define h_L as $h_L \in \{h_{1L},\cdots,h_{dL}\}$ and $h_L = O\Big(h_{jL}^C\Big)$ for $0 < C \leq 1$. Then, in the asymptotic analysis of the second-step estimator $\tilde{m}_k(\mathbf{x}_k)$, we need the following assumption of the bandwidth vector $\mathbf{g}$.

(B8) All $g_{ij} > 0$, $g_{ij} \to 0$ and $ng_{\pi} \to \infty$ as $\mathbf{n} \to \infty$; moreover, when estimating $m_k(.)$, $\mathbf{g}_k$ tends to zero in the manner as $n = O\Big((g_{\pi k} g_{kL}^4)^{-1}\Big)$, where g_{kL} is an element in $\mathbf{g}_k$ satisfying that $g_{ij}/g_{kL} \to 0$ as $\mathbf{n} \to \infty$ for all $g_{ij} \neq g_{kL}$; at last, $h_L/g_{kL} \to 0$.

Theorem 3.3 *If the conditions in Theorem 3.1 and (B8) are satisfied, moreover, some consistent estimator of* $\boldsymbol{\theta}$ *is available, then*

$$(ng_{\pi k})^{1/2}\Big(\tilde{m}_k(\mathbf{x}_k) - m(\mathbf{x}_k) - \text{Bias}(\mathbf{x}_k, \mathbf{g}_k)\Big) \xrightarrow{D} N\Big(0,\, \Sigma_\tau(\mathbf{x}_k)\Big), \tag{3.13}$$

where $\Sigma_\tau(\mathbf{x}_k) = \sigma_\tau^2 f_k(\mathbf{x}_k) \int K_\pi^2(\boldsymbol{\delta}) d\boldsymbol{\delta}\, E\left[f_{-k}^{-2}(\mathbf{X}_{\mathbf{s}}^{(-k)}) f^{-2}(\mathbf{X}_{\mathbf{s}}) \,|\, \mathbf{X}_{k\mathbf{s}} = \mathbf{x}_k\right]$, *and* $\text{Bias}(\mathbf{x}_k, \mathbf{g}_k) = \frac{1}{2}\, tr[m_k''(\mathbf{x}_k) \int K_\pi(\boldsymbol{\delta})(\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T) \odot (\mathbf{g}_k \mathbf{g}_k^T)\, d\boldsymbol{\delta}]$.

3.4 Simulations

To evaluate our method, a simulation will be conducted in this section. We choose the same testing model in Simulation 2.2, which has additive structure, so that some cross comparison between the fitting results in this chapter and those in Chapter 2 can be executed. Observations will be generated in 3 different sample sizes, say 10×10, 15×15 and 20×20, with error terms satisfying model (2.2) -(2.4). For each sample size, we will replicate 100 times.

Simulation 3.1

$$Y_{\mathbf{s}} = sin(\pi X_{1,\mathbf{s}}) + cos(\pi X_{2,\mathbf{s}}) + \epsilon_{\mathbf{s}}\,, \tag{3.14}$$

where $X_{i,\mathbf{s}} \sim U(0, 4)$, $i = 1, 2$, $\epsilon_{\mathbf{s}}$ is modeled as (2.2) -(2.4) with $\tau_{\mathbf{s}} \sim N(0, 1)$ and coefficients $\boldsymbol{\theta} = (\theta_1, \theta_2)$ being $(0.38, -0.1)$, $(0.3, -0.2)$ and $(0.4, 0.3)$ respectively.

To estimate $m_k(x_k)$ in the additive model via the marginal integration technic, one need to fit the whole trend surface of the function $m(.)$ first and then sum up all the estimated values $\hat{m}(\mathbf{X}_{\mathbf{s}}^{(k)})$ with x_k keeping unchanged. When the number of explanatory variables is large, the fitting of the trend surface may suffer from the curse of dimensionality. Moreover, the calculation amount will be much larger than that of the nonparametric model considered in Chapter 2. To reduce the calculation expense to an acceptable level, we adopt some simplification in the bandwidth selection. First, we use a single bandwidth for both explanatory variables in the testing model. The reasons for this are that $X_{1,\mathbf{s}}$ and $X_{2\,\mathbf{s}}$ have identical distribution, and their functions are similar in the sense of function type. Hence, the optimal bandwidth for $X_{1,\mathbf{s}}$ and $X_{2\,\mathbf{s}}$ should be close in value, and it is appropriate to use a single bandwidth instead. Second, we determine the optimal bandwidths in the two steps separatively. This means, in each replication of the simulation, we first determine the optimal bandwidth for the first-step estimator in (3.3), then form the new process in (3.7), and at last determine the optimal bandwidth for the second-step estimator in (3.8). More details can be seen in Section 2.4.1. Although these simplifications of bandwidth selection need to be applied, we can see from the simulation results that our method still performs pretty good.

The main results of the simulation are shown in Table 3.1 and Table 3.2. We can see that our two-step estimation performs much better than the traditional estimation with marginal integration technique. All the values of MSE obtained from our method are much smaller, and the differences are outstanding in the cases with torus SGAR-type and separable SGAR-type errors. We also present the results of two typical fitting graphically in Figure 3.1 to Figure 3.4, where the error terms satisfy the torus SGAR model

.

Table 3.1: Simulation 3.1. MSE for the estimators of $m(.)$.

Sample Size	10×10	15×15	20×20
Torus$^{(a)}$			
mean of $\text{MSE}_o^{(d)}$	1.3007	0.6986	0.4108
mean of $\text{MSE}_m^{(e)}$	0.4144	0.1673	0.1016
var of MSE_o	1.4508	0.1156	0.0318
var of MSE_m	0.0376	0.0060	0.0031
Separable$^{(b)}$			
mean of MSE_o	5.2749	22.2180	2.1208
mean of MSE_m	0.4393	0.3991	0.0952
var of MSE_o	47.0274	3.36×10^3	1.8404
var of MSE_m	0.0443	0.3356	0.0021
Unilateral$^{(c)}$			
mean of MSE_o	0.3417	0.1559	0.1049
mean of MSE_m	0.3385	0.1428	0.1035
var of MSE_o	0.0263	0.0052	0.0022
var of MSE_m	0.0335	0.0055	0.0028

$^{(a)}$ Torus: ϵ_s generated by (2.2).
$^{(b)}$ Separable: with ϵ_s generated by (2.3).
$^{(c)}$ Unilateral: with ϵ_s generated by (2.4).
$^{(d)}$ MSE_o: MSE given by the traditional estimation with marginal integration technique.
$^{(e)}$ MSE_m: MSE given by our proposed method.

.

Table 3.2: Simulation 3.1. Estimators of $\boldsymbol{\theta}=(\theta_1,\theta_2)$.

Sample Size	10×10	15×15	20×20
Torus$^{(*)}$			
mean of $\hat{\theta}_1$	0.3307	0.3675	0.3852
mean of $\hat{\theta}_2$	-0.1472	-0.1262	-0.1161
Separable$^{(*)}$			
mean of $\hat{\theta}_1$	0.2939	0.2902	0.2853
mean of $\hat{\theta}_2$	-0.2088	-0.2041	-0.2101
Unilateral$^{(*)}$			
mean of $\hat{\theta}_1$	0.2891	0.3673	0.3522
mean of $\hat{\theta}_2$	0.2067	0.2721	0.3082

$^{(*)}$ Torus, Separable and Unilateral are same defined as the ones in Table 3.1.

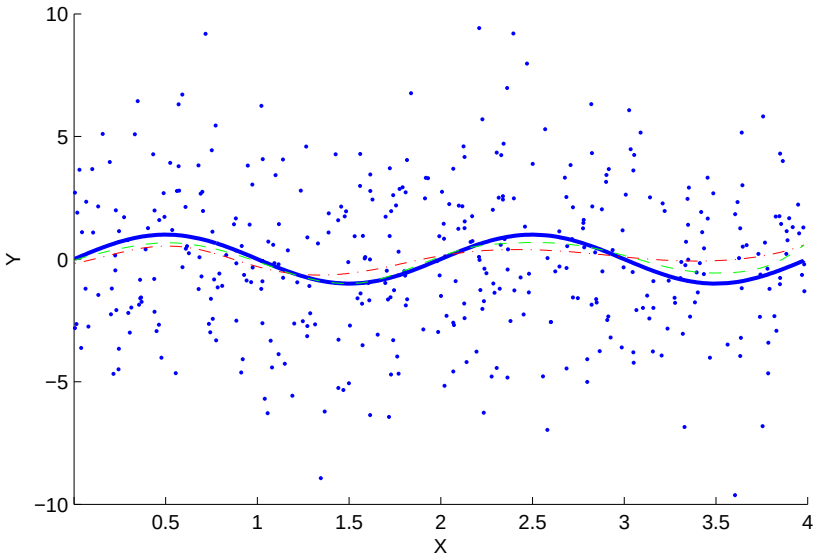


Figure 3.1: Simulation 3.1. The sin function fitting in the additive model with torus SGAR-type error. Solid line refers to the true values of the sin function; dashed line refers to the estimates given by our proposed method; and dash-dot line refers to those given by the traditional estimation with marginal integration technique. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

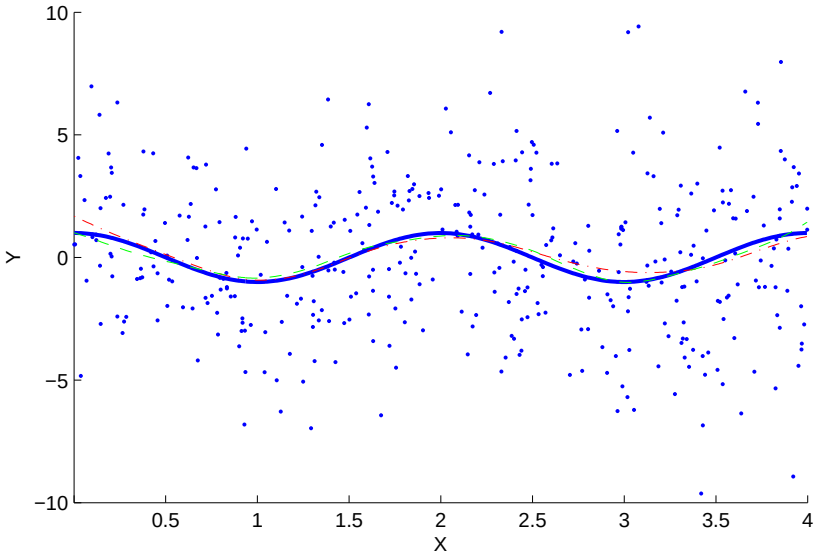


Figure 3.2: Simulation 3.1. The cos function fitting in the additive model with torus SGAR-type error. Solid line refers to the true values of the cos function; dashed line refers to the estimates given by our proposed method; and dash-dot line refers to those given by the traditional estimation with marginal integration technique. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

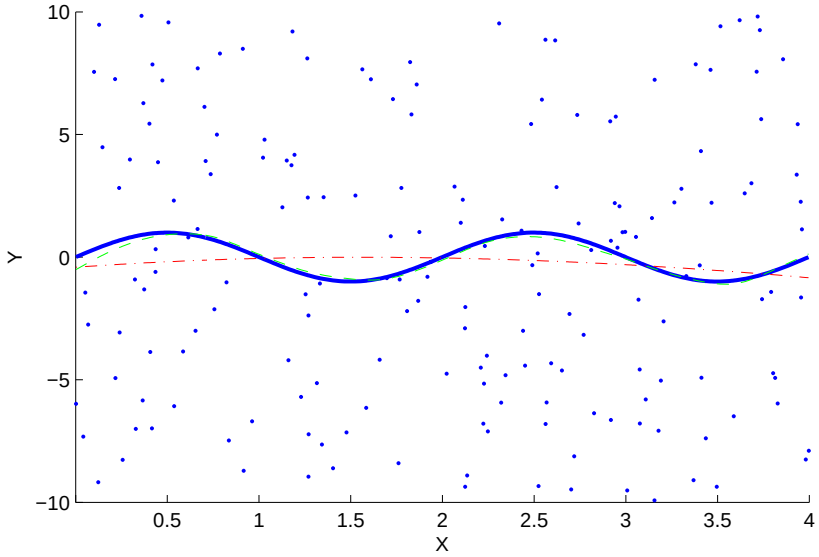


Figure 3.3: Simulation 3.1. The sin function fitting in the additive model with separable SGAR-type error. Solid line refers to the true values of the sin function; dashed line refers to the estimates given by our proposed method; and dash-dot line refers to those given by the traditional estimation with marginal integration technique. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

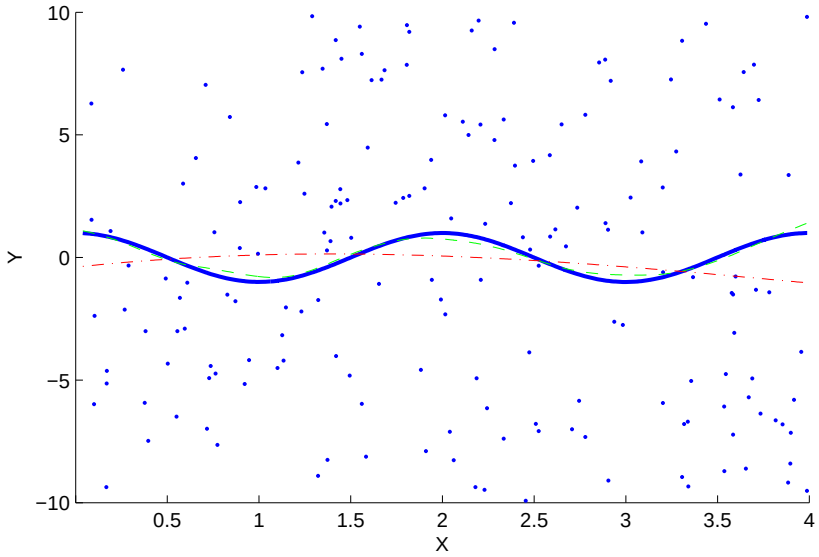


Figure 3.4: Simulation 3.1. The cos function fitting in the additive model with separable SGAR-type error. Solid line refers to the true values of the cos function; dashed line refers to the estimates given by our proposed method; and dash-dot line refers to those given by the traditional estimation with marginal integration technique. Sample size: 20×20. Note: as our estimates are too close to the true values, the dashed line may not be shown clearly in a black and white graph.

(2.2) and the separable SGAR model (2.3) respectively. From the figures, it is easy to see that the estimated curves obtain by our method are closer to the true curves than the one obtained by the traditional method.

In this simulation, we use the same testing model as the one in Simulation 2.3. For the later, the two-step estimation introduced in Chapter 2 is used. Therefore, we can compare the performance of the methods in Chapter 2 and in this chapter, given that the testing model does have additive structure. From Table 2.6 and Table 3.1, we can find that the MSE values obtained by the method in this chapter are smaller than the other. The results imply that the method considered in this chapter is more suitable for the case that additive structure of the regression function exists.

3.5 Proofs

Proof of Theorem 3.1: Let $\mathbf{F}(\mathbf{X}_\mathbf{s}^{(k)}) = \left(\overline{Y} + m(\mathbf{X}_\mathbf{s}^{(k)}), (m'(\mathbf{X}_\mathbf{s}^{(k)}) \odot \mathbf{h})^T \right)^T$. Note that we have $\eta^T\left(\vec{\mathbf{X}}^T(\mathbf{x})W(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x})\right)^{-1}\vec{\mathbf{X}}^T(\mathbf{x})W(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x}) = (1, \mathbf{0})$, then

$$\hat{m}_k(\mathbf{x}_k) - m_k(\mathbf{x}_k) = \frac{1}{n}\sum_{\mathbf{s}\in\Omega} \eta^T\left(\vec{\mathbf{X}}^T(\mathbf{X}_\mathbf{s}^{(k)})W(\mathbf{X}_\mathbf{s}^{(k)})\vec{\mathbf{X}}(\mathbf{X}_\mathbf{s}^{(k)})\right)^{-1}\vec{\mathbf{X}}^T(\mathbf{X}_\mathbf{s}^{(k)})W(\mathbf{X}_\mathbf{s}^{(k)})\left[\mathbf{Y} - \vec{\mathbf{X}}(\mathbf{X}_\mathbf{s}^{(k)})\mathbf{F}(\mathbf{X}_\mathbf{s}^{(k)})\right] + O_p(n^{-1/2}). \tag{3.15}$$

By Lemma 2.1 and the symmetry property of $K(.)$,

$$\vec{\mathbf{X}}^T(\mathbf{X}_\mathbf{s}^{(k)})W(\mathbf{X}_\mathbf{s}^{(k)})\vec{\mathbf{X}}(\mathbf{X}_\mathbf{s}^{(k)}) = f(\mathbf{X}_\mathbf{s}^{(k)})\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \int \boldsymbol{\delta}\boldsymbol{\delta}^T K_\pi(\boldsymbol{\delta}) d\boldsymbol{\delta} \end{pmatrix} + o_p(1), \tag{3.16}$$

where $\boldsymbol{\delta}$ is a D-dimensional vector. Then,

$$\eta^T\Big(\vec{\mathbf{X}}^T(\mathbf{X}_\mathbf{s}^{(k)})W(\mathbf{X}_\mathbf{s}^{(k)})\vec{\mathbf{X}}(\mathbf{X}_\mathbf{s}^{(k)})\Big)^{-1} = \eta^T\Big(f^{-1}(\mathbf{X}_\mathbf{s}^k)+o_p(1)\Big). \tag{3.17}$$

Substitute this into (3.15), then

$$\begin{aligned}\hat{m}_k(\mathbf{x}_k)-m_k(\mathbf{x}_k) \;=\; & \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\sum_{\mathbf{t}\in\Omega}(nh_\pi)^{-1}f^{-1}(\mathbf{X}_\mathbf{s}^k)(1+o_p(1))\,K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_\mathbf{t}-\mathbf{X}_\mathbf{s}^{(k)})\Big)\,(\mu-\overline{Y}\,) \\ & +\frac{1}{n}\sum_{\mathbf{s}\in\Omega}\sum_{\mathbf{t}\in\Omega}(nh_\pi)^{-1}f^{-1}(\mathbf{X}_\mathbf{s}^k)(1+o_p(1))\,K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_\mathbf{t}-\mathbf{X}_\mathbf{s}^{(k)})\Big)\,\mathbf{R}_\mathbf{t} \\ & +\frac{1}{n}\sum_{\mathbf{s}\in\Omega}\sum_{\mathbf{t}\in\Omega}(nh_\pi)^{-1}f^{-1}(\mathbf{X}_\mathbf{s}^k)(1+o_p(1))\,K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_\mathbf{t}-\mathbf{X}_\mathbf{s}^{(k)})\Big)\epsilon_\mathbf{t} + O_p(n^{-1/2}) \\ \hat{=}\; & Q_1 + Q_2 + Q_3 + O_p(n^{-1/2}),\end{aligned} \tag{3.18}$$

where $\mathbf{R}_\mathbf{t} = \frac{1}{2}tr\Big(m_k''(\mathbf{x}_k)(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)^T\Big) + \frac{1}{2}\sum_{j\neq k} tr\Big(m_j''(\mathbf{x}_j)(\mathbf{X}_{j\mathbf{t}}-\mathbf{X}_{j\mathbf{s}})(\mathbf{X}_{j\mathbf{t}}-\mathbf{X}_{j\mathbf{s}})^T\Big)$, and $m_j''(\mathbf{x}_j)$ is the second derivative matrix of $m_j(.)$ at $\mathbf{x}_j$.

With assumption (B7), it is sufficient for us to establish (3.10) by showing

$$Q_1 = O_p(n^{-1/2}), \tag{3.19}$$

$$Q_2 = \frac{1}{2}tr\Big[m_k''(\mathbf{x}_k)\int K_\pi(\boldsymbol{\delta})(\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T)\odot(\mathbf{h}_k\mathbf{h}_k^T)\,d\boldsymbol{\delta}\Big] + o_p(h_{k\mathcal{L}}^2), \tag{3.20}$$

$$Q_3 = (nh_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega} f^{-1}(\mathbf{X}_\mathbf{t}^{(k)})f_{-k}(\mathbf{X}_\mathbf{t}^{(-k)})\,K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k))\,\epsilon_\mathbf{t} + o_p((nh_{\pi k}^{-1/2})). \tag{3.21}$$

First, we consider Q_1. It is easy to show that $(nh_\pi)^{-1}\sum_{\mathbf{t}\in\Omega} K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_\mathbf{t}-\mathbf{X}_\mathbf{s}^{(k)}))= f(\mathbf{X}_\mathbf{s}^{(k)}) + O_p(h_{k\mathcal{L}})$ and $\mu-\overline{Y} = O_p(n^{-1/2})$. Therefore, (3.19) follows directly.

Next, let us focus on Q_2. Define

$$H_k = (nh_\pi)^{-1}\sum_{\mathbf{t}\in\Omega} K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_\mathbf{t}-\mathbf{X}_\mathbf{s}^{(k)})\Big)\,tr\Big[m_k''(\mathbf{x}_k)(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)^T\Big], \tag{3.22}$$

and for $j \neq k$, we define H_j similar to H_k but with $\mathbf{x}_k$ substituted by $\mathbf{X}_{j\mathbf{s}}$.

Then, same as Lemma 2.6, we can show that

$$E(H_k) = f(\mathbf{X}_{\mathbf{s}}^{(k)})\, tr\Big[m_k''(\mathbf{x}_k)\int K_\pi(\boldsymbol{\delta})(\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T)\odot(\mathbf{h}_k\mathbf{h}_k^T)\, d\boldsymbol{\delta}\Big] + o(h_{k\mathcal{L}}^2), \tag{3.23}$$

$$Var(H_k) \leq (nh_\pi)^{-1}\Big[f(\mathbf{X}_{\mathbf{s}}^{(k)})\int tr^2\Big(m_k''(\mathbf{x}_k)(\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T)\odot(\mathbf{h}_k\mathbf{h}_k^T)\Big)\, K_\pi^2(\boldsymbol{\delta})\, d\boldsymbol{\delta} + o(h_{k\mathcal{L}}^4)\Big], \tag{3.24}$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_1^T, \cdots, \boldsymbol{\delta}_d^T)^T$, $\boldsymbol{\delta}_i$ is a b_i-dimensional vector, $i = 1, \cdots, d$. Hence, we have

$$H_k = f(\mathbf{X}_{\mathbf{s}}^{(k)})\, tr\Big[m_k''(\mathbf{x}_k)\int K_\pi(\boldsymbol{\delta})(\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T)\odot(\mathbf{h}_k\mathbf{h}_k^T)\, d\boldsymbol{\delta}\Big] + o_p(h_{k\mathcal{L}}^2)\,. \tag{3.25}$$

which is equal to $O_p(h_{k\mathcal{L}}^2)$. With the same procedure above, we get

$$H_j = f(\mathbf{X}_{\mathbf{s}}^{(k)})\, tr\Big[m_j''(\mathbf{X}_{j\mathbf{s}})\int K_\pi(\boldsymbol{\delta})(\boldsymbol{\delta}_j\,\boldsymbol{\delta}_j^T)\odot(\mathbf{h}_j\mathbf{h}_j^T)\, d\boldsymbol{\delta}\Big] + o_p(h_{k\mathcal{L}}^2)\,, \tag{3.26}$$

for $j \neq k$. Due to the assumption that $\mathbf{h}_j/h_{k\mathcal{L}} = o(\mathbf{1})$, we can see $H_j = o_p(h_{k\mathcal{L}}^2)$ for any $j \neq k$. This, together with (3.25), is sufficient for us to show (3.20).

At last, we turn to Q_3. Define

$$C_{\mathbf{t}k} = (nh_{\pi(-k)})^{-1}\sum_{\mathbf{s}\in\Omega} f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\, K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{t}}^{(-k)} - \mathbf{X}_{\mathbf{s}}^{(-k)})\Big), \tag{3.27}$$

where $h_{\pi(-k)} = \prod_{j\neq k} h_{\pi j}$. Then Q_3 can be expressed as

$$Q_3 = (nh_{\pi k})^{-1}(1+o_p(1))\sum_{\mathbf{t}\in\Omega} K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)\, C_{\mathbf{t}k}\, \epsilon_{\mathbf{t}}. \tag{3.28}$$

It is easy to show that $C_{\mathbf{t}k} = f^{-1}(\mathbf{X}_{\mathbf{t}}^{(k)}) f_{-k}(\mathbf{X}_{\mathbf{t}}^{(-k)}) + o_p(h_{k\mathcal{L}})$, then we have

$$\begin{aligned} Q_3 &= \Big[(nh_{\pi k})^{-1} \sum_{\mathbf{t}\in\Omega} f^{-1}(\mathbf{X}_{\mathbf{t}}^{(k)}) f_{-k}(\mathbf{X}_{\mathbf{t}}^{(-k)})\, K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)\, \epsilon_{\mathbf{t}}\Big] \cdot (1 + o_p(1)) \\ &\quad + \Big[(nh_{\pi k})^{-1} \sum_{\mathbf{t}\in\Omega} K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)\, \epsilon_{\mathbf{t}}\Big] \cdot o_p(h_{k\mathcal{L}}) \\ &\hat{=} Q_{31} \cdot (1 + o_p(1)) + Q_{32} \cdot o_p(h_{k\mathcal{L}}). \end{aligned} \tag{3.29}$$

Obviously Q_{31} has zero mean, while its variance is given as follows.

$$\begin{aligned} Var(Q_{31}) &= (nh_{\pi k})^{-2} \sigma_\epsilon^2 \sum_{\mathbf{t}\in\Omega} E\Big[f^{-2}(\mathbf{X}_{\mathbf{t}}^{(k)}) f_{-k}^2(\mathbf{X}_{\mathbf{t}}^{(-k)})\, K_\pi^2\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)\Big] \\ &\quad + (nh_{\pi k})^{-2} \sum_{\substack{\mathbf{t}'\neq\mathbf{t}'' \\ \mathbf{t}',\mathbf{t}''\in\Omega}} Cov(\epsilon_{\mathbf{t}'}, \epsilon_{\mathbf{t}''})\, E^2\Big[f^{-1}(\mathbf{X}_{\mathbf{t}}^{(k)}) f_{-k}(\mathbf{X}_{\mathbf{t}}^{(-k)}) K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)\Big] \end{aligned} \tag{3.30}$$

Let $\mathbf{u} = \Big(\mathbf{u}_k^T, (\mathbf{u}^{(-k)})^T\Big)^T$, where $\mathbf{u}_k$ is a b_k dimensional random vector and $\mathbf{u}^{(-k)}$ is a $D - b_k$ dimensional random vector. Then

$$\begin{aligned} & E\Big[f^{-2}(\mathbf{X}_{\mathbf{t}}^{(k)}) f_{-k}^2(\mathbf{X}_{\mathbf{t}}^{(-k)})\, K_\pi^2\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)\Big] \\ &= \int f^{-2}(\mathbf{x}_k, \mathbf{u}^{(-k)})\, f_{-k}^2(\mathbf{u}^{(-k)})\, K_\pi^2\Big(\boldsymbol{\zeta}_1(\mathbf{u}_k - \mathbf{x}_k)\Big)\, f(\mathbf{u})\, d\mathbf{u} \\ &= h_{\pi k} \int f^{-2}(\mathbf{x}_k, \mathbf{u}^{(-k)})\, f_{-k}^2(\mathbf{u}^{(-k)}) \Big[\int K_\pi^2(\boldsymbol{\delta}) f(\boldsymbol{\delta} \odot \mathbf{h}_k + \mathbf{x}_k, \mathbf{u}^{(-k)})\, d\boldsymbol{\delta}\Big]\, d\mathbf{u}^{(-k)} \\ &= h_{\pi k} \Big[\int K_\pi^2(\boldsymbol{\delta})\, d\boldsymbol{\delta} \int f^{-1}(\mathbf{x}_k, \mathbf{u}^{(-k)}) f_{-k}^2(\mathbf{u}^{(-k)})\, d\mathbf{u}^{(-k)} + o(1)\Big], \end{aligned} \tag{3.31}$$

which is equal to $O(h_{\pi k})$. Similarly, we can show

$$E\Big[f^{-1}(\mathbf{X}_{\mathbf{t}}^{(k)}) f_{-k}(\mathbf{X}_{\mathbf{t}}^{(-k)}) K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)\Big] = O(h_{\pi k}). \tag{3.32}$$

Moreover, via Lemma 2.3 or Lemma 2.5, we have

$$\begin{aligned}\sum_{\substack{\mathbf{t}'\neq\mathbf{t}''\\ \mathbf{t}',\mathbf{t}''\in\Omega}} Cov(\epsilon_{\mathbf{t}'},\epsilon_{\mathbf{t}''}) &= \sum_{|l_1|<n_1}\sum_{|l_2|<n_2}(n_1-|l_1|)(n_2-|l_2|)\Big|R(l_1,l_2)\Big| - n\sigma_\epsilon^2 \\ &\leq Cn\sum_{l_1=0}^{n_1}\sum_{l_2=0}^{n_2}\Big|R(l_1,l_2)\Big| = O(n). \end{aligned} \tag{3.33}$$

Substitute (3.31) - (3.33) into (3.30), we can show $Var(Q_{31}) = O\Big((nh_{\pi k})^{-1}\Big)$, and hence $Q_{31} = O_p\Big((nh_{\pi k})^{-1/2}\Big)$. By the similar procedure as (3.30)-(3.33), we can also show that $Q_{32} = O_p\Big((nh_{\pi k})^{-1/2}\Big)$. Via (3.29), (3.21) can then be shown, with the first term on its right hand side being $O_p\Big((nh_{\pi k})^{-1/2}\Big)$. The proof of Theorem 3.1 is completed. □

Theorem 3.2 can be established directly with the two lemma below. We define

$$A_{\mathbf{n}} = (nh_{\pi k})^{-1/2}\sum_{\mathbf{s}\in\Omega} f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\epsilon_{\mathbf{s}}. \tag{3.34}$$

From equation (3.10), we can see that it suffices for us establish the asymptotic normality of $\hat{m}_k(\mathbf{x}_k)$ by showing A_n is asymptotically normal. To get this, we may use the technique of truncated MA model. By Lemma 2.4, we can transform (2.3) into some 2-dimensional moving average model, where the coefficients satisfy $|\psi_{\mathbf{t}}| \leq C\|\mathbf{t}\|^{-\alpha}$, $\alpha = 3+\delta$, δ is some positive constant and $\mathbf{t} = (t_1, t_2) \in Z^2$. Define

$$\epsilon_{\mathbf{s}}^m = \sum_{|t_1|,|t_2|<m} \psi_{\mathbf{t}}\tau_{\mathbf{s}-\mathbf{t}} \quad \text{and} \quad \tilde{\epsilon}_{\mathbf{s}} = \epsilon_{\mathbf{s}}^m + \eta_{\mathbf{s}},$$

where m is some amount satisfying $m \to \infty$ as $\mathbf{n} \to \infty$, and $\eta_{\mathbf{s}}$ is independent random variable with zero mean and finite variance. Furthermore, we assume that $\{\eta_{\mathbf{s}}\}$ is independent of $\{\tau_{\mathbf{s}}\}$ and $\{\mathbf{X}_{\mathbf{s}}\}$, and $\eta_{\mathbf{s}} \backsim \epsilon_{\mathbf{s}} - \epsilon_{\mathbf{s}}^m$ where $\backsim$ denote equality in distribution. Now

define $\tilde{A}_{\mathbf{n}}$ similar to $A_{\mathbf{n}}$, but with $\epsilon_{\mathbf{s}}$ in (3.34) substituted by $\tilde{\epsilon}_{\mathbf{s}}$, that is

$$\tilde{A}_{\mathbf{n}} = (nh_{\pi k})^{-1/2} \sum_{\mathbf{s}\in\Omega} f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_{\pi}\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\, \tilde{\epsilon}_{\mathbf{s}}. \tag{3.35}$$

Then following lemma shows the equivalence of the asymptotic normality of $A_{\mathbf{n}}$ and $\tilde{A}_{\mathbf{n}}$.

Lemma 3.1 *If $\epsilon_{\mathbf{s}}$ can be expressed in the form of (2.35) with coefficient $\psi_{\mathbf{k}}$ satisfying (2.36), and assumptions* (B1), (B2), (B4), (B5) *and* (B7) *hold, then*

$$|A_{\mathbf{n}} - \tilde{A}_{\mathbf{n}}| = o_p(1)\,. \tag{3.36}$$

Proof of Lemma 3.1:

$$\begin{aligned} Pr\Big[|A_{\mathbf{n}} - \tilde{A}_{\mathbf{n}}| > \eta\Big] &= Pr\Big[\,\Big|\sum_{\mathbf{s}\in\Omega} (\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}})\, f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_{\pi}\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\Big| > (nh_{\pi k})^{1/2}\eta\Big] \\ &\le \sum_{\mathbf{s}\in\Omega} Pr\Big[\,\Big|(\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}})\, f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_{\pi}\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\Big| > n^{-1/2}h_{\pi k}^{1/2}\eta\,\Big] \\ &\le n^2 h_{\pi k}^{-1}\eta^2\, Var\Big[(\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}}) f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_{\pi}\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\Big]. \end{aligned} \tag{3.37}$$

Define $\Gamma = \{\mathbf{t} = (t_1, t_2) : |t_i| \ge m \text{ for at least one } i, i = 1, 2\}$, then $\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}} = \sum_{\mathbf{t}\in\Gamma} \psi_{\mathbf{t}}\tau_{\mathbf{s}-\mathbf{t}} + \eta_{\mathbf{s}}$. As $\{\mathbf{X}_{\mathbf{s}}\}$ is independent of $\{\tau_{\mathbf{s}}\}$ and $\{\eta_{\mathbf{s}}\}$, by (3.37) we have

$$Pr\Big[|A_{\mathbf{n}} - \tilde{A}_{\mathbf{n}}| > \eta\Big] \le Cn^2 h_{\pi k}^{-1} E[(\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}})^2] E\Big[f^{-2}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}^2(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_{\pi}^2\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\Big].$$

From (3.31), we can see that $h_{\pi k}^{-1} E\Big[f^{-2}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}^2(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_{\pi}^2\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\Big]$ is bounded. By the proof of Lemma 2.7, we also get $E[(\epsilon_{\mathbf{s}} - \tilde{\epsilon}_{\mathbf{s}})^2] \le Cnm^{-2\alpha}$, where $m \to \infty$ as $\mathbf{n} \to \infty$ and $\alpha = 3 + \delta$, δ is some positive constant. So (3.37) is finally bounded by $Cn^3 m^{-2\alpha}$. Let $m = \Big(n(h_{\pi k} h_{kL}^4)^{c_1}\Big)^{1/2}$, c_1 is some constant and $0 < c_1 < \delta/(3 + \delta)$, we can ensure that $m \to \infty$ and $n^3 m^{-2\alpha} \to 0$ as $\mathbf{n} \to \infty$. The proof is then completed. □

Lemma 3.2 *If $\epsilon_{\mathbf{s}}$ can be expressed in the form of (2.35) with coefficient $\psi_{\mathbf{k}}$ satisfying (2.36), and assumptions* (B1), (B3), (B4)-(B7) *hold, then we have*

$$\tilde{A}_{\mathbf{n}} \xrightarrow{D} N\Big(0,\, \Sigma_{\epsilon}(\mathbf{x}_k)\Big),$$

where $\Sigma_{\epsilon}(\mathbf{x}_k)$ is defined in Theorem 3.2.

Proof of Lemma 3.2: Let $\Delta_{\mathbf{s}} = f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\, f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\, K_{\pi}\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}} - \mathbf{x}_k)\Big)\tilde{\epsilon}_{\mathbf{s}}$, and therefore $\tilde{A}_{\mathbf{n}} = \sum_{\mathbf{s}\in\Omega} (nh_{\pi k})^{-1/2}\Delta_{\mathbf{s}}$. Due to the fact that $m \to \infty$ as $\mathbf{n} \to \infty$, we have $\tilde{\epsilon}_{\mathbf{s}} \to \epsilon_{\mathbf{s}}$ as $\mathbf{n} \to \infty$. Hence, similar to (3.30) - (3.33), the asymptotical variance of $\tilde{A}_{\mathbf{s}}$ can be given as

$$\lim_{\mathbf{n}\to\infty} Var\,(\tilde{A}_{\mathbf{n}}) = \sigma_{\epsilon}^2 f_k(\mathbf{x}_k) \int K_{\pi}^2(\boldsymbol{\delta}) d\boldsymbol{\delta}\, E\,[f_{-k}^2(\mathbf{X}_{\mathbf{s}}^{(-k)}) f^{-2}(\mathbf{X}_{\mathbf{s}}) \Big|\, \mathbf{X}_{k\mathbf{s}} = \mathbf{x}_k] \hat{=} \Sigma_{\epsilon}(\mathbf{x}_k). \tag{3.38}$$

We decompose $\tilde{A}_{\mathbf{n}}$ into pieces involving "large" and "small" blocks. Let (l_1, l_2) be the length of the large blocks, satisfying $n_i = r_i(2m + l_i)$, $l_i/m \to \infty$ as $\mathbf{n} \to \infty$, and r_i is some integer tending to infinity as well. When r_i cannot take integer value, some remaining block will exist. However, the consideration of such case is similar to that presented in the following, and hence is omitted here. We let $l_i = [n(h_{\pi k}h_{kL}^4)^{c_2}]^{1/2}$, where $0 < c_2 < c_1$, c_1 is used in the proof of Lemma 3.1, and c_2 can take different values with distinct $l_i, i = 1, 2$. Now denote $\mathbf{j} = (j_1, j_2)$ and

$$U(1, \mathbf{n}, \mathbf{j}) = \sum_{\substack{s_i = j_i(l_i+2m)+1 \\ i=1,2}}^{j_i(l_i+2m)+l_i} (nh_{\pi k})^{-1/2}\Delta_{\mathbf{s}},$$

which is the sum of random variables over the $\mathbf{j}$-th large block. To sum up the random variables over all large blocks, we define

$$T(\mathbf{n}, 1) = \sum_{\substack{0 \le j_i \le r_i - 1 \\ i=1,2}} U(1, \mathbf{n}, \mathbf{j}).$$

We also define

$$U(2,\mathbf{n},\mathbf{j})=\sum_{s_1=j_1(l_1+2m)+1}^{j_1(l_1+2m)+l_1}\ \sum_{s_2=j_2(l_2+2m)+l_2+1}^{(j_2+1)(l_2+2m)}(nh_{\pi k})^{-1/2}\Delta_{\mathbf{s}},$$

$$U(3,\mathbf{n},\mathbf{j})=\sum_{s_1=j_1(l_1+2m)+l_1+1}^{(j_1+1)(l_1+2m)}\ \sum_{s_2=j_2(l_2+2m)+1}^{j_2(l_2+2m)+l_2}(nh_{\pi k})^{-1/2}\Delta_{\mathbf{s}},$$

$$U(4,\mathbf{n},\mathbf{j})=\sum_{\substack{s_i=j_i(l_i+2m)+l_i+1\\ i=1,2}}^{(j_i+1)(l_i+2m)}(nh_{\pi k})^{-1/2}\Delta_{\mathbf{s}},$$

which are the sum of random variables over different types of small blocks. Then $T(\mathbf{n},k)$, $2\le k\le 4$, can be defined similarly as $T(\mathbf{n},1)$. Hence we have $\tilde{A}_{\mathbf{n}}=\sum_{k=1}^{4}T(\mathbf{n},k)$. We then process to show

(b1) $T(\mathbf{n},k)=o_p(1),$ for $2\le k\le 4,$

(b2) $Var\,[T(\mathbf{n},1)]\,\hat{=}\,S_{\mathbf{n}}^2\to\Sigma_\epsilon(\mathbf{x}_k),$ as $\mathbf{n}\to\infty,$

(b3) $T(\mathbf{n},1)/S_{\mathbf{n}}\xrightarrow{D}N(0,\mathbf{1}).$

Proof of (b1): Without loss of generality, we only prove $T(\mathbf{n},2)=o_p(1)$, while the proofs for the rest are similar. It is sufficient for us to show the variance of $T(\mathbf{n},2)$ tends to zero. Define

$$S(2,\mathbf{j})=\{\,\mathbf{s}:\ j_1(l_1+2m)+1\le s_1\le j_1(l_1+2m)+l_1,\ j_2(l_2+2m)+l_2+1\le s_2\le(j_2+1)(l_2+2m)\},$$

which is the location set of the random variables in $U(2,\mathbf{n},\mathbf{j})$, and it contains $2l_1m$

elements . Notice that distinct $U(2,\mathbf{n},\mathbf{j})$, respecting to $\mathbf{j}$, are independent, thus we have

$$\begin{aligned} Var[T(\mathbf{n},2)] &= \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} Var[U(2,\mathbf{n},\mathbf{j})] \\ &= (nh_{\pi k})^{-1}\sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}}\left(\sum_{\mathbf{s}\in S(2,\mathbf{j})} Var(\Delta_{\mathbf{s}}) + \sum_{\mathbf{s}',\mathbf{s}''\in S(2,\mathbf{j})} Cov(\Delta_{\mathbf{s}'},\Delta_{\mathbf{s}''})\right) \\ &\le (nh_{\pi k})^{-1}(2l_1m\times r_1\times r_2)Var(\Delta_{\mathbf{s}}) + (nh_{\pi k})^{-1}\sum_{\substack{\mathbf{s}',\mathbf{s}''\in\Omega\\ \mathbf{s}'\neq\mathbf{s}''}} Cov(\Delta_{\mathbf{s}'},\Delta_{\mathbf{s}''}). \end{aligned}$$

It is easy to show $h_{\pi k}^{-1}Var(\Delta_{\mathbf{s}})$ is bounded, then with the fact that $m/l_2\to 0$, the first term on the right hand side of the last inequality tends to zero. Similar to the proof of (3.30)-(3.33), we can show that the second term is equal to $O(h_{\pi k})$. So we complete the proof of (b1).

Proof of (b2): By (b1) and (3.38), (b2) follows straightforwardly.

Proof of (b3): Notice that all $U(1,\mathbf{n},\mathbf{j})$ are independent. By the Linderbery central limit theorem, (b3) is held if we can show

$$\sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} E\left[(U(1,\mathbf{n},\mathbf{j}))^2 I\Big\{|U(1,\mathbf{n},\mathbf{j})|>\epsilon S_{\mathbf{n}}\Big\}\right]\longrightarrow 0, \tag{3.39}$$

for any $\epsilon>0$. As $\tilde{\epsilon}_{\mathbf{s}}$ is not bounded, we define the truncated variables $\tilde{\epsilon}_{\mathbf{s}}^L=\tilde{\epsilon}_{\mathbf{s}}I\{|\tilde{\epsilon}_{\mathbf{s}}|\le L\}$, L is some positive constant. Similarly, let $\Delta_{\mathbf{s}}^L=f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})\,f_{-k}(\mathbf{X}_{\mathbf{s}}^{(-k)})\,K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{s}}-\mathbf{x}_k)\Big)\tilde{\epsilon}_{\mathbf{s}}^L$ and

$$\begin{aligned} U^L(1,\mathbf{n},\mathbf{j}) &= \sum_{\mathbf{s}\in S(1,\mathbf{j})}(nh_{\pi k})^{-1/2}\Delta_{\mathbf{s}}^L, \\ T^L(\mathbf{n},1) &= \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} U^L(1,\mathbf{n},\mathbf{j}), \end{aligned}$$

where $S(1,\mathbf{j})=\{\,\mathbf{s}: j_i(l_i{+}2m){+}1\le s_i\le j_i(l_i{+}2m){+}l_i,\ i=1,2\,\}$. As the asymptotic variance of $T(\mathbf{n},1)$ exists, the one of $T^L(\mathbf{n},1)$ also exists. We denote $S_{\mathbf{n}}^L$ as the asymptotic standard

deviation of $T^L(\mathbf{n},1)$. Now we first show that $T^L(\mathbf{n},1)\Big/S^L_{\mathbf{n}} \xrightarrow{D} N(0,1)$, or equivalently

$$J \triangleq \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} E\left[\left(U^L(1,\mathbf{n},\mathbf{j})\right)^2 I\Big\{|U^L(1,\mathbf{n},\mathbf{j})| > \epsilon S^L_{\mathbf{n}}\Big\}\right] \longrightarrow 0\,, \tag{3.40}$$

and then generalize the result to (b3).

$$\begin{aligned} J &\le CL^2(nh_{\pi k})^{-1} l_1^2 l_2^2 \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} Pr\Big[|U^L(1,\mathbf{n},\mathbf{j})| > \epsilon\, S^L_{\mathbf{n}}\Big] \\ &\le C(nh_{\pi k})^{-1} l_1^2 l_2^2 \sum_{\substack{0\le j_i\le r_i-1\\ i=1,2}} (\epsilon\, S^L_{\mathbf{n}})^{-2} E\left[\left(U^L(1,\mathbf{n},\mathbf{j})\right)^2\right]. \end{aligned} \tag{3.41}$$

Examine the following expectation,

$$E\left[\left(U^L(1,\mathbf{n},\mathbf{j})\right)^2\right] = (nh_{\pi k})^{-1} E\left\{\Big[\sum_{\mathbf{s}\in S(1,\mathbf{j})} \Delta^L_{\mathbf{s}}\Big]^2\right\} \le C(nh_{\pi k})^{-1}\Big\{ l_1 l_2\, E[(\Delta^L_{\mathbf{s}})^2] + l_1^2\, l_2^2\, E^2[\Delta^L_{\mathbf{s}}]\Big\}.$$

Since $E[(\Delta^L_{\mathbf{s}})^2] = O(h_{\pi k})$ and $E[\Delta^L_{\mathbf{s}}] = O(h_{\pi k})$, we have

$$E\left[\left(U^L(1,\mathbf{n},\mathbf{j})\right)^2\right] \le C\Big(n^{-1} l_1 l_2 (1 + h_{\pi k} l_1 l_2)\Big).$$

Substitute this into the expression of J, and rewrite n, l_1 and l_2 in terms of h_π and h_L, we then have

$$J \le C\Big((h_{\pi k}^{c_2-1} h_{kL}^{4c_2-2})^2 + (h_{\pi k} h_{kL}^4)^{3c_2-2}\Big),$$

which will tend to zero with some c_2 close enough to 1.

Now define $T^{L*}(\mathbf{n},1) = T(\mathbf{n},1) - T^L(\mathbf{n},1)$, then

$$\begin{aligned}
&\left\| E\left[\exp\left(i\lambda T(\mathbf{n},1)\right) - \exp\left(-\lambda^2 S_{\mathbf{n}}^2/2\right)\right]\right\| \\
\leq\ &\left\| E\left[\exp\left(i\lambda T^L(\mathbf{n},1)\right) - \exp\left(-\lambda^2 (S_{\mathbf{n}}^L)^2/2\right)\right]\right\| + E\left\| \exp(i\lambda T^{L*}(\mathbf{n},1)) - 1\right\| \\
&+ \left| \exp(-\lambda^2 (S_{\mathbf{n}}^L)^2/2) - \exp(-\lambda^2 S_{\mathbf{n}}^2/2)\right| \\
\hat{=}\ &E_1 + E_2 + E_3.
\end{aligned}$$

As we have shown $T^L(\mathbf{n},1)\Big/S_{\mathbf{n}}^L \xrightarrow{D} N(0,1)$ above, therefore, E_1 tends to zero. E_3 also tends to zero by dominated convergence theorem. It remains for us to show E_2 converges, and this is held if $\mathrm{Var}\left[T^{L*}(\mathbf{n},1)\right] \to 0$. Similar to establishing the asymptotic variance of $T(\mathbf{n},1)$, we have

$$\lim_{\mathbf{n}\to\infty} \mathrm{Var}\left[T^{L*}(\mathbf{n},1)\right] = \mathrm{Var}\Big(\tilde{\epsilon}_{\mathbf{s}} I\{|\tilde{\epsilon}_{\mathbf{s}}| > L\}\Big) f_k(\mathbf{x}_k) \int K_\pi^2(\boldsymbol{\delta}) d\boldsymbol{\delta}\, E\Big[f_{-k}^2(\mathbf{X}_{\mathbf{s}}^{(-k)}) f^{-2}(\mathbf{X}_{\mathbf{s}}) \Big| \mathbf{X}_{k\mathbf{s}} = \mathbf{x}_k\Big]$$

which converges to zero as $L \to \infty$ by dominated convergence theorem. The proof is then completed. □

Proof of Theorem 3.3: Recall that $\hat{\mathbf{P}}$ can be expressed as $\hat{\mathbf{P}} = \hat{\mu} + \sum_{i=1}^{d} \hat{\mathbf{M}}_k + (\mathbf{I} - \mathbf{B}(\hat{\boldsymbol{\theta}}))\hat{\mathbf{E}}$. If we substitute $\hat{\boldsymbol{\theta}}$ by $\boldsymbol{\theta}$, we can form another process as $\check{\mathbf{P}} = \hat{\mu} + \sum_{i=1}^{d} \hat{\mathbf{M}}_k + (\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\hat{\mathbf{E}}$. Moreover, denote the elements in $\check{\mathbf{P}}$ by $\check{P}_{\mathbf{s}}$, $\mathbf{s} \in \Omega$, and define

$$\check{m}_k(\mathbf{x}_k) = \frac{1}{n}\sum_{\mathbf{s}\in\Omega} \check{m}(\mathbf{X}_{\mathbf{s}}^{(k)}) - \frac{1}{n}\sum_{\mathbf{s}\in\Omega} \check{P}_{\mathbf{s}},$$

$$\check{m}(\mathbf{X}_{\mathbf{s}}^{(k)}) = \eta^T\Big[\vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)})\, \vec{\mathbf{X}}(\mathbf{X}_{\mathbf{s}}^{(k)})\Big]^{-1} \vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)})\, \check{\mathbf{P}}.$$

Then, we see that $\check{m}_k(\mathbf{x}_k) - m_k(\mathbf{x}_k) = \Big(\tilde{m}_k(\mathbf{x}_k) - \check{m}_k(\mathbf{x}_k)\Big) + \Big(\check{m}_k(\mathbf{x}_k) - m_k(\mathbf{x}_k)\Big)$. To establish Theorem 3.3, we will show that $\tilde{m}_k(\mathbf{x}_k) - \check{m}_k(\mathbf{x}_k) = o_p((ng_{\pi k})^{-1/2})$ and $(ng_{\pi k})^{-1/2}(\check{m}_k(\mathbf{x}_k) -$

$m_k(\mathbf{x}_k)$) is asymptotically normal.

We first consider the former. From Theorem 3.1 we have $\hat{\mathbf{M}}_k = \mathbf{M}_k + O_p(h_{kL}^2)$, together with $\hat{\mu} = \mu + O_p(n^{-1/2})$ and assumption (B8), we have $\hat{\mathbf{E}} = \mathbf{E} + O_p(h_L^2)$. Define $\mathbf{G}$ as a vector with the i-th element being $G_i = \sum_{j \in N(\mathbf{s}_i)} \epsilon_{\mathbf{s}_j}$, where $N(\mathbf{s}_i)$ is the neighborhood set of $\epsilon_{\mathbf{s}_i}$. Also note that there are finite nonzero elements in each row of $\mathbf{B}(\boldsymbol{\theta})$ and $\mathbf{B}(\hat{\boldsymbol{\theta}})$. Then

$$
\begin{aligned}
&\tilde{m}_k(\mathbf{x}_k) - \check{m}_k(\mathbf{x}_k) \\
&= \frac{1}{n} \sum_{\mathbf{s}\in\Omega} \Big[\tilde{m}(\mathbf{X}_{\mathbf{s}}^{(k)}) - \check{m}_k(\mathbf{X}_{\mathbf{s}}^{(k)}) \Big] - \frac{1}{n} \mathbf{1}^T (\hat{\mathbf{P}} - \check{\mathbf{P}}) \\
&= \frac{1}{n} \sum_{\mathbf{s}\in\Omega} \eta^T \Big[\vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)})\, \vec{\mathbf{X}}(\mathbf{X}_{\mathbf{s}}^{(k)}) \Big]^{-1} \vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)}) (B(\boldsymbol{\theta}) - B(\hat{\boldsymbol{\theta}})) \hat{\mathbf{E}} - \frac{1}{n} \mathbf{1}^T (B(\boldsymbol{\theta}) - B(\hat{\boldsymbol{\theta}}))\, \hat{\mathbf{E}} \\
&= \frac{1}{n} \eta^T \Big(f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)}) + o_p(1) \Big) \vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)}) \Big(\mathbf{G} + O_p(h_L^2) \Big) \cdot o_p(1) - \frac{1}{n} \mathbf{1}^T \Big(\mathbf{G} + O_p(h_L^2) \Big) \cdot o_p(1) \\
&= \frac{1}{n} \sum_{\mathbf{s}\in\Omega} \eta^T f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)}) \vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)}) \mathbf{G} \cdot o_p(1) + \frac{1}{n} \sum_{\mathbf{s}\in\Omega} \eta^T f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)}) \vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\, W(\mathbf{X}_{\mathbf{s}}^{(k)}) \mathbf{1} \cdot o_p(h_L^2) \\
&\quad - \frac{1}{n} \mathbf{1}^T \mathbf{G} \cdot o_p(1) + o_p(h_L^2) \\
&\hat{=} I_{11} \cdot o_p(1) + I_{12} \cdot o_p(h_L^2) - I_{13} \cdot o_p(1) + o_p(h_L^2).
\end{aligned} \tag{3.42}
$$

Via assumption (B8) and the equation above, it is sufficient for us to show $\tilde{m}_k(\mathbf{x}_k) - \check{m}_k(\mathbf{x}_k) = o_p((ng_{\pi k})^{-1/2})$ by proving $I_{11} = O_p\Big((ng_{\pi k})^{-1/2}\Big)$, $I_{12} = O_p\Big((ng_{\pi k})^{-1/2}\Big)$ and $I_{13} = O_p(n^{-1/2})$. As the proof about I_{11} is most involved, we will omit the proofs for the others in the following. Recall $C_{\mathbf{t}k}$ defined in (3.27), and $C_{\mathbf{t}k} = f^{-1}(\mathbf{X}_{\mathbf{t}}^{(k)}) f_{-k}(\mathbf{X}_{\mathbf{t}}^{(-k)}) + o_p(g_{kL})$, we

then have

$$\begin{aligned} I_{11} &= \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\sum_{\mathbf{t}\in\Omega}(ng_\pi)^{-1}f^{-1}(\mathbf{X}_\mathbf{s}^{(k)})\,K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_\mathbf{t}-\mathbf{X}_\mathbf{s}^{(k)})\Big)G_\mathbf{t} \\ &= (ng_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega}K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\,C_{\mathbf{t}k}\,G_\mathbf{t} \\ &= (ng_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega}f^{-1}(\mathbf{X}_\mathbf{t}^{(k)})f_{-k}(\mathbf{X}_\mathbf{t}^{(-k)})K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\,G_\mathbf{t} \\ &\quad +(ng_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega}K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\,G_\mathbf{t}\cdot o_p(g_{kL}) \\ &\hat{=}\ I_{111}+I_{112}\cdot o_p(g_{kL}). \end{aligned} \tag{3.43}$$

Clearly,

$$I_{111} = (ng_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega}\sum_{\mathbf{r}\in N(\mathbf{t})}f^{-1}(\mathbf{X}_\mathbf{t}^{(k)})f_{-k}(\mathbf{X}_\mathbf{t}^{(-k)})K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\epsilon_\mathbf{r} \tag{3.44}$$

has zero mean, and its variance can be given as

$$\begin{aligned} Var(I_{111}) = (ng_{\pi k})^{-2}&\sum_{\mathbf{t}\in\Omega}\sum_{\mathbf{r}'\in N(\mathbf{t})}\sum_{\mathbf{r}''\in N(\mathbf{t})}E\Big[f^{-2}(\mathbf{X}_\mathbf{t}^{(k)})f_{-k}^2(\mathbf{X}_\mathbf{t}^{(-k)})K_\pi^2\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\Big]\,Cov(\epsilon_{\mathbf{r}'},\epsilon_{\mathbf{r}''}) \\ &+(ng_{\pi k})^{-2}E^2\Big[f^{-1}(\mathbf{X}_\mathbf{t}^{(k)})f_{-k}(\mathbf{X}_\mathbf{t}^{(-k)})K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\Big]\sum_{\substack{\mathbf{t}'\neq\mathbf{t}''\\ \mathbf{t}',\mathbf{t}''\in\Omega}}\sum_{\mathbf{r}'\in N(\mathbf{t}')}\sum_{\mathbf{r}''\in N(\mathbf{t}'')}Cov(\epsilon_{\mathbf{r}'},\epsilon_{\mathbf{r}''}). \end{aligned} \tag{3.45}$$

Let $\mathbf{u} = \Big(\mathbf{u}_k^T, (\mathbf{u}^{(-k)})^T\Big)^T$, where $\mathbf{u}_k$ is a b_k dimensional random vector and $\mathbf{u}^{(-k)}$ is a $D-b_k$ dimensional random vector . We can show that

$$\begin{aligned} &E\Big[f^{-2}(\mathbf{X}_\mathbf{t}^{(k)})\,f_{-k}^2(\mathbf{X}_\mathbf{t}^{(-k)})\,K_\pi^2\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\Big] \\ =\ &\int f^{-2}(\mathbf{x}_k,\mathbf{u}^{(-k)})\,f_{-k}^2(\mathbf{u}^{(-k)})\,K_\pi^2\Big(\boldsymbol{\zeta}_1(\mathbf{u}_k-\mathbf{x}_k)\Big)f(\mathbf{u})\,d\mathbf{u} \\ =\ &g_{\pi k}\int f^{-2}(\mathbf{x}_k,\mathbf{u}^{(-k)})\,f_{-k}^2(\mathbf{u}^{(-k)})\int K_\pi^2(\boldsymbol{\delta})f(\mathbf{x}_k+\boldsymbol{\delta}\odot\mathbf{g}_k,\mathbf{u}^{(-k)})\,d\boldsymbol{\delta}\,d\mathbf{u}^{(-k)} \\ =\ &g_{\pi k}\int f^{-1}(\mathbf{u}^{(k)})f_{-k}^2(\mathbf{u}^{(-k)})\,d\mathbf{u}^{(-k)}\int K_\pi^2(\boldsymbol{\delta})\,d\boldsymbol{\delta}+o(g_{\pi k}), \end{aligned} \tag{3.46}$$

which is equal to $O(g_{\pi k})$. Similarly, we have $E\Big[f^{-1}(\mathbf{X}_\mathbf{t}^{(k)})f_{-k}(\mathbf{X}_\mathbf{t}^{(-k)})K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\Big]=$

$O(g_{\pi k})$. Therefore, due to the fact that there are finite elements in each set $N(\mathbf{t})$, we obtain

$$\begin{aligned} Var(I_{111}) &\le Cn^{-2}g_{\pi k}^{-1}\sum_{\mathbf{t}\in\Omega}\sum_{\mathbf{r}'\in N(\mathbf{t})}\sum_{\mathbf{r}''\in N(\mathbf{t})} Cov(\epsilon_{\mathbf{r}'},\epsilon_{\mathbf{r}''}) + Cn^{-2}\sum_{\substack{\mathbf{t}'\neq\mathbf{t}''\\ \mathbf{t}',\mathbf{t}''\in\Omega}}\sum_{\mathbf{r}'\in N(\mathbf{t}')}\sum_{\mathbf{r}''\in N(\mathbf{t}'')} Cov(\epsilon_{\mathbf{r}'},\epsilon_{\mathbf{r}''}) \\ &\le C(ng_{\pi k})^{-1}\sigma_\epsilon^2 + Cn^{-1}\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\left|R(k_1,k_2)\right|, \end{aligned} \tag{3.47}$$

where $R(\mathbf{k}) = Cov(\epsilon_{\mathbf{s}},\epsilon_{\mathbf{s}+\mathbf{k}})$ for $\mathbf{k} = (k_1,k_2) \in \mathbb{Z}^2$. By Lemma 2.3 or Lemma 2.5, we obtain $Var(I_{111}) = O\Big((ng_{\pi k})^{-1}\Big)$, and consequently $I_{111} = O_p\Big((ng_{\pi k})^{-1/2}\Big)$.

With the same procedure as (3.45)-(3.47), we can also show $I_{112} = O_p\Big((ng_{\pi k})^{-1/2}\Big)$. Then, by (3.43), we can conclude $I_{11} = O_p\Big((ng_{\pi k})^{-1/2}\Big)$. With similar technique and procedure, we can show $I_{12} = O_p\Big((ng_{\pi k})^{-1/2}\Big)$ and $I_{13} = O_p(n^{-1/2})$. The proof of $\tilde{m}_k(\mathbf{x}_k) - \check{m}_k(\mathbf{x}_k) = o_p((ng_{\pi k})^{-1/2})$ is then completed.

We go on to consider the asymptotic normality of $\check{m}_k(\mathbf{x}_k)$. Define $\check{\mathbf{F}}(\mathbf{X}_{\mathbf{s}}^{(k)}) = \Big(\frac{1}{n}\sum_{\mathbf{r}\in\Omega}\check{P}_{\mathbf{r}} + m(\mathbf{X}_{\mathbf{s}}^{(k)}),\ (m'(\mathbf{X}_{\mathbf{s}}^{(k)})\odot\mathbf{g})^T\Big)^T$, then

$$\begin{aligned} &\check{m}_k(\mathbf{x}_k) - m_k(\mathbf{x}_k) \\ =\ & \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\eta^T\Big(\vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\,\mathbf{W}(\mathbf{X}_{\mathbf{s}}^{(k)})\,\vec{\mathbf{X}}(\mathbf{X}_{\mathbf{s}}^{(k)})\Big)^{-1}\vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\,\mathbf{W}(\mathbf{X}_{\mathbf{s}}^{(k)})\Big[\check{\mathbf{P}} - \vec{\mathbf{X}}(\mathbf{X}_{\mathbf{s}}^{(k)})\check{\mathbf{F}}(\mathbf{X}_{\mathbf{s}}^{(k)})\Big] + O_p(n^{-1/2}) \\ =\ & \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\eta^T\Big(f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)}) + o_p(1)\Big)\vec{\mathbf{X}}^T(\mathbf{X}_{\mathbf{s}}^{(k)})\,\mathbf{W}(\mathbf{X}_{\mathbf{s}}^{(k)})\Big[\check{\mathbf{P}} - \vec{\mathbf{X}}(\mathbf{X}_{\mathbf{s}}^{(k)})\check{\mathbf{F}}(\mathbf{X}_{\mathbf{s}}^{(k)})\Big] + O_p(n^{-1/2}) \\ =\ & \Bigg[\frac{1}{n}\sum_{\mathbf{s}\in\Omega}\sum_{\mathbf{t}\in\Omega}(ng_\pi)^{-1}f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})K_\pi\Big(\zeta_1(\mathbf{X}_{\mathbf{t}} - \mathbf{X}_{\mathbf{s}}^{(k)})\Big)\cdot O_p(h_L^2) \\ &+ \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\sum_{\mathbf{t}\in\Omega}(ng_\pi)^{-1}f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})K_\pi\Big(\zeta_1(\mathbf{X}_{\mathbf{t}} - \mathbf{X}_{\mathbf{s}}^{(k)})\Big)\mathbf{R}_{\mathbf{t}} \\ &+ \frac{1}{n}\sum_{\mathbf{s}\in\Omega}\sum_{\mathbf{t}\in\Omega}(ng_\pi)^{-1}f^{-1}(\mathbf{X}_{\mathbf{s}}^{(k)})K_\pi\Big(\zeta_1(\mathbf{X}_{\mathbf{t}} - \mathbf{X}_{\mathbf{s}}^{(k)})\Big)\tau_{\mathbf{t}}\Bigg](1 + o_p(1)) + O_p(n^{-1/2}) \\ \hat{=}\ & \Big[I_{21}\cdot O_p(h_L^2) + I_{22} + I_{23}\Big](1 + o_p(1)) + O_p(n^{-1/2}), \end{aligned} \tag{3.48}$$

where $\mathbf{R_t}$ is defined in (3.18). As $(ng_\pi)^{-1}\sum_{\mathbf{t}\in\Omega} K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X_t} - \mathbf{X_s^{(k)}})\Big) = f(\mathbf{X_s^{(k)}}) + O_p(g_{kL})$, it is readily to see that $I_{21} = O_p(1)$. By (3.20) and (3.22)-(3.26), we can show

$$I_{22} = \frac{1}{2}tr\Big[m_k''(\mathbf{x}_k)\int K_\pi(\boldsymbol{\delta})(\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T)\odot(\mathbf{g}_k\mathbf{g}_k^T)\,d\boldsymbol{\delta}\Big] + o_p(g_{kL}^2). \tag{3.49}$$

Then, by (3.48) and assumption (B8), it remains for us to show that $(ng_{\pi k})^{1/2}I_{23} \xrightarrow{D} N\Big(0,\ \Sigma_\tau(\mathbf{x}_k)\Big)$. Consider that

$$\begin{aligned} I_{23} &= (ng_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega} K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}} - \mathbf{x}_k)\Big)C_{\mathbf{t}k}\tau_{\mathbf{t}} \\ &= (ng_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega} f^{-1}(\mathbf{X_t^{(k)}})f_{-k}(\mathbf{X_t^{(-k)}})K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\tau_{\mathbf{t}} \\ &\quad +\Big[(ng_{\pi k})^{-1}\sum_{\mathbf{t}\in\Omega} K_\pi\Big(\boldsymbol{\zeta}_1(\mathbf{X}_{k\mathbf{t}}-\mathbf{x}_k)\Big)\tau_{\mathbf{t}}\Big]\cdot o_p(g_{kL}) \\ &\hat{=}\ I_{231} + I_{232}\cdot o_p(g_{kL}) \end{aligned} \tag{3.50}$$

It is easy to show that the variance of I_{232} is equal to $O\Big((ng_{\pi k})^{-1}\Big)$, and its mean is zero. This indicates $I_{232} = O_p\Big((ng_{\pi k})^{-1/2}\Big)$. Moreover, since $\{\mathbf{X_s}\}$ and $\{\tau_{\mathbf{s}}\}$ are *i.i.d.*, and mutually independent, the asymptotic normality of I_{231} obviously hold. We can show that I_{231} has zero mean, and

$$Var(I_{231}) = (ng_{\pi k})^{-1}\sigma_\tau^2\int f^{-1}(\mathbf{u}^{(k)})f_{-k}^2(\mathbf{u}^{(-k)})\,d\mathbf{u}^{(-k)}\int K_\pi^2(\boldsymbol{\delta})d\boldsymbol{\delta} + o\Big((ng_{\pi k})^{-1}\Big),$$

where $\boldsymbol{\delta}$ is a b_k- dimensional vector. This conclude the proof of Theorem 3.3. □

3.6 Summary and Remark

In this chapter, we focus on the nonparametric additive model, where the errors are assumed to satisfy the torus SGAR model, separable SGAR model and unilateral SGAR

model. The two-step estimation procedure introduced in Martins-Filho & Yao (2009) is adopted. For the first-step estimator, we show its convergence rate with the three types of error structures, and establish the asymptotic normality when ϵ_s satisfies the separable and unilateral SGAR model. For the second-step estimator, the asymptotic normality with all three error structures is established. Simulations are conducted to assess the performance of our estimation. It can be seen that our two-step estimation performs better than the original estimation of additive model with the marginal integration technique. Furthermore, comparing the fitting results presented in Chapter 2 and Chapter 3, we find that the two-step estimation with additive model achieves better results than the one with general nonparametric model in Chapter 2, given that the testing model has additive structure.

Chapter 4

Partially Linear Model with SGAR-type Error

4.1 Literature Review

Nonparametric methods are powerful tools for nonlinear regression and nonlinear time series analysis. However, such methods always face a serious problem, the curse of dimensionality. Because of this, only a few explanatory variables can be introduced into the model, unless the amount of the observations is extremely large. On the other hand, although the parametric models do not suffer the curse of dimensionality, the assumption of linearity is usually unreasonable in practical research. Therefore, as a combination of nonparametric and parametric models, the partially linear model has been broadly considered in recent years. A great deal of literature focuses on this type of model, for instance, Heckman (1986), Rice (1986), Chen (1988), Robinson (1988), Speckman (1988), Chen & Shiau (1991, 1994) and Hamilton (1997). Several kinds of

methods are used in these literature, such as the spline method, the kernel method and the local polynomial estimation.

The *i.i.d*. assumption of the errors is usually applied in the partially linear model, including the literature mentioned above. However, this assumption is not always appropriate, especially with financial data where serial dependency of errors usually exists. So in some literature the partially linear model with correlated errors is also studied. For example, Engle et al.(1986) and Schick (1994, 1996, 1998) considered the models with AR-type errors; Sun et al. (2002) investigated the convergence rates of both the parametric and nonparametric estimators with MA(∞) errors and fixedly designed regressors; Lu & Gijbels (2001) studied the consistency of the estimators in the partially linear regression model with dependent observations and ARCH-type errors, and local polynomial fitting was applied.

In the area of spatial process analysis, the nonparametric methods are less considered. As mentioned in Gao et al. (2006), this is mainly due to the curse of dimensionality. On a lattice, a nonparametric fitting of spatial data given its closest neighbors requires a four-dimensional model. Obviously, the amount of spatial data is seldomly adequate enough for such study. Therefore, Gao et al. (2006) considered an estimation of the partially linear model by combining the marginal integration technic with local linear fitting. They assumed additive structure for the nonparametric function in the model, so that the problem of the curse of dimensionality becomes less severe. Different from the model with correlated errors mentioned above, they considered the case with dependent observations instead. Under some spatially mixing condition, the asymptotic normality of the estimator was established.

In this chapter, we also consider the partially linear model with spatial data. Different from the idea of Gao et al. (2006), we consider the model with spatially correlated errors. By treating the spatial dependency with a parametric model for the errors, we can greatly avoid the curse of dimensionality. The two-step estimation idea will be used again.

The outline of this chapter is as follow. The proposed model as well as the estimation procedure are introduced in the second section. Asymptotic properties of the estimators are given in the third section. Simulations are conducted to show the performance of our method, and the results are shown in section four. Proofs of the theoretical results are given thereafter, and at last is a short conclusion.

4.2 Model and Estimation

The model with interest is

$$Y_{\mathbf{s}} = \mathbf{Z}_{\mathbf{s}}^T \boldsymbol{\lambda} + m(\mathbf{X}_{\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{4.1}$$

where $\mathbf{s} = (s_1, s_2) \in \mathbb{Z}^2$ is the index of position from a $n_1 \times n_2$ rectangular lattice Ω, $m(.)$ is some unknown function, $\boldsymbol{\lambda}$ is a b-dimensional parameter vector, $Y_{\mathbf{s}}$ is the response variable, $\mathbf{Z}_{\mathbf{s}}$ and $\mathbf{X}_{\mathbf{s}}$ are the explanatory vectors being b-dimensional and d-dimensional respectively, $\{\mathbf{X}_{\mathbf{s}}\}$ is independent of $\{\epsilon_{\mathbf{s}}\}$, and $\epsilon_{\mathbf{s}}$ satisfy the torus, separable and unilateral SGAR models defined in (2.2)-(2.4).

Arrange the points $\mathbf{s} \in \Omega$ in arbitrage order and denote them by $\mathbf{s}_1, \cdots, \mathbf{s}_n$ respectively. Then our model can be expressed in the matrix form as

$$\mathbf{Y} = \vec{\mathbf{Z}}\boldsymbol{\lambda} + \mathbf{M} + \mathbf{E}, \tag{4.2}$$

$$(\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\mathbf{E} = \Upsilon, \tag{4.3}$$

where $\mathbf{Y} = (Y_{\mathbf{s}_1}, \cdots, Y_{\mathbf{s}_n})$, $\vec{\mathbf{Z}} = (\mathbf{Z}_{\mathbf{s}_1}, \cdots, \mathbf{Z}_{\mathbf{s}_n})^T$, $\mathbf{M} = (m(\mathbf{X}_{\mathbf{s}_1}), \cdots, m(\mathbf{X}_{\mathbf{s}_n}))$, $\mathbf{E} = (\epsilon_{\mathbf{s}_1}, \cdots, \epsilon_{\mathbf{s}_n})$, $\mathbf{B}(\boldsymbol{\theta})$ is the matrix made up of the SGAR coefficients, and is defined in the same manner as in Chapter 2.

The purpose of our method is to estimate $\boldsymbol{\lambda}$ and $m(.)$, given that $\epsilon_{\mathbf{s}}$ is spatially correlated. Following the previous research, we first assume that $\boldsymbol{\lambda}$ is known. Let $\mathbf{h} = (h_1, \cdots, h_d)^T$, h_i is some bandwidth, and $\boldsymbol{\beta}(\mathbf{x}) = \left(m(\mathbf{x}), (\mathbf{h} \odot m'(\mathbf{x}))^T\right)^T$. Then, by local linear fitting, $\boldsymbol{\beta}(\mathbf{x})$ can be estimated as

$$\hat{\boldsymbol{\beta}}(\mathbf{x}, \boldsymbol{\lambda}) = \left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x})\right)^{-1}\left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\mathbf{Y}_\lambda\right). \tag{4.4}$$

Here, $\mathbf{Y}_\lambda$ is a n-dimensional vector with the i-th element being $Y_{\mathbf{s}_i} - Z_{\mathbf{s}_i}^T\boldsymbol{\lambda}$, $\vec{\mathbf{X}}(\mathbf{x})$ is a $n \times (d+1)$ matrix with the i-th row being $\boldsymbol{\zeta}(\mathbf{X}_{\mathbf{s}_i} - \mathbf{x}) = \left(1, \boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_i} - \mathbf{x})\right)$ and $\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_i} - \mathbf{x}) = ((\mathbf{X}_{\mathbf{s}_i} - \mathbf{x}) \odot (1/\mathbf{h}))^T$, $\mathbf{W} = (nh_\pi)^{-1}\mathrm{diag}\left(K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_1} - \mathbf{x})), \cdots, K_\pi(\boldsymbol{\zeta}_1(\mathbf{X}_{\mathbf{s}_n} - \mathbf{x}))\right)$ with $h_\pi = \prod_{i=1}^{d} h_i$ and $K_\pi(\boldsymbol{\delta}) = \prod_{i=1}^{d} K(\delta_i)$, $\boldsymbol{\delta} = (\delta_1, \cdots, \boldsymbol{\delta}_d)^T$.

As we can express $\mathbf{Y}_\lambda$ as $\mathbf{Y}_\lambda = \mathbf{Y} - \vec{\mathbf{Z}}\boldsymbol{\lambda}$, an estimator of $m(\mathbf{x})$ can then be given by

$$\begin{aligned}\hat{m}(\mathbf{x},\boldsymbol{\lambda}) &= \eta^T\left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x})\right)^{-1}\left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\mathbf{Y}\right) - \eta^T\left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x})\right)^{-1}\left(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{Z}}\right)\boldsymbol{\lambda} \\ &\hat{=} \hat{Q}_{\mathbf{Y}}(\mathbf{x}) - \hat{Q}_{\mathbf{Z}}^T(\mathbf{x})\,\boldsymbol{\lambda},\end{aligned} \tag{4.5}$$

where η is a $(d+1)$-dimensional vector with the first element being 1 and 0 otherwise. Note that if we take conditional expectation of model (4.1), given $\mathbf{X}_{\mathbf{s}} = \mathbf{x}$, we have

$$E(Y_{\mathbf{s}} \mid \mathbf{X}_{\mathbf{s}} = \mathbf{x}) = E(\mathbf{Z}_{\mathbf{s}}^T \mid \mathbf{X}_{\mathbf{s}} = \mathbf{x})\boldsymbol{\lambda} + m(\mathbf{x}). \tag{4.6}$$

Marching (4.5) with (4.6), we can see that $\hat{Q}_{\mathbf{Y}}(\mathbf{x})$ is the local linear estimator of $E(Y_{\mathbf{s}} \,|\, \mathbf{X}_{\mathbf{s}} = \mathbf{x})$, and $\hat{Q}_{\mathbf{Z}}(\mathbf{x})$ is the local linear estimator of $E(\,\mathbf{Z}_{\mathbf{s}} \,|\, \mathbf{X}_{\mathbf{s}} = \mathbf{x})$. For convenience, denote $Q_{\mathbf{Y}}(\mathbf{x}) = E(Y_{\mathbf{s}} \,|\, \mathbf{X}_{\mathbf{s}} = \mathbf{x})$ and $Q_{\mathbf{Z}}(\mathbf{x}) = E(\,\mathbf{Z}_{\mathbf{s}} \,|\, \mathbf{X}_{\mathbf{s}} = \mathbf{x})$.

Now we turn back to estimate $\boldsymbol{\lambda}$. Define $C_{\mathbf{s}} = Y_{\mathbf{s}} - \hat{Q}_{\mathbf{Y}}(\mathbf{X}_{\mathbf{s}})$ and $\mathbf{D}_{\mathbf{s}} = \mathbf{Z}_{\mathbf{s}} - \hat{Q}_{\mathbf{Z}}(\mathbf{X}_{\mathbf{s}})$. As $\hat{m}(\mathbf{x},\boldsymbol{\lambda})$ is linear in $\boldsymbol{\lambda}$, an estimator of $\boldsymbol{\lambda}$ can then be given by

$$\underset{\lambda \in R^b}{\operatorname{argmin}} \sum_{\mathbf{s}\in\Omega} \Big(Y_{\mathbf{s}} - \mathbf{Z}_{\mathbf{s}}^T \boldsymbol{\lambda} - \hat{m}(\mathbf{X}_{\mathbf{s}}, \boldsymbol{\lambda})\Big)^2 = \underset{\lambda \in R^b}{\operatorname{argmin}} \sum_{\mathbf{s}\in\Omega} \Big(C_{\mathbf{s}} - \mathbf{D}_{\mathbf{s}}^T \boldsymbol{\lambda}\Big)^2.$$

Therefore, we have

$$\hat{\boldsymbol{\lambda}} = (\vec{\mathbf{D}}^T \vec{\mathbf{D}})^{-1} (\vec{\mathbf{D}}^T \mathbf{C}), \tag{4.7}$$

where $\vec{\mathbf{D}} = (\mathbf{D}_{\mathbf{s}_1}, \ldots, \mathbf{D}_{\mathbf{s}_n})^T$ and $\mathbf{C} = (C_{\mathbf{s}_1}, \ldots, C_{\mathbf{s}_n})^T$. Then the estimator of $m(\mathbf{x})$ can be given by

$$\hat{m}(\mathbf{x}) = \hat{Q}_Y(\mathbf{x}) - \hat{Q}_{\mathbf{Z}}^T(\mathbf{x}) \hat{\boldsymbol{\lambda}}. \tag{4.8}$$

So far what we have done is the standard estimation procedure for partially linear model. Notice that the error $\{\epsilon_{\mathbf{s}}\}$ is not *i.i.d.* but spatially correlated. However, we have not made use of the information of such spatial dependency. Comparing with (4.2), if we form a new process $\{P_{\mathbf{s}}\}$ as

$$\mathbf{P} = \vec{\mathbf{Z}} \boldsymbol{\lambda} + \mathbf{M} + (\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\mathbf{E}\,, \tag{4.9}$$

where $\mathbf{P} = (P_{\mathbf{s}_1}, \cdots, P_{\mathbf{s}_n})^T$, then $\mathbf{P}$ has the same conditional mean as $\mathbf{Y}$, but $\mathbf{P}$ has *i.i.d.* errors instead. Therefore, we can apply the partially linear estimation demonstrated above to $\mathbf{P}$ to obtain some improved estimators of $\boldsymbol{\lambda}$ and $m(\mathbf{x})$. However, as $\boldsymbol{\lambda}$, $\mathbf{M}$, $\mathbf{B}(\boldsymbol{\theta})$ and $\mathbf{E}$ are all unknown, we may substitute some consistent estimators for them. It will be shown in the next section that $\hat{\boldsymbol{\lambda}}$ and $\hat{m}(\mathbf{x})$ given in (4.7) and (4.8) are consistent,

consequently $\epsilon_{\mathbf{s}}$ may be estimated consistently as well. Therefore, with some proper estimator of $\mathbf{B}(\boldsymbol{\theta})$, we can form

$$\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{B}(\hat{\boldsymbol{\theta}}))\mathbf{Y} + \mathbf{B}(\hat{\boldsymbol{\theta}})(\vec{\mathbf{Z}}\hat{\boldsymbol{\lambda}} + \hat{\mathbf{M}}). \tag{4.10}$$

$\hat{\mathbf{P}} = (\hat{P}_{\mathbf{s}_1}, \ldots, \hat{P}_{\mathbf{s}_n})^T$. It then remains for us to conduct the partially linear fitting to $\hat{\mathbf{P}}$. Let $L_{\mathbf{s}} = \hat{P}_{\mathbf{s}} - \hat{Q}_{\mathbf{P}}(\mathbf{X}_{\mathbf{s}})$, where $\hat{Q}_{\mathbf{P}}(\mathbf{X}_{\mathbf{s}}) = \eta^T(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x}))^{-1}(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\hat{\mathbf{P}})$, and $\mathbf{L} = (L_{\mathbf{s}_1}, \ldots, L_{\mathbf{s}_n})^T$. Moreover, let $\mathbf{g} = (g_1, \cdots, g_d)^T$ be the bandwidth vector for the second-step estimation, and consequently $g_\pi = \prod_{i=1}^{d} g_i$. Substitute $\mathbf{h}$ and h_π by $\mathbf{g}$ and g_π in the expression of $\vec{\mathbf{X}}(\mathbf{x})$ and $\mathbf{W}(\mathbf{x})$, then the improved estimators of $\boldsymbol{\lambda}$ and $m(\mathbf{x})$ are given by

$$\tilde{\boldsymbol{\lambda}} = (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}\vec{\mathbf{D}}^T\mathbf{L}, \tag{4.11}$$

$$\tilde{m}(\mathbf{x}) = \hat{Q}_{\mathbf{P}}(\mathbf{x}) - \hat{Q}_{\mathbf{Z}}^T(\mathbf{x})\tilde{\boldsymbol{\lambda}}. \tag{4.12}$$

Now let us sum up the whole estimation procedure, which we denote as PLR-SCE throughout this thesis.

Step 1. Obtain the initial estimators of $\boldsymbol{\lambda}$ and $m(\mathbf{x})$ by (4.7) and (4.8). As such estimators are consistent, $\epsilon_{\mathbf{s}}$ may be estimated consistently, too. Hence some appropriate estimator of $\mathbf{B}(\boldsymbol{\theta})$ is also available.

Step 2. Form $\hat{\mathbf{P}}$ defined in (4.10), then apply the partially linear fitting to the new process, so that the improved estimators given in (4.11) and (4.12) can be obtained.

4.3 Theoretical Results

In this section we mainly study the asymptotic properties of the first- and second-step estimators of $\boldsymbol{\lambda}$ and $m(\mathbf{x})$, which are given in (4.7)-(4.8) and (4.11)-(4.12) respectively. Several assumptions are in order.

(C1) The random vectors $\mathbf{X_s}, \mathbf{s} \in \Omega$, are *i.i.d.* with joint density $f(\mathbf{x})$, $|f(\mathbf{x})| < \infty$. Moreover, $\{\mathbf{X_s}\}$ is independent of $\{\epsilon_\mathbf{s}\}$ and $\{\tau_\mathbf{s}\}$, where $\tau_\mathbf{s}$ is defined in (2.2)-(2.4).

(C2) The second derivatives of $m(.)$ exist and are continuous at all $\mathbf{x}$.

(C3) $\mathbf{Z_s} = Q_\mathbf{Z}(\mathbf{X_s}) + \sigma_\mathbf{Z}\, \mathbf{e_s}$, where $Q_\mathbf{Z}(.) = \Big(Q_1(.), \ldots, Q_b(.)\Big)$, $\mathbf{e_s}$ is a *i.i.d.* random vector with zero mean and unit covariance matrix, $\{\mathbf{e_s}\}$ is independent of $\{\mathbf{X_s}\}$, $\{\epsilon_\mathbf{s}\}$ and $\{\tau_\mathbf{s}\}$.

(C4) The second derivatives of $Q_i(.)$, $i = 1, \cdots, b$, exist and continuous at all $\mathbf{x}$.

(C5) The kernel function $K(.)$ is symmetric, with bounded support, and Lipschitz continuous.

(C6) The random field $\{\epsilon_\mathbf{s}, \mathbf{s} \in \mathbb{Z}^2\}$ is strictly stationary; $\epsilon_\mathbf{s}$ has zero mean and finite variance.

(C6a) The coefficients in (2.2) satisfy $|\theta_1| + |\theta_2| < 1/2$.

(C6b) The coefficients in (2.3) satisfy $|\theta_1| < 1/2$, $|\theta_2| < 1/2$.

(C6c) The coefficients in (2.4) satisfy $|\theta_1| + |\theta_2| < 1$.

(C7) $h_i > 0, h_i \to 0$ as $\mathbf{n} \to \infty$; moreover, denote h_L as the bandwidth with the slowest convergence rate, such as $h_L \in \{h_1, \cdots, h_d\}$, $h_L = O(h_k^C)$ for $0 < C \leq 1$ and $k = 1, \cdots, d$, then $\mathbf{h}$ converges to zero in the manner that $n = O((h_\pi h_L^4)^{-1})$.

(C8) $g_i > 0$, $g_i \to 0$ as $\mathbf{n} \to \infty$; denote g_L the bandwidth with the slowest convergence rate, such as $g_L \in \{g_1, \cdots, g_d\}, g_L = O(g_k^C)$ for $0 < C \leq 1$ and $k = 1, \cdots, d$; $\mathbf{g}$ converges to zero in the manner that $n = O((g_\pi g_L^4)^{-1})$; moreover, $\lim_{\mathbf{n}\to\infty} h_L/g_L \to 0$.

(C9) There exist some sequences as: $l_1 \to \infty$, $l_2 \to \infty$ and $m \to \infty$ as $\mathbf{n} \to \infty$; $m/l_i \to 0$ and $l_i/n_i \to 0$ for $i = 1, 2$, as $\mathbf{n} \to \infty$.

We first consider the asymptotic properties of the first-step estimators $\hat{\boldsymbol{\lambda}}$ and $\hat{m}(\mathbf{x})$.

Theorem 4.1 *If $\epsilon_\mathbf{s}$ follows any model in (2.2)-(2.4), with coefficients satisfying assumptions* (C6a)-(C6c) *respectively, and assumptions* (C1)-(C7) *hold, then*

$$n^{1/2}(\,\hat{\boldsymbol{\lambda}} - \lambda\,) \xrightarrow{D} N\Big(0,\, \sigma_\epsilon^2(\sigma_\mathbf{Z}\sigma_\mathbf{Z}^T)^{-1}\Big).$$

Recall $\mathbf{U}_\mathbf{n}(\mathbf{x})$ and $\mathbf{V}_\mathbf{n}^*(\mathbf{x})$ defined in (2.15)-(2.16), and then the following theorem can be founded.

Theorem 4.2 *If the conditions in Theorem 4.1 hold, then*

$$\hat{m}(\mathbf{x}) - m(\mathbf{x}) - \text{Bias}(\mathbf{x}, \mathbf{h}) = \eta^T \mathbf{U}_\mathbf{n}^{-1}(\mathbf{x})\mathbf{V}_\mathbf{n}^*(\mathbf{x}) + o_p(h_L^2), \tag{4.13}$$

where $\text{Bias}(\mathbf{x}, \mathbf{h}) = \frac{1}{2}\,\text{tr}\Big[m''(\mathbf{x})\int(\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{h}\mathbf{h}^T)\, K_\pi(\boldsymbol{\delta})\, d\boldsymbol{\delta}\Big]$, *$\eta$ is a vector with the first element valued 1 and 0 elsewise. Furthermore, as* $\text{Bias}(\mathbf{x}, \mathbf{h}) = O(h_L^2)$ *and* $\mathbf{U}_\mathbf{n}^{-1}(\mathbf{x})\mathbf{V}_\mathbf{n}^*(\mathbf{x}) = O_p((nh_\pi)^{-1/2})$, *we have*

$$\hat{m}(\mathbf{x}) - m(\mathbf{x}) = O_p(h_L^2). \tag{4.14}$$

When $\epsilon_\mathbf{s}$ satisfies model (2.3) or (2.4), we can also establish the asymptotic normality of $\hat{m}(\mathbf{x})$. We only focus on the case with $\epsilon_\mathbf{s}$ satisfying (2.3), whose analyses are more

involved.

Theorem 4.3 *If $\epsilon_{\mathbf{s}}$ follows model (2.3), and assumptions* (C1)-(C7), (C6b) *and* (C9) *hold, then*

$$(nh_\pi)^{1/2}\Big(\hat{m}(\mathbf{x}) - m(\mathbf{x}) - \text{Bias}(\mathbf{x}, \mathbf{h})\Big) \xrightarrow{D} N(0, \Sigma_\epsilon(\mathbf{x})), \tag{4.15}$$

where $\text{Bias}(\mathbf{x}, \mathbf{h})$ *is defined in Theorem 4.2, and* $\Sigma_\epsilon(\mathbf{x}) = f^{-1}(\mathbf{x})\,\sigma_\epsilon^2 \int K_\pi^2(\boldsymbol{\delta})\,d\boldsymbol{\delta}$.

Next, we shall establish the asymptotic normality of the second-step estimators $\tilde{\boldsymbol{\lambda}}$ and $\tilde{m}(\mathbf{x})$. Given some consistent estimator of the SGAR coefficient $\boldsymbol{\theta}$, the following two theorems can be founded.

Theorem 4.4 *If $\epsilon_{\mathbf{s}}$ follows any model in (2.2)-(2.4) with coefficients satisfying assumptions* (C6a)-(C6c) *respectively, assumptions* (C1)-(C8) *hold, and some consistent estimator of $\boldsymbol{\theta}$ is available, then*

$$n^{1/2}(\,\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\,) \xrightarrow{D} N\Big(0,\, \sigma_\tau^2(\sigma_{\mathbf{Z}}\sigma_{\mathbf{Z}}^T)^{-1}\Big). \tag{4.16}$$

Theorem 4.5 *If the conditions in Theorem 4.4 are satisfied, then we have*

$$(ng_\pi)^{1/2}\Big(\tilde{m}(\mathbf{x}) - m(\mathbf{x}) - \text{Bias}(\mathbf{x}, \mathbf{g})\Big) \xrightarrow{D} N(0, \Sigma_\tau(\mathbf{x})), \tag{4.17}$$

where $\text{Bias}(\mathbf{x}, \mathbf{g}) = \frac{1}{2}\,\text{tr}\Big[m''(\mathbf{x}) \int (\boldsymbol{\delta}\boldsymbol{\delta}^T) \odot (\mathbf{g}\mathbf{g}^T)\, K_\pi(\boldsymbol{\delta})\,d\boldsymbol{\delta}\Big]$ *and* $\Sigma_\tau(\mathbf{x}) = f^{-1}(\mathbf{x})\,\sigma_\tau^2 \int K_\pi^2(\boldsymbol{\delta})\,d\boldsymbol{\delta}$.

4.4 Simulations

To assess the performance of our approach considered in this chapter, we will conduct some simulations below. Observations $\{Y_{\mathbf{s}}, \mathbf{Z}_{\mathbf{s}}, \mathbf{X}_{\mathbf{s}}\}$ will be generated with errors $\{\epsilon_{\mathbf{s}}\}$ satisfying (2.2)-(2.4). Simulations will be conducted with 3 different sample sizes, namely

10×10, 15×15 and 20×20, and with 100 replication each. Results will be compared with those obtained by traditional partially linear fitting. For the bandwidth determination, we choose the first- and second-step bandwidth $\mathbf{h}$ and $\mathbf{g}$ iteratively. See more about the methods of bandwidth selection in Section 2.4.1.

Simulation 4.1 The testing model is

$$Y_{\mathbf{s}} = 0.5Z_{\mathbf{s}} + \sin(\pi X_{\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{4.18}$$

where $Z_{\mathbf{s}} \sim U(0, 8)$, $X_{\mathbf{s}} \sim U(0, 4)$, $\epsilon_{\mathbf{s}}$ is modeled as (2.2) -(2.4) with $\tau_{\mathbf{s}} \sim N(0, 1)$ and coefficients $\boldsymbol{\theta} = (\theta_1, \theta_2)$ being $(0.38, -0.1)$, $(0.3, -0.2)$ and $(0.4, 0.3)$ respectively.

From Table 4.1 we can see that as the sample size increases, most of the MSE of the estimators given by both the partially linear fitting and PLR-SCE decrease. This is coincident with the results that the estimators of $m(.)$ in both steps are consistent. Due to the fact that taking the spatial dependency of the errors into account, the MSE obtained by PLR-SCE are smaller, indicating that our method gains an improvement over the traditional method. Table 4.2 presents the fitting results of $\boldsymbol{\lambda}$. It shows that the estimators of $\boldsymbol{\lambda}$ given by partially linear fitting as well as PLR-SCE are close to the true values, and the results converge as the sample size increase. In Table 4.3, the estimators of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ obtained in Step 2 of the estimation are given. It can be seen that the results are reasonably good.

We are also interested to compare the results of Simulation 2.1 and 4.1. One can find that the results given in Table 2.2 and Table 4.1 are quite close. That means the additional parametric component in Simulation 4.1 does not affect the fitting results of the nonparametric section.

.

Table 4.1: Simulation 4.1. MSE for the estimators of $m(.)$.

Sample Size	10×10	15×15	20×20
Torus[(a)]			
mean of $\text{MSE}_o^{(d)}$	0.5493	0.3509	0.1750
mean of $\text{MSE}_m^{(e)}$	0.0937	0.0481	0.0320
var of MSE_o	0.1208	0.0306	0.0091
var of MSE_m	0.0022	0.0005	0.0002
Separable[(b)]			
mean of MSE_o	3.6804	5.8272	1.0422
mean of MSE_m	0.1580	0.0861	0.0353
var of MSE_o	30.1403	262.9937	0.7275
var of MSE_m	0.0131	0.0145	0.0003
Unilateral[(c)]			
mean of MSE_o	0.1087	0.0571	0.0370
mean of MSE_m	0.0943	0.0457	0.0299
var of MSE_o	0.0026	0.0007	0.0003
var of MSE_m	0.0026	0.0004	0.0001

[(a)] Torus: $\epsilon_{\mathbf{s}}$ generated by (2.2).
[(b)] Separable: $\epsilon_{\mathbf{s}}$ generated by (2.3).
[(c)] Unilateral: $\epsilon_{\mathbf{s}}$ generated by (2.4).
[(d)] MSE_o refers to the MSE given by partially linear fitting.
[(e)] MSE_m refers to the MSE given by PLR-SCE.

.

Table 4.2: Simulation 4.1. Estimators of λ.

Sample Size	10×10	15×15	20×20
Torus$^{(a)}$			
mean of $\hat{\lambda}_o^{(b)}$	0.4735	0.4832	0.4897
mean of $\hat{\lambda}_m^{(c)}$	0.4548	0.4813	0.4919
var of $\hat{\lambda}_o$	0.0284	0.0120	0.0050
var of $\hat{\lambda}_m$	0.0024	0.0010	0.0005
Separable$^{(a)}$			
mean of $\hat{\lambda}_o$	0.5298	0.5309	0.4612
mean of $\hat{\lambda}_m$	0.4681	0.4833	0.4918
var of $\hat{\lambda}_o$	0.2721	0.4114	0.0486
var of $\hat{\lambda}_m$	0.0051	0.0030	0.0007
Unilateral$^{(a)}$			
mean of $\hat{\lambda}_o$	0.4635	0.4829	0.4907
mean of $\hat{\lambda}_m$	0.4637	0.4815	0.4905
var of $\hat{\lambda}_o$	0.0026	0.0008	0.0007
var of $\hat{\lambda}_m$	0.0024	0.0006	0.0005

$^{(a)}$ Torus, Separable and Unilateral are same defined as those in Table 4.1.
$^{(b)}$ $\hat{\lambda}_o$ refers to the estimators of λ obtained by the partially linear fitting.
$^{(c)}$ $\hat{\lambda}_m$ refers to the estimators of λ obtained by PLR-SCE.

.

Table 4.3: Simulation 4.1. Estimators of $\boldsymbol{\theta}=(\theta_1,\theta_2)$.

Sample Size	10×10	15×15	20×20
Torus$^{(*)}$			
mean of $\hat{\theta}_1$	0.3589	0.3857	0.3947
mean of $\hat{\theta}_2$	-0.1315	-0.1157	-0.1103
Separable$^{(*)}$			
mean of $\hat{\theta}_1$	0.2855	0.2891	0.2894
mean of $\hat{\theta}_2$	-0.2155	-0.2072	-0.2079
Unilateral$^{(*)}$			
mean of $\hat{\theta}_1$	0.3109	0.3665	0.3710
mean of $\hat{\theta}_2$	0.2309	0.2670	0.2893

$^{(*)}$ Torus, Separable and Unilateral are same defined as those in Table 4.1

Simulation 4.2 The testing model is

$$Y_{\mathbf{s}} = 0.5Z_{\mathbf{s}} + \sin(\pi X_{1,\mathbf{s}}) + \cos(\pi X_{2,\mathbf{s}}) + \epsilon_{\mathbf{s}}, \tag{4.19}$$

where $X_{i\mathbf{s}} \sim U(0,4)$, $i = 1, 2$, and other settings are the same as those in Simulation 4.1.

.

Table 4.4: Simulation 4.2. MSE for the estimators of $m(.)$.

Sample Size	10×10	15×15	20×20
Torus$^{(a)}$			
mean of $\text{MSE}_o^{(d)}$	1.6181	1.0743	0.7219
mean of $\text{MSE}_m^{(e)}$	0.6835	0.3048	0.1890
var of MSE_o	1.3553	0.1313	0.0231
var of MSE_m	0.1276	0.0111	0.0011
Separable$^{(b)}$			
mean of MSE_o	17.411	26.4761	2.8441
mean of MSE_m	0.9200	0.5619	0.2093
var of MSE_o	1.35×10^3	3.60×10^3	3.2634
var of MSE_m	0.4560	0.1048	0.0029
Unilateral$^{(c)}$			
mean of MSE_o	0.4977	0.2898	0.1813
mean of MSE_m	0.4473	0.2410	0.1567
var of MSE_o	0.0196	0.0067	0.0014
var of MSE_m	0.0132	0.0023	0.0009

$^{(a)}$ Torus: $\epsilon_{\mathbf{s}}$ generated by (2.2).
$^{(b)}$ Separable: $\epsilon_{\mathbf{s}}$ generated by (2.3).
$^{(c)}$ Unilateral: $\epsilon_{\mathbf{s}}$ generated by (2.4).
$^{(d)}$ MSE_o refers to the MSE given by partially linear fitting.
$^{(e)}$ MSE_m refers to the MSE given by PLR-SCE.

In Simulation 4.2, we consider a model with two explanatory variables in the nonparametric part. The results are shown in Table 4.4 to Table 4.6. Similar to the results of Simulation 4.1, we can find that PLR-SCE gains a better performance than the traditional estimation approach. By a cross comparison between the results presented in Table 4.1 and Table 4.4, we also see that the MSE of the estimators increase as the number of the nonparametric components increase. This is because of the curse of the

.

Table 4.5: Simulation 4.2. Estimators of λ.

Sample Size	10×10	15×15	20×20
Torus$^{(a)}$			
mean of $\hat{\lambda}_o^{(b)}$	0.4481	0.4622	0.4670
mean of $\hat{\lambda}_m^{(c)}$	0.3762	0.4151	0.4353
var of $\hat{\lambda}_o$	0.0276	0.0122	0.0053
var of $\hat{\lambda}_m$	0.0084	0.0016	0.0009
Separable$^{(a)}$			
mean of $\hat{\lambda}_o$	0.4314	0.4898	0.4756
mean of $\hat{\lambda}_m$	0.3924	0.4360	0.4380
var of $\hat{\lambda}_o$	0.4609	0.8071	0.0535
var of $\hat{\lambda}_m$	0.0330	0.0080	0.0011
Unilateral$^{(a)}$			
mean of $\hat{\lambda}_o$	0.3670	0.4181	0.4377
mean of $\hat{\lambda}_m$	0.3587	0.4113	0.4339
var of $\hat{\lambda}_o$	0.0062	0.0013	0.0007
var of $\hat{\lambda}_m$	0.0051	0.0011	0.0005

$^{(a)}$ Torus, Separable and Unilateral are same defined as those in Table 4.4.
$^{(b)}$ $\hat{\lambda}_o$ refers to the estimators of λ obtained by the partially linear fitting.
$^{(c)}$ $\hat{\lambda}_m$ refers to the estimators of λ obtained by PLR-SCE.

.

Table 4.6: Simulation 4.2. Estimators of $\boldsymbol{\theta}=(\theta_1,\theta_2)$.

Sample Size	10×10	15×15	20×20
Torus$^{(*)}$			
mean of $\hat{\theta}_1$	0.3036	0.3439	0.3559
mean of $\hat{\theta}_2$	-0.1346	-0.1362	-0.1300
Separable$^{(*)}$			
mean of $\hat{\theta}_1$	0.2856	0.2806	0.2808
mean of $\hat{\theta}_2$	-0.2187	-0.2123	-0.2139
Unilateral$^{(*)}$			
mean of $\hat{\theta}_1$	0.2368	0.3103	0.3093
mean of $\hat{\theta}_2$	0.1651	0.2305	0.2598

$^{(*)}$ Torus, Separable and Unilateral are same defined as those in Table 4.4.

dimensionality. Moreover, the estimation of the parametric coefficient λ is also less satisfied than that in Simulation 4.1. Hence, to increase the estimation performance in practice, we may remain only the primary factors in the nonparametric function, while considered the secondary factors as some parametric components.

4.5 Proofs

Before presenting the proof of Theorem 4.1, we need to establish two lemma in advance.

Lemma 4.1 *If assumptions* (C1), (C3)-(C5) *and* (C7) *are satisfied, then*

$$\hat{Q}_i(\mathbf{x}) - Q_i(\mathbf{x}) = O_p(h_L^2), \tag{4.20}$$

$i = 1, \ldots, b$.

Proof of Lemma 4.1: Lemma 4.1 can be seen as a special case of Theorem 2.1, with *i.i.d*. errors. Therefore, the proof of Lemma 4.1 can be given similarly as those of the later. □

For the sake of convenience, define $\mathbf{J_n}(\mathbf{x}) = \eta^T(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x}))^{-1}\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})$.

Lemma 4.2 *If assumptions* (C1), (C2), (C5) *and* (C7) *hold, then for any* $\mathbf{x}$*,*

$$m(\mathbf{x}) - \mathbf{J_n}(\mathbf{x})\mathbf{M} = O_p(h_L^2). \tag{4.21}$$

Proof of Lemma 4.2: Recall that $\mathbf{M} = \vec{\mathbf{X}}(\mathbf{x})\boldsymbol{\beta}(\mathbf{x}) + O_p(h_L^2)$, hence Lemma 4.2 can be established by simply showing $\mathbf{J_n}(\mathbf{x})\,\mathbf{1} = O_p(1)$. By Lemma 2.1 and the systematic

property of $K(.)$, we have

$$\eta^T(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x}))^{-1} = \eta^T(f^{-1}(\mathbf{x}) + o_p(1)).$$

Therefore,

$$\mathbf{J_n}(\mathbf{x})\,\mathbf{1} = (nh_\pi)^{-1}\sum_{\mathbf{s}\in\Omega}(f^{-1}(\mathbf{x}) + o_p(1))K_\pi(\zeta_1(\mathbf{X_s} - \mathbf{x})).$$

It is easy to show $(nh_\pi)^{-1}\sum_{\mathbf{s}\in\Omega} f^{-1}(\mathbf{x})K_\pi(\zeta_1(\mathbf{X_s} - \mathbf{x})) = O_p(1)$, then Lemma 4.2 follows. □

Proof of Theorem 4.1: Define $\vec{\mathbf{J}} = \left(\mathbf{J}_\mathbf{n}^T(\mathbf{X}_{\mathbf{s}_1}), \ldots, \mathbf{J}_\mathbf{n}^T(\mathbf{X}_{\mathbf{s}_n})\right)^T$, then we have

$$\mathbf{C} = (\mathbf{I} - \vec{\mathbf{J}})\mathbf{Y} = (\mathbf{I} - \vec{\mathbf{J}})(\vec{\mathbf{Z}}\boldsymbol{\lambda} + \mathbf{M} + \mathbf{E}) = \vec{\mathbf{D}}\boldsymbol{\lambda} + (\mathbf{I} - \vec{\mathbf{J}})(\mathbf{M} + \mathbf{E}). \tag{4.22}$$

Substitute it into (4.7), then

$$\hat{\boldsymbol{\lambda}} = (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}(\vec{\mathbf{D}}^T\mathbf{C}) = \lambda + (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}\left[\vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})(\mathbf{M} + \mathbf{E})\right]. \tag{4.23}$$

Hence, to show Theorem 4.1, we will prove

$$\text{(c1)} \qquad n^{-1}\vec{\mathbf{D}}^T\vec{\mathbf{D}} \xrightarrow{P} \sigma_\mathbf{Z}\sigma_\mathbf{Z}^T,$$

$$\text{(c2)} \qquad n^{-1/2}\,\vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})(\mathbf{M} + \mathbf{E}) \xrightarrow{D} N(0, \sigma_\epsilon^2\sigma_\mathbf{Z}\sigma_\mathbf{Z}^T).$$

Proof of (c1):

$$\begin{aligned}
n^{-1}\vec{\mathbf{D}}^T\vec{\mathbf{D}} &= n^{-1}\sum_{\mathbf{s}\in\Omega}\left(\mathbf{Z_s} - \hat{Q}_\mathbf{Z}(\mathbf{X_s})\right)\left(\mathbf{Z_s} - \hat{Q}_\mathbf{Z}(\mathbf{X_s})\right)^T \\
&= n^{-1}\sum_{\mathbf{s}\in\Omega}\left(\mathbf{Z_s} - Q_\mathbf{Z}(\mathbf{X_s})\right)\left(\mathbf{Z_s} - Q_\mathbf{Z}(\mathbf{X_s})\right)^T + n^{-1}\sum_{\mathbf{s}\in\Omega}\left(\mathbf{Z_s} - Q_\mathbf{Z}(\mathbf{X_s})\right)\left(Q_\mathbf{Z}(\mathbf{X_s}) - \hat{Q}_\mathbf{Z}(\mathbf{X_s})\right)^T \\
&\quad + n^{-1}\sum_{\mathbf{s}\in\Omega}\left(Q_\mathbf{Z}(\mathbf{X_s}) - \hat{Q}_\mathbf{Z}(\mathbf{X_s})\right)\left(\mathbf{Z_s} - Q_\mathbf{Z}(\mathbf{X_s})\right)^T \\
&\quad + n^{-1}\sum_{\mathbf{s}\in\Omega}\left(Q_\mathbf{Z}(\mathbf{X_s}) - \hat{Q}_\mathbf{Z}(\mathbf{X_s})\right)\left(Q_\mathbf{Z}(\mathbf{X_s}) - \hat{Q}_\mathbf{Z}(\mathbf{X_s})\right)^T.
\end{aligned}$$

By Lemma 4.1, we have $Q_{\mathbf{Z}}(\mathbf{X_s}) - \hat{Q}_{\mathbf{Z}}(\mathbf{X_s}) = O_p(h_L^2)$. Together with assumptions (C3) and (C7), we can see that the last 3 summands on the right hand side of the last equality are all equal to $o_p(1)$. (c1) then follows straightforwardly.

To show (c2), it suffices to show

$$\begin{aligned} q_1 &= n^{-1/2}\vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})\mathbf{M} = o_p(1), \\ q_2 &= n^{-1/2}\vec{\mathbf{D}}^T\vec{\mathbf{J}}\,\mathbf{E} = o_p(1), \\ q_3 &= n^{-1/2}\vec{\mathbf{D}}^T\mathbf{E} \xrightarrow{D} N\Big(0, \sigma_\epsilon^2\sigma_{\mathbf{Z}}\sigma_{\mathbf{Z}}^T\Big). \end{aligned}$$

First, we have

$$q_1 = n^{-1/2}\sum_{\mathbf{s}\in\Omega}\mathbf{D_s}\Big(m(\mathbf{X_s}) - \mathbf{J_n}(\mathbf{X_s})\mathbf{M}\Big) \le n^{-1/2}\gamma_1\sum_{\mathbf{s}\in\Omega}\mathbf{D_s},$$

where $\gamma_1 = \sup_{\mathbf{s}\in\Omega}\{|\, m(\mathbf{X_s}) - \mathbf{J_n}(\mathbf{X_s})\mathbf{M}\,|\}$. By Lemma 4.2, we have $\gamma_1 = O_p(h_L^2)$. Moreover,

$$\begin{aligned} \sum_{\mathbf{s}\in\Omega}\mathbf{D_s} &= \sum_{\mathbf{s}\in\Omega}\Big(\mathbf{Z_s} - Q_{\mathbf{Z}}(\mathbf{X_s})\Big) + \sum_{\mathbf{s}\in\Omega}\Big(Q_{\mathbf{Z}}(\mathbf{X_s}) - \hat{Q}_{\mathbf{Z}}(\mathbf{X_s})\Big) \\ &\le \sum_{\mathbf{s}\in\Omega}\sigma_{\mathbf{Z}}e_{\mathbf{s}} + \Big(\sum_{\mathbf{s}\in\Omega}\Big(Q_{\mathbf{Z}}(\mathbf{X_s}) - \hat{Q}_{\mathbf{Z}}(\mathbf{X_s})\Big)\odot\Big(Q_{\mathbf{Z}}(\mathbf{X_s}) - \hat{Q}_{\mathbf{Z}}(\mathbf{X_s})\Big)\Big)^{1/2}. \end{aligned} \quad (4.24)$$

The first term on the right hand side of the last inequality is obviously equal to $O_p(n^{1/2})$, while the second term is equal to $O_p(n^{1/2}h_L^2)$ by Lemma 4.1. Hence, with assumption (C7), we can obtain that $q_1 = o_p(1)$.

Consider q_2, we have

$$q_2 = n^{-1/2}\sum_{\mathbf{s}\in\Omega}\mathbf{D_s}\,\mathbf{J_n}(\mathbf{X_s})\,\mathbf{E} \le n^{-1/2}\gamma_2\sum_{\mathbf{s}\in\Omega}\mathbf{D_s},$$

where $\gamma_2 = \sup_{\mathbf{s}\in\Omega}\{|\,\mathbf{J_n}(\mathbf{x})\,\mathbf{E}\,|\}$. By Lemma 2.1 and Lemma 2.2, we can show that $\gamma_2 = O_p((nh_\pi)^{-1/2})$. Therefore, together with (4.24) and assumption (C7), we have $q_2 = o_p(1)$.

At last, we focus on q_3.

$$q_3 = n^{-1/2}\sum_{\mathbf{s}\in\Omega}\Big(\mathbf{Z}_\mathbf{s} - Q_\mathbf{Z}(\mathbf{X}_\mathbf{s})\Big)\epsilon_\mathbf{s} + n^{-1/2}\sum_{\mathbf{s}\in\Omega}\Big(Q_\mathbf{Z}(\mathbf{X}_\mathbf{s}) - \hat{Q}_\mathbf{Z}(\mathbf{X}_\mathbf{s})\Big)\epsilon_\mathbf{s} \triangleq q_{31} + q_{32}.$$

Let $\gamma_3 = \sup_{\mathbf{s}\in\Omega}\{|Q_\mathbf{Z}(\mathbf{X}_\mathbf{s}) - \hat{Q}_\mathbf{Z}(\mathbf{X}_\mathbf{s})|\}$, then $q_{32} \le n^{-1/2}\gamma_3 \sum_{\mathbf{s}\in\Omega}\epsilon_\mathbf{s}$. By Lemma 4.1, we have $\gamma_3 = O_p(h_L^2)$. So if we can prove $\sum_{\mathbf{s}\in\Omega}\epsilon_{\mathbf{s}_i} = O_p(n^{1/2})$, we then have $q_{32} = o_p(1)$. For this purpose, we only need to consider the variance of $\sum_{\mathbf{s}\in\Omega}\epsilon_{\mathbf{s}_i}$. Recall that $R(\mathbf{k}) = \mathrm{Cov}(\epsilon_\mathbf{s}, \epsilon_{\mathbf{s}+\mathbf{k}})$, $\mathbf{k} = (k_1, k_2)$, then

$$\mathrm{Var}\Big(\sum_{\mathbf{s}\in\Omega}\epsilon_\mathbf{s}\Big) = \sum_{|k_1|<n_1}\sum_{|k_2|<n_2}(n_1 - |k_1|)(n_2 - |k_2|)\Big|R(k_1,k_2)\Big| \le Cn\sum_{k_1=0}^{n_1}\sum_{k_2=0}^{n_2}\Big|R(k_1,k_2)\Big|.$$

With Lemma 2.3 or Lemma 2.5, we obtain $\mathrm{Var}\Big(\sum_{i=1}^{n}\epsilon_{\mathbf{s}_i}\Big) = O(n)$, hence $\sum_{\mathbf{s}\in\Omega}\epsilon_\mathbf{s} = O_p(n^{1/2})$. Now it remains for us to show $q_{31} = n^{-1/2}\sum_{\mathbf{s}\in\Omega}\sigma_\mathbf{Z} e_\mathbf{s}\epsilon_\mathbf{s}$ is asymptotically normal. This is naturally satisfied, since both $\{\mathbf{e}_\mathbf{s}\}$ and $\{\epsilon_\mathbf{s}\}$ are *i.i.d.* and mutually independent. At last, it is easy to show $\mathrm{Var}(q_{31}) \to \sigma_\epsilon^2\sigma_\mathbf{Z}\sigma_\mathbf{Z}^T$. That conclude the proof of Theorem 4.1. □

Proof of Theorem 4.2: Consider

$$\begin{aligned}\hat{m}(\mathbf{x}) - m(\mathbf{x}) &= \Big(\hat{Q}_\mathbf{Y}(\mathbf{x}) - \hat{Q}_\mathbf{Z}^T(\mathbf{x})\boldsymbol{\lambda} - m(\mathbf{x})\Big) + (\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})\hat{Q}_\mathbf{Z}(\mathbf{x}) \\ &= \Big(\mathbf{J}_\mathbf{n}(\mathbf{x})(\mathbf{M} + \mathbf{E}) - m(\mathbf{x})\Big) + (\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})\hat{Q}_\mathbf{Z}(\mathbf{x}). \qquad (4.25)\end{aligned}$$

By Theorem 4.1, we have $\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}} = O_p(n^{-1/2})$. By Lemma 4.1, we have $\hat{Q}_\mathbf{Z}(\mathbf{x}) = O_p(1)$. Hence, together with (4.25) and assumption (C7), Theorem 4.2 can be established by showing that

$$\mathbf{J}_\mathbf{n}(\mathbf{x})(\mathbf{M} + \mathbf{E}) - m(\mathbf{x}) - \mathrm{Bias}(\mathbf{x}, \mathbf{h}) = \eta^T\mathbf{U}_\mathbf{n}^{-1}(\mathbf{x})\mathbf{V}_\mathbf{n}^*(\mathbf{x}) + o_p(h_L^2),$$

which is proved in Theorem 2.1. The proof of Theorem 4.2 is completed. □

Proof of Theorem 4.3: From (4.25) and assumption (C7), we can see that the asymptotic normality of $\hat{m}(\mathbf{x})$ can be established by showing

$$(nh_\pi)^{1/2}\Big(\mathbf{J_n}(\mathbf{x})(\mathbf{M}+\mathbf{E}) - m(\mathbf{x}) - \text{Bias}(\mathbf{x},\mathbf{h})\Big) \xrightarrow{D} N(0, \Sigma_\epsilon(\mathbf{x})). \tag{4.26}$$

Such result is satisfied by Theorem 2.2. □

Proof of Theorem 4.4: Recall that $\hat{\mathbf{P}}$ can be expressed as $\hat{\mathbf{P}} = \vec{\mathbf{Z}}\hat{\boldsymbol{\lambda}} + \hat{\mathbf{M}} + (\mathbf{I} - \mathbf{B}(\hat{\boldsymbol{\theta}}))\hat{\mathbf{E}}$. Similar to $\hat{\mathbf{P}}$, we defined a n-dimensional vector $\check{\mathbf{P}} = (\check{P}_{\mathbf{s}_1}, \cdots, \check{P}_{\mathbf{s}_n})^T$ as $\check{\mathbf{P}} = \vec{\mathbf{Z}}\hat{\boldsymbol{\lambda}} + \hat{\mathbf{M}} + (\mathbf{I} - \mathbf{B}(\boldsymbol{\theta}))\hat{\mathbf{E}}$. Moreover, denote $\check{Q}_{\mathbf{P}}(\mathbf{x}) = \eta^T\Big(\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\vec{\mathbf{X}}(\mathbf{x})\Big)^{-1}\vec{\mathbf{X}}^T(\mathbf{x})\mathbf{W}(\mathbf{x})\check{\mathbf{P}}$, and $\check{L}_{\mathbf{s}} = \check{\mathbf{P}}_{\mathbf{s}} - \check{Q}_{\mathbf{P}}(\mathbf{X}_{\mathbf{s}})$ for $\mathbf{s} \in \Omega$. Then, like (4.11), we let

$$\check{\boldsymbol{\lambda}} = (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}\vec{\mathbf{D}}^T\check{\mathbf{L}}, \tag{4.27}$$

where $\check{\mathbf{L}} = (\check{L}_{\mathbf{s}_1}, \ldots, \check{L}_{\mathbf{s}_n})^T$. As we have $\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = (\tilde{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}}) + (\check{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$, Theorem 4.4 can be established by proving

$$\text{(c3)} \quad \tilde{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}} = o_p(n^{-1/2}),$$

$$\text{(c4)} \quad n^{1/2}(\check{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{D} N\Big(0,\ \sigma_\tau^2(\sigma_{\mathbf{Z}}\sigma_{\mathbf{Z}}^T)^{-1}\Big).$$

Proof of (c3):

$$\tilde{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}} = (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}\vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})(\hat{\mathbf{P}} - \check{\mathbf{P}}) = (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}\vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})(\mathbf{B}(\boldsymbol{\theta}) - \mathbf{B}(\hat{\boldsymbol{\theta}}))\hat{\mathbf{E}}.$$

From (c1), we have $\vec{\mathbf{D}}^T\vec{\mathbf{D}} = O_p(n)$. So to prove (c3), it suffices to show

$$I_1 = \vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})(\mathbf{B}(\boldsymbol{\theta}) - \mathbf{B}(\hat{\boldsymbol{\theta}}))\hat{E} = o_p(n^{1/2}).$$

Define $\mathbf{G} = (G_{\mathbf{s}_1}, \ldots, G_{\mathbf{s}_n})^T$ with the i-th element being $G_{\mathbf{s}_i} = \sum_{\mathbf{t}\in N(\mathbf{s}_i)} \epsilon_{\mathbf{t}}$, $N(\mathbf{s})$ is the neighborhood set of $\epsilon_{\mathbf{s}}$. By Theorem 4.1 and Theorem 4.2, we have $\hat{\mathbf{E}} = \mathbf{E} + O_p(h_L^2)$. Note that there are finite nonzero elements in $\mathbf{B}(\boldsymbol{\theta})$, therefore

$$\begin{aligned} I_1 &= \vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})(\mathbf{G} + O_p(h_L^2)) \cdot o_p(1) \\ &= \left(\vec{\mathbf{D}}^T\mathbf{G} + \vec{\mathbf{D}}^T\vec{\mathbf{J}}\mathbf{G}\right) \cdot o_p(1) + \vec{\mathbf{D}}^T(\mathbf{I} - \vec{\mathbf{J}})\mathbf{1} \cdot o_p(h_L^2) \\ &= (I_{11} + I_{12}) \cdot o_p(1) + I_{13} \cdot o_p(h_L^2). \end{aligned} \tag{4.28}$$

For I_{11}, we have

$$\begin{aligned} I_{11} &= \sum_{\mathbf{s}\in\Omega} \mathbf{D}_{\mathbf{s}} G_{\mathbf{s}} = \sum_{\mathbf{s}\in\Omega} (\mathbf{Z}_{\mathbf{s}} - \hat{Q}_{\mathbf{Z}}(\mathbf{X}_{\mathbf{s}}))G_{\mathbf{s}} \\ &= \sum_{\mathbf{s}\in\Omega} (\mathbf{Z}_{\mathbf{s}} - Q_{\mathbf{Z}}(\mathbf{X}_{\mathbf{s}}))G_{\mathbf{s}} + \sum_{\mathbf{s}\in\Omega} (Q_{\mathbf{Z}}(\mathbf{X}_{\mathbf{s}}) - \hat{Q}_{\mathbf{Z}}(\mathbf{X}_{\mathbf{s}}))G_{\mathbf{s}} \\ &\hat{=} I_{111} + I_{112}. \end{aligned} \tag{4.29}$$

Due to the fact that $\{\mathbf{e}_{\mathbf{s}}\}$ is independent of $\{G_{\mathbf{s}}\}$ and $\mathbf{e}$ is *i.i.d.*, we can show that $\text{var}(I_{111}) = O(n)$, and hence $I_{111} = O_p(n^{1/2})$. By Lemma 4.1, we have $Q_{\mathbf{Z}}(\mathbf{X}_{\mathbf{s}}) - \hat{Q}_{\mathbf{Z}}(\mathbf{X}_{\mathbf{s}}) = O_p(g_L^2)$, and by the same procedure of (2.57)-(2.60), we can show $\text{var}\,(\sum_{\mathbf{s}\in\Omega} \mathbf{G}_{\mathbf{s}}) = O(n)$. These will leads to the result that $I_{112} = o_p(n^{1/2})$. Therefore, we have $I_{11} = O_p(n^{1/2})$.

For I_{12}, it can be presented that

$$I_{12} = \sum_{\mathbf{s}\in\Omega} \mathbf{D}_{\mathbf{s}}\mathbf{J}_{\mathbf{n}}(\mathbf{X}_{\mathbf{s}})\mathbf{G} \leq \gamma_4 \sum_{\mathbf{s}\in\Omega} \mathbf{D}_{\mathbf{s}},$$

where $\gamma_4 = \sup_{\mathbf{s}\in\Omega}\{\,|\mathbf{J_n}(\mathbf{X_s})\mathbf{G}|\,\}$. By the same procedure of (2.57)-(2.60) again, we can show that $\gamma_4 = O_p((ng_\pi)^{-1/2})$. Moreover, by (4.24) and assumption (C8), we have $\sum_{\mathbf{s}\in\Omega}\mathbf{D_s} = O_p(n^{1/2}) + O_p(g_\pi^{-1/2})$. Hence, we can conclude that $I_{12} = O_p(g_\pi^{-1/2})$.

Similar to the proof of I_{11} and I_{12}, we can show $I_{13} = O_p(n^{1/2})$. So by (4.28), we can finally get $I_1 = o_p(n^{1/2})$, which conclude the proof of (c3).

Proof of (c4): By Theorem 4.1 and Theorem 4.2, we have

$$\begin{aligned}\check{\mathbf{L}} &= (\mathbf{I}-\vec{\mathbf{J}})\check{\mathbf{P}} = (\mathbf{I}-\vec{\mathbf{J}})\left[\vec{\mathbf{Z}}\hat{\boldsymbol{\lambda}} + \hat{\mathbf{M}} + (\mathbf{I}-\mathbf{B}(\boldsymbol{\theta}))\hat{\mathbf{E}}\right]\\ &= (\mathbf{I}-\vec{\mathbf{J}})\Big(\vec{\mathbf{Z}}\boldsymbol{\lambda} + \mathbf{M} + \Upsilon + O_p(h_L^2)\Big)\\ &= \vec{\mathbf{D}}\boldsymbol{\lambda} + (\mathbf{I}-\vec{\mathbf{J}})(\mathbf{M}+\Upsilon) + (\mathbf{I}-\vec{\mathbf{J}})\mathbf{1}\cdot O_p(h_L^2). \qquad (4.30)\end{aligned}$$

Substitute this into (4.27), then

$$\check{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}\vec{\mathbf{D}}(\mathbf{I}-\vec{\mathbf{J}})(\mathbf{M}+\Upsilon) + (\vec{\mathbf{D}}^T\vec{\mathbf{D}})^{-1}\vec{\mathbf{D}}(\mathbf{I}-\vec{\mathbf{J}})\mathbf{1}\cdot O_p(h_L^2). \qquad (4.31)$$

As we have shown $\vec{\mathbf{D}}(\mathbf{I}-\vec{\mathbf{J}})\mathbf{1} = O_p(n^{-1/2})$ and (c1), hence the second term on the right hand side of the equation above is equal to $o_p(n^{-1/2})$. Then with (c1) again, it remains for us to show

$$\text{(c5)} \qquad n^{1/2}\vec{\mathbf{D}}(\mathbf{I}-\vec{\mathbf{J}})(\mathbf{M}+\Upsilon) \xrightarrow{D} N\Big(0,\ \sigma_\tau^2\sigma_{\mathbf{Z}}\sigma_{\mathbf{Z}}^T\Big),$$

which can be proved similarly as (c2). Theorem 4.4 follows. □

Proof of Theorem 4.5: Recall that $\hat{\mathbf{P}} = \vec{\mathbf{Z}}\hat{\boldsymbol{\lambda}} + \hat{\mathbf{M}} + (\mathbf{I}-\mathbf{B}(\hat{\boldsymbol{\theta}}))\hat{\mathbf{E}}$, hence

$$\begin{aligned}(ng_\pi)^{1/2}\Big(\tilde{m}(\mathbf{x}) - m(\mathbf{x})\Big) &= (ng_\pi)^{1/2}\Big(\,\hat{Q}_{\mathbf{P}}(\mathbf{x}) - \hat{Q}_{\mathbf{Z}}(\mathbf{x})\hat{\boldsymbol{\lambda}} - m(\mathbf{x}) + (\hat{\boldsymbol{\lambda}}-\tilde{\boldsymbol{\lambda}})\hat{Q}_{\mathbf{Z}}(\mathbf{x})\,\Big)\\ &= (ng_\pi)^{1/2}\Big[\,\mathbf{J_n}(\mathbf{x})\Big(\hat{\mathbf{M}} + (\mathbf{I}-\mathbf{B}(\hat{\boldsymbol{\theta}}))\hat{\mathbf{E}}\Big) - m(\mathbf{x})\Big] + (ng_\pi)^{1/2}(\hat{\boldsymbol{\lambda}}-\tilde{\boldsymbol{\lambda}})\hat{Q}_{\mathbf{Z}}(\mathbf{x}).\end{aligned}$$

By Theorem 4.1 and Theorem 4.4, we have $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = O_p(n^{-1/2})$ and $\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = O_p(n^{-1/2})$. Therefore, $\hat{\boldsymbol{\lambda}} - \tilde{\boldsymbol{\lambda}} = (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) - (\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) = O_p(n^{-1/2})$. Moreover, by Lemma 4.1 we have $\hat{Q}_{\mathbf{Z}}(\mathbf{x}) = O_p(1)$. Hence, the second term on the right hand side of the last equality is equal to $o_p(1)$. Now, it remains to show the first term is asymptotically normal, which has been proved in Theorem 2.3. The proof of Theorem 4.5 is completed. □

4.6 Summary and Remark

This chapter is another extended research of Chapter 2. The partially linear model is considered, where the errors are assumed to satisfy the torus, separable or unilateral SGAR model. The two-step estimation approach is proposed again. We establish the asymptotic normality of the first- and second-step estimators of the parametric coefficients. For the first-step estimator of the $m(.)$ function, we present the convergence rate with all three types of error structures and the asymptotic normality with separable and unilateral SGAR-type errors. For the second-step estimator of the $m(.)$ function, we establish the asymptotic normality when errors satisfy the three types of structures. Simulations are conducted, showing that our method obtains great improvement in estimation when spatially dependent errors exist. Comparing with the approach in Chapter 2, the method considered here partly avoid the curse of dimensionality. Hence, more regressors can be introduced into the nonparametric component, leading the estimation more feasible and flexible in practice.

Chapter 5

Case Study with Hong Kong Real Estate Market Data

The assessment of prices of house is an important problem in the real estate market. There are two groups of factors that largely determine square-unit-price of a house. The first one contains the basic properties of the house, such as the area of the house, the age of the house, etc. The other one reflect the influences from the nearby blocks. It is natural that a luxurious house in the suburbs may not be quite expensive, while an old house in the business district or at the center of a city can be of astonishingly high price. Therefore, with the consideration of the effects of both the exogenous variables and spatial dependency, our models considered in Chapter 2 to Chapter 4 may be appropriate for the house pricing analysis. In this chapter, we will examine the Hong Kong real estate market, and the organization is as follow. In the first section, we give a brief introduction to the target problem as well as the data. In the three sections following, we apply the two-step fitting considered in Chapter 2 to Chapter 4 to the case study

respectively. In the last section, a short summary is given.

5.1 The Hong Kong Real Estate Market Data

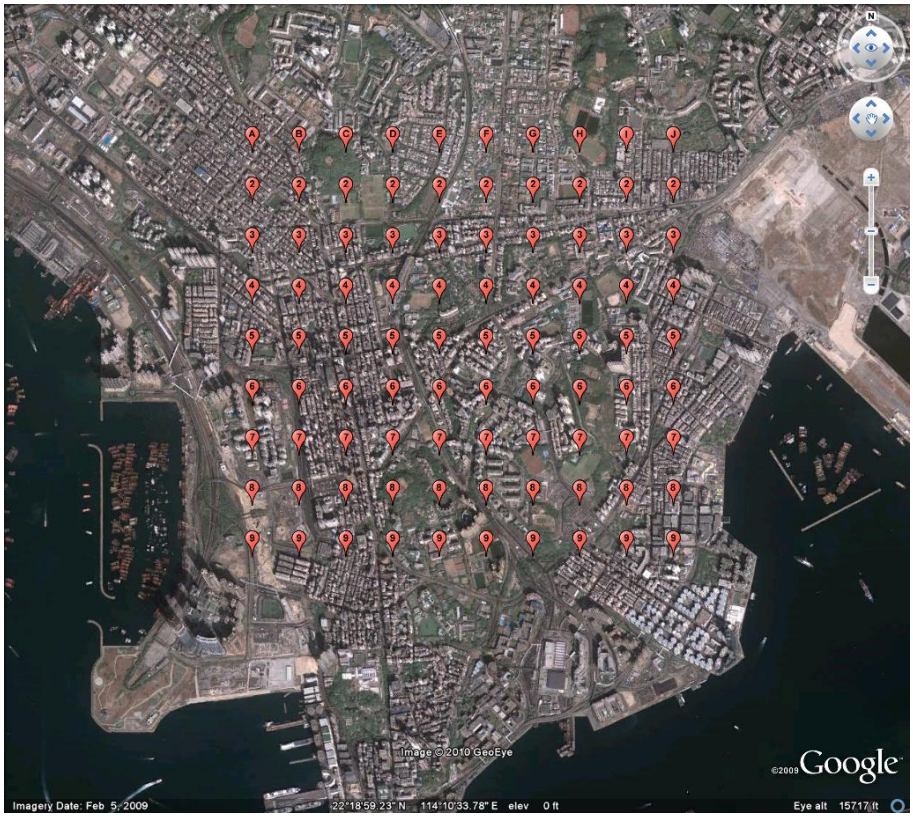


Figure 5.1: The map for the Hong Kong real estate market data.

Kowloon is a main district in Hong Kong. As a large and regular residential area can be found in this district, we choose it for our study. We consider observations from a 9×10 regular grid, see Figure 5.1. All the positions are equally spaced in area. However, only data at 86 positions are available at last. Each observation contains 2 pieces of information. The first part is the information of the target house, including the average price (AP) per feet^2 of the house, the total area (Area) of the house and the age of the building (AoB). The second part contains the information of the community where the house locates, including the average household size (AHS), the

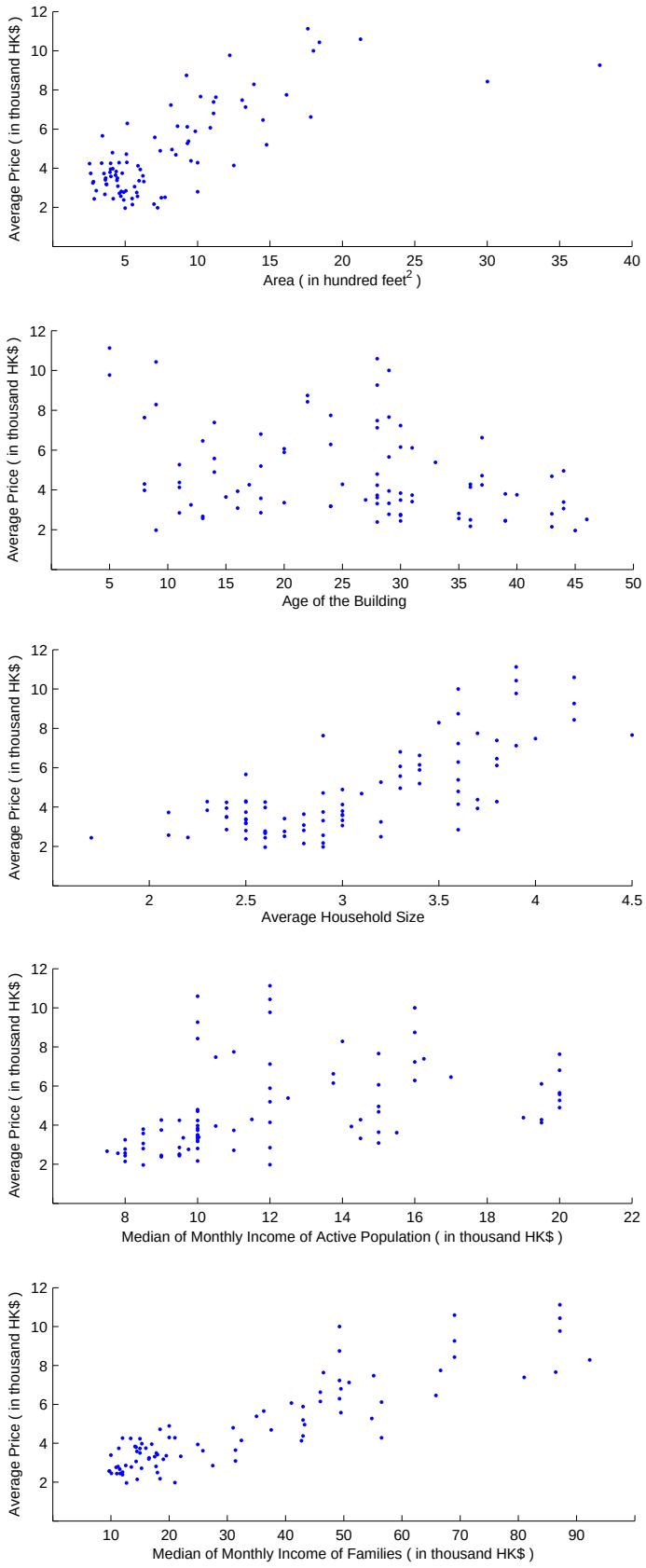


Figure 5.2: Scattergrams of the Hong Kong real estate data.

median of monthly income of active population (MMIAP) and the median of monthly income of families (MMIF). The data of the first part comes from the trading record during April 2009 to September 2009, while the data of the second part comes from the Hong Kong census in 2006. All data is available on the website of Centaline Property (http://hk.centanet.com/icms/template.aspx?series=1).

In Figure 5.2, we plot the scattergrams for the AP variable and each of the other five variables separatively. From the second and fourth graphs of Figure 5.2, it is hard for us to find any trend by eye. However, from the other three graphs, we can see that AP value increases generally as the values of Area, AHS and MMIF increase, and such increment may be linear. Note that some linear relationship may also exist between Area, AHS and MMIF mutually, as it is quite natural that a family with more income will buy a bigger house, which can accommodate more people. We may also note, in the first graph of Figure 5.2, that there are two points isolated from the others. That may lead to a bad result in kernel estimation.

5.2 Fitting with LLR-SCE

In this section, we will apply the estimation procedure LLR-SCE introduced in Chapter 2 to this case study. As the amount of observation is not quite adequate, we only consider models with no more than two explanatory variables. Before the fitting, we should conduct some preparation. Since there are 4 missing data in the study, we will give some initial estimators to them as follow. We first obtain some estimates of the $m(.)$ function by the local linear fitting, and subsequently the estimates of the errors can be given. Then the errors for the missing data can be estimated as the one-step predictor

Table 5.1: Case Study. MSE given by LLR-SCE$^{(a)}$.

d = 1	**LR-SCE**$^{(a)}$	**LLR**$^{(a)}$	**Torus**$^{(b)}$	**Separable**$^{(b)}$	**Unilateral**$^{(b)}$
Area	2.0979×10^6	1.6982×10^6	1.5298×10^6	1.6921×10^6	1.6928×10^6
AoB	3.3609×10^6	3.7014×10^6	2.1747×10^6	2.0037×10^6	2.1729×10^6
AHS	2.3258×10^6	1.7958×10^6	1.7646×10^6	1.7734×10^6	1.7742×10^6
MMIAP	3.6579×10^6	3.3427×10^6	2.4637×10^6	2.4085×10^6	2.4782×10^6
MMIF	1.0277×10^6	1.0595×10^6	1.0033×10^6	9.9356×10^5	9.9731×10^5
d = 2	**LR-SCE**$^{(a)}$	**LLR**	**Torus**	**Separable**	**Unilateral**
Area & AoB	1.8862×10^6	7.7731×10^5	7.3859×10^5	7.4528×10^5	7.5133×10^5
Area & AHS	1.7558×10^6	1.0411×10^6	8.0838×10^5	8.7777×10^5	8.7068×10^5
Area & MMIAP	1.7887×10^6	1.1531×10^6	1.1052×10^6	1.1476×10^6	1.1496×10^6
Area & MMIF	9.5500×10^5	7.9500×10^5	5.9471×10^5	6.0350×10^5	6.5401×10^5
AoB & AHS	2.0077×10^6	1.4371×10^6	1.2468×10^6	1.2690×10^6	1.3024×10^6
AoB & MMIAP	3.4258×10^6	2.7509×10^6	1.8381×10^6	1.7897×10^6	1.9068×10^6
AoB & MMIF	1.0598×10^6	6.2371×10^5	5.6655×10^5	5.5781×10^5	5.5328×10^5
AHS & MMIAP	2.2917×10^6	1.1441×10^6	1.0793×10^6	1.0715×10^6	1.0940×10^6
AHS & MMIF	1.0433×10^6	9.9428×10^5	9.5270×10^5	9.4140×10^5	9.4406×10^5
MMIAP & MMIF	1.0444×10^6	4.8321×10^5	4.8336×10^5	4.6791×10^5	4.7162×10^5
All variables	9.8050×10^5				

$^{(a)}$ LLR-SCE refers to the two-step estimation considered in Chapter 2. LR-SCE refers to the estimation of linear regression model with unilateral SGAR-type errors. LLR refers to the local linear fitting.
$^{(b)}$ Torus: LLR-SCE with ϵ_s satisfying (2.2). Separable: LLR-SCE with ϵ_s satisfying (2.3). Unilateral: LLR-SCE with ϵ_s satisfying (2.4).

Table 5.2: Case Study. Estimators of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ given by LLR-SCE$^{(*)}$.

d = 1	**Torus**$^{(*)}$		**Separable**$^{(*)}$		**Unilateral**$^{(*)}$	
Area	0.0484	-0.0157	0.0181	-0.0159	0.0404	-0.0303
AoB	0.2059	0.1907	0.1885	0.1556	0.3999	0.3203
AHS	0.0493	-0.0050	0.0431	0.0085	0.0837	0.0103
MMIAP	0.0373	0.2088	0.0572	0.2094	0.1177	0.4171
MMIF	0.0074	0.0812	0.0048	0.0928	0.0329	0.1792
d = 2	**Torus**		**Separable**		**Unilateral**	
Area&AoB	0.0858	0.0659	0.0836	0.0542	0.1547	0.0888
Area&AHS	0.1127	-0.1037	0.0863	-0.0774	0.1825	-0.1816
Area&MMIAP	0.0588	-0.0061	0.0264	0.0083	0.0463	0.0085
Area&MMIF	0.1509	-0.0274	0.1410	-0.0061	0.2591	-0.0327
AoB&AHS	0.1062	0.0909	0.0898	0.0798	0.1740	0.1506
AoB&MMIAP	0.1406	0.2136	0.1161	0.2031	0.2326	0.4133
AoB&MMIF	0.0462	0.0327	0.0472	0.0591	0.1158	0.1256
AHS&MMIAP	0.0095	0.0668	0.0185	0.0793	0.0475	0.1525
AHS&MMIF	0.0184	0.0699	0.0162	0.0836	0.0528	0.1648
MMIAP&MMIF	-0.0116	0.0215	-0.0143	0.0504	0.0090	0.0872

$^{(*)}$ LLR-SCE, Torus, Separable and Unilateral are same defined as the ones in Table 5.1.

with the corresponding SGAR model, and the AP values for the missing data can be roughly estimated as the estimators of $m(\mathbf{x})$ plus the estimators of the errors.

Now we can begin the fitting. For comparison, two traditional models/methods are also adopted in this case. The first one is the linear regression model with spatially correlated errors. Such model is considered in many literature, for example, Basu & Reinsel (1994). The second one is the non-parametric regression model where local linear fitting is used. The estimation results are shown in Table 5.1 and Table 5.2. It is readily to find that our fitting overmatch the one with linear regression model. Therefore, we mainly compare the fitting results given by local linear fitting and our method in the following.

We first focus on the results with single explanatory variable. From the upper section of Table 5.1, we can see that the mean squared errors given by our proposed method are smaller than those given by the local linear fitting. The differences are especially large for the cases with AoB and MMIAP. For example, with single explanatory variable AoB, the mean squared error given by local linear fitting is 3.7014×10^6, while those given by our method are less that 2.2×10^6. In comparison, in the cases with Area, AHS and MMIF, the differences are smaller. We may explain such phenomenon with Figure 5.2. As we have mentioned above, linear trends can be seen in the graphs with Area, AHS and MMIF in Figure 5.2. In such situation, the local linear fitting may perform well enough, so that our modified fitting makes little improvement. In contrast, when no obvious trend exists, our method performs much better than the local linear fitting. Another explanation might be that AoB and MMIAP are not the key explanatory variables, or at least they are not sufficient to determine the average price of houses.

Hence, the spatial autoregression complements to improve the model. In Table 5.2, the estimators of $\boldsymbol{\theta}$ are shown. It can be seen that the estimated values, correlated with AoB and MMIAP, are larger than the others. That means the errors in these two cases are more likely to be spatially correlated, and therefore our method plays its role. This is coincident with the results in Table 5.1.

We continue to examine the results with two explanatory variables. From the lower section of Table 5.1, we have several findings. First, the greatest improvement of the results made by our method, comparing with the local linear fitting, takes place in the case with AoB & MMIAP. This is not surprising, as we have mentioned above that our method also performs much better than local linear fitting with single AoB or MMIAP. Second, our method also decreases the MSE greatly in the cases with Area & AHS and Area & MMIF. Remember that our method does not make large improvement in the cases with Area, AHS or MMIF alone. This shows that that Area, AHS and MMIF is not able to determine the average price of houses alone, however, a combination of them may do. Next, comparing the results with MMIF and AHS & MMIF, we can see that the additional explanatory variable AHS does not lead to a significant decreasement of the MSE. However, together with other explanatory variables, AHS does contribute to an obvious decrease of the MSE. For instance, see the results with AoB and AoB & AHS. Such phenomenon indicates that AHS and MMIF may be highly correlated. This verifies our suspicion before, and AHS may be deleted from the model fitting. Furthermore, we find that the mean squared error given by local linear fitting with AoB & MMIAP is 2.7509×10^6, larger than the ones given by our method with single AoB or MMIAP, which are around 2.1×10^6 and 2.4×10^6 respectively. Hence, our method is

especially meaningful here. At last, we note that the MSE with MMIAP & MMIF is the smallest one, and it is the best result so far.

5.3 Fitting with ADD-SCE

In the last section, we apply LLR-SCE considered in Chapter 2 to our empirical research. Multiple regressors have been introduced into modeling. We may be curious to know whether some additive structure exists for the explanatory variables. If the answer is positive, then the additive model should be more appropriate to this case study than the nonparametric model considered in the last section. This is due to the faster convergent rate of the estimators in additive model, see Chapter 3. Therefore, in the following, we conduct the fitting with every pair of explanatory variables using the method considered in Chapter 3.

In Table 5.3, the first column shows the MSE of the fitting with the method of marginal integration technique. For comparison, the rest columns show the MSE obtained by ADD-SCE. It can be seen that our method gets better results in all setup. On the other hand, we may also compare the results in Table 5.3 with the ones in Table 5.1. By matching line by line, we find that the MSE values given in Table 5.1 are smaller than those given in Table 5.3. This may indicate that the assumption of additive structure is not quite suitable in this empirical research. And therefore, the nonparametric model fits the data better.

At last, we show part of the fitting results graphically in Figure 5.3. It can be found clearly that the graphs of MMIF fit the observations very well. This indicates that MMIF is an important regressor in our analysis. By contrast, we can see that the contribution

of other variables are less significant.

Table 5.3: Case Study. MSE given by ADD-SCE[(a)].

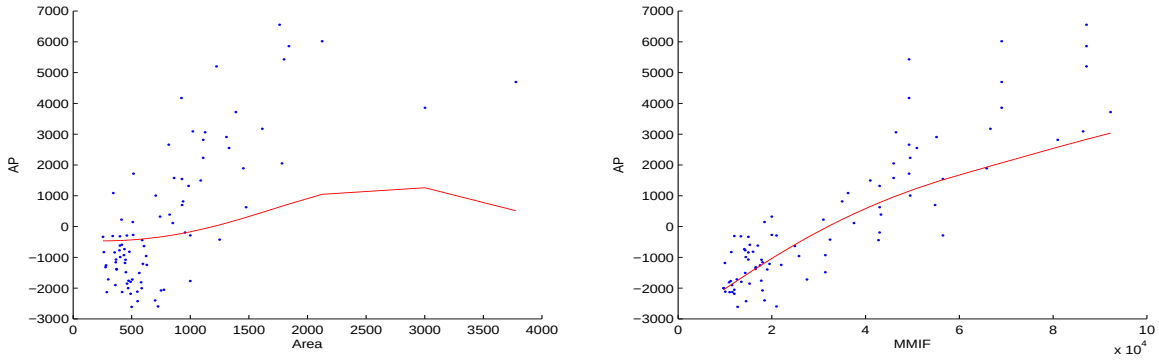
	ADD[(b)]	**Torus**[(c)]	**Separable**[(d)]	**Unilateral**[(e)]
Area & AoB	1.2659×10^6	1.1965×10^6	1.2223×10^6	1.2218×10^6
Area & AHS	1.5673×10^6	1.5293×10^6	1.5385×10^6	1.5345×10^6
Area & MMIAP	1.5115×10^6	1.4605×10^6	1.4783×10^6	1.4733×10^6
Area & MMIF	1.2500×10^6	1.0954×10^6	1.0388×10^6	9.5106×10^5
AoB & AHS	1.8354×10^6	1.6932×10^6	1.7083×10^6	1.7155×10^6
AoB & MMIAP	3.5163×10^6	2.4779×10^6	2.3914×10^6	2.4826×10^6
AoB & MMIF	1.0769×10^6	9.7906×10^5	9.7614×10^5	9.7465×10^5
AHS & MMIAP	1.9140×10^6	1.8211×10^6	1.8081×10^6	1.7893×10^6
AHS & MMIF	1.2708×10^6	1.1088×10^6	1.0946×10^6	1.0751×10^6
MMIAP & MMIF	1.1207×10^6	1.0103×10^6	9.9022×10^5	9.8214×10^5

[(a)] ADD-SCE refers to two-step estimation considered in Chapter 3.
[(b)] ADD refers to the fitting of additive model with marginal integration technique.
[(c)] Torus: ADD-SCE with ϵ_s satisfying (2.2).
[(d)] Separable: ADD-SCE with ϵ_s satisfying (2.3).
[(e)] Unilateral: ADD-SCE with ϵ_s satisfying (2.4).

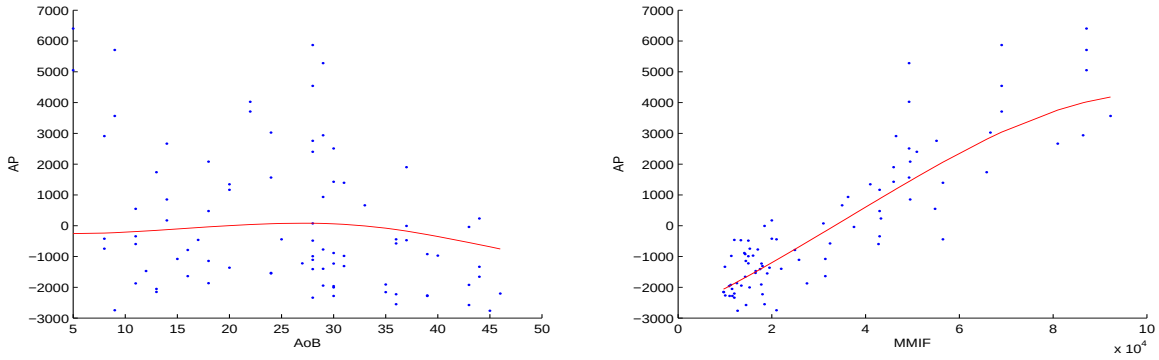
Table 5.4: Case Study. Estimators of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ given by ADD-SCE[(*)].

	Torus[(*)]		**Separable**[(*)]		**Unilateral**[(*)]	
Area&AoB	0.0870	0.0254	0.0709	0.0175	0.1534	0.0277
Area&AHS	0.0436	0.0054	0.0305	0.0220	0.0525	0.0069
Area&MMIAP	0.0584	-0.0032	0.0289	0.0079	0.0500	-0.0221
Area&MMIF	0.1373	0.1390	0.1147	0.1586	0.1101	0.1013
AoB&AHS	0.0674	0.0932	0.0643	0.0814	0.1335	0.1531
AoB&MMIAP	0.0977	0.2206	0.0970	0.2091	0.2141	0.3844
AoB&MMIF	0.0375	0.1051	0.0278	0.1152	0.0383	0.1617
AHS&MMIAP	0.0587	0.0739	0.0621	0.0702	0.0685	0.0989
AHS&MMIF	0.0561	0.1186	0.0514	0.1378	0.0524	0.1621
MMIAP&MMIF	0.0661	0.1159	0.0685	0.1166	0.0591	0.1629

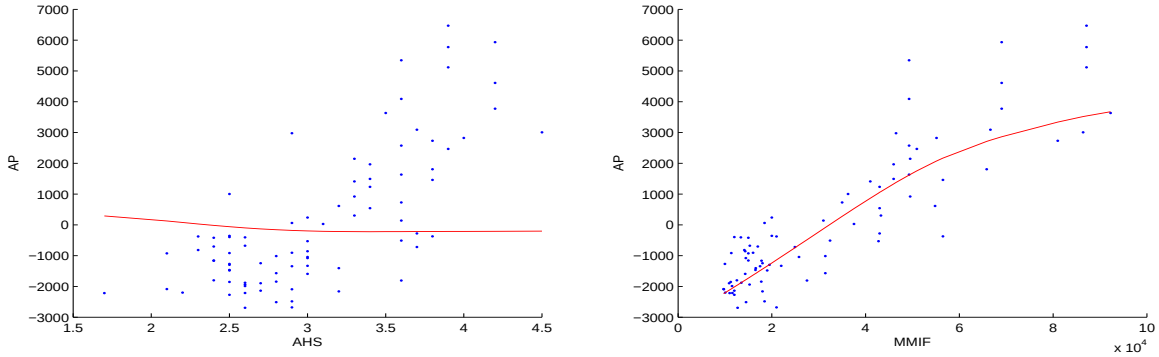
[(*)] ADD-SCE, Torus, Separable and Unilateral are same defined as the ones in Table 5.3.



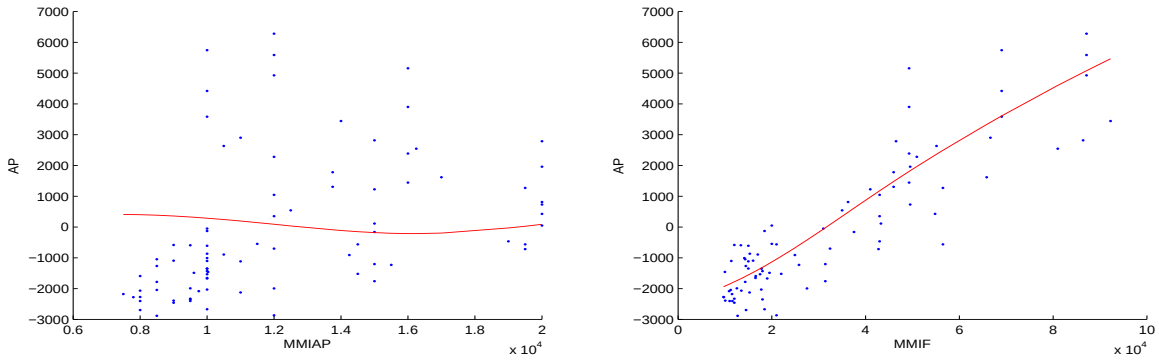
(a) AP vs. Area&MMIF



(b) AP vs. AoB&MMIF



(c) AP vs. AHS&MMIF



(d) AP vs. MMIAP&MMIF

Figure 5.3: Case study. Graphs for part of the estimators with ADD-SCE fitting.

5.4 Fitting with PLR-SCE

Suffering from the curse of dimensionality, we should avoid adding more explanatory variables into the unknown regression function. Therefore, if we want to introduce more explanatory variables into modelling, we may consider the partially linear regression model. As we have mentioned, the AP variable may increase with Area, AHS and MMIF linearly or approximately linearly. Hence, we considered the models with one explanatory variable in the linear part and two explanatory variables in the nonparametric function $m(.)$. All the estimation results are given in Table 5.5, Table 5.6 and Table 5.7. In Table 5.5, the MSE of the fitting are given. The estimators of λ, which is the coefficient of the linear variable, are given in Table 5.6. And the estimators of $\boldsymbol{\theta}$, which contains the coefficients of the SGAR model, are given in Table 5.7. Each table is divided into three sections according to different linear explanatory variables in the fitting. From these tables, we have the following comments.

First, it is not a good choice to put MMIF as a linear variable. Comparing the lower section of Table 5.1 and the last section of Table 5.5, we can see that the MSE given in the later are even larger than the ones given in the former. For example, with explanatory variables MMIAP & MMIF, the MSE given by local linear fitting is 4.8321×10^5, see the last line of Table 5.1. However, in the last section of Table5.5 where MMIF is used as a linear explanatory variable, we can see that the MSE with MMIAP, MMIF and any one more other variable are much larger than 4.8321×10^5. This indicates that using MMIF as a linear variable will greatly worsen the estimation results.

Second, the fitting results are improved by adding Area as a linear variable. This can be seen by a quick comparison between the lower section of Table 5.1 and the first section

of Table 5.5. Moreover, as we have mentioned, there are two isolated points for the Area variable. This causes some trouble for the bandwidth selection in the kernel estimation. So using Area as a linear variable may be a wise choice.

Third, when MMIF is used as a variable in the non-parametric part, adding AHS as a linear variable will not reduce the MSE. In Table 5.1, the MSE given by local linear fitting with AoB & MMIF is 6.2371×10^5, while in Table 5.5, the MSE given by partially linear fitting with AHS being a linear explanatory variable and AoB & MMIF being the non-parametric ones is 6.3841×10^5. The later is even larger. Such result also happens in the case with MMIAP & MMIF.

Next, AoB and MMIAP contribute less than others to the fitting. This conclusion can be drawn from Table 5.6. We can see that the estimates of λ are very large when AoB & MMIAP are the explanatory variables in the non-parametric part. The large value of λ indicates that the contribution of the linear explanatory variable to the fitting is large, and therefore the contribution of variables in the non-parametric part is comparatively small.

At last, we can see that our modified fitting performs better than the traditional fitting of partially linear model in all cases. The best fitting is with Area, MMIAP and MMIF, Area being the linear explanatory variable.

Table 5.5: Case Study. MSE given by PLR-SCE[a].

Linear : Area	**PLR**[b]	**Torus**[c]	**Separable**[d]	**Unilateral**[e]
AoB & AHS	1.2350×10^6	1.1765×10^6	1.1817×10^6	1.1989×10^6
AoB & MMIAP	1.1039×10^6	1.0829×10^6	1.0814×10^6	1.0716×10^6
AoB & MMIF	6.5051×10^5	5.5511×10^5	5.4509×10^5	5.4622×10^5
AHS & MMIAP	8.8348×10^5	7.6977×10^5	7.6543×10^5	7.6059×10^5
AHS & MMIF	7.9095×10^5	7.2979×10^5	7.0057×10^5	6.0755×10^5
MMIAP & MMIF	4.5226×10^5	4.7097×10^5	4.1717×10^5	4.1440×10^5
Linear : AHS	**PLR**	**Torus**	**Separable**	**Unilateral**
Area & AoB	8.9588×10^5	8.1833×10^5	8.1856×10^5	8.1758×10^5
Area & MMIAP	1.0888×10^6	1.0387×10^6	1.0418×10^6	1.0612×10^6
Area & MMIF	6.7824×10^5	5.7568×10^5	5.8360×10^5	6.2828×10^5
AoB & MMIAP	1.3716×10^6	1.0100×10^6	1.0210×10^6	1.0401×10^6
AoB & MMIF	6.3841×10^5	6.2657×10^5	6.3014×10^5	6.2122×10^5
MMIAP & MMIF	5.0032×10^5	5.1220×10^5	4.3948×10^5	4.3292×10^5
Linear : MMIF	**PLR**	**Torus**	**Separable**	**Unilateral**
Area & AoB	8.6521×10^5	8.1427×10^5	8.2329×10^5	8.0463×10^5
Area & AHS	1.1729×10^6	1.1103×10^6	1.1115×10^6	1.1496×10^6
Area & MMIAP	1.1636×10^6	1.1168×10^6	1.1192×10^6	1.1468×10^6
AoB & AHS	1.2460×10^6	1.1329×10^6	1.1341×10^6	1.1382×10^6
AoB & MMIAP	1.2558×10^6	1.0284×10^6	1.0276×10^6	1.0289×10^6
AHS & MMIAP	7.7978×10^5	7.4317×10^5	7.4668×10^5	7.3826×10^5

[a] PLR-SCE refers to the two-step estimation considered in Chapter 4.
[b] PLR refers to the partially linear fitting.
[c] Torus: PLR-SCE with ϵ_s satisfying (2.2).
[d] Separable: PLR-SCE with ϵ_s satisfying (2.3).
[e] Unilateral: PLR-SCE with ϵ_s satisfying (2.4).

Table 5.6: Case Study. Estimators of λ given by PLR-SCE$^{(*)}$.

Linear : Area	**PLR**$^{(*)}$	**Torus**$^{(*)}$	**Separable**$^{(*)}$	**Unilateral**$^{(*)}$
AoB & AHS	0.6149	0.5004	0.5133	0.5083
AoB & MMIAP	2.0669	2.0106	2.0474	1.9957
AoB & MMIF	0.2094	0.1630	0.1599	0.1638
AHS & MMIAP	0.3064	0.2535	0.2509	0.2433
AHS & MMIF	0.1478	0.1397	0.1337	0.0956
MMIAP & MMIF	0.0358	0.0335	0.0357	0.0305
Linear : AHS	**PLR**	**Torus**	**Separable**	**Unilateral**
Area & AoB	356.8348	342.6098	346.8671	332.5619
Area & MMIAP	369.3316	368.4477	369.7063	367.1023
Area & MMIF	53.6007	37.4678	40.4183	32.6920
AoB & MMIAP	1501.4675	1274.5000	1266.0801	1269.5656
AoB & MMIF	−6.5104	−7.6320	−6.2872	−11.7225
MMIAP & MMIF	−33.2386	−38.0151	−53.5872	−55.0966
Linear : MMIF	**PLR**	**Torus**	**Separable**	**Unilateral**
Area & AoB	0.0136	0.0131	0.0133	0.0128
Area & AHS	0.0176	0.0164	0.0165	0.0166
Area & MMIAP	0.0184	0.0176	0.0176	0.0177
AoB & AHS	0.0151	0.0141	0.0141	0.0139
AoB & MMIAP	0.0468	0.0433	0.0432	0.0431
AHS & MMIAP	0.0121	0.0118	0.0117	0.0116

$^{(*)}$ PLR-SCE, PLR, Torus, Separable and Unilateral are same defined as the ones in Table 5.5.

Table 5.7: Case Study. Estimators of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ given by PLR-SCE$^{(*)}$.

Linear : Area	**Torus**$^{(*)}$		**Separable**$^{(*)}$		**Unilateral**$^{(*)}$	
AoB & AHS	0.0827	0.0575	0.0741	0.0512	0.0691	0.0484
AoB & MMIAP	0.0643	0.0387	0.0567	0.0301	0.0820	0.0394
AoB & MMIF	0.0571	0.0408	0.0563	0.0435	0.0610	0.0588
AHS & MMIAP	-0.0153	0.0526	-0.0216	0.0517	-0.0020	0.0713
AHS & MMIF	0.0451	0.0295	0.0480	0.0219	0.0635	0.0266
MMIAP & MMIF	0.0055	0.0223	-0.0567	0.0430	-0.0435	0.0699
Linear : AHS	**Torus**		**Separable**		**Unilateral**	
Area & AoB	0.0645	0.0577	0.0626	0.0558	0.0704	0.0502
Area & MMIAP	0.0565	-0.0151	0.0537	-0.0137	0.0291	-0.0006
Area & MMIF	0.1559	-0.0273	0.1517	-0.0218	0.1379	-0.0026
AoB & MMIAP	0.0943	0.1598	0.0943	0.1536	0.0783	0.1495
AoB & MMIF	0.0464	0.0285	0.0385	0.0263	0.0446	0.0472
MMIAP & MMIF	0.0134	0.0119	-0.0299	0.0395	-0.0213	0.0664
Linear : MMIF	**Torus**		**Separable**		**Unilateral**	
Area & AoB	0.0716	0.0467	0.0700	0.0473	0.0744	0.0450
Area & AHS	0.0593	-0.0327	0.0585	-0.0318	0.0445	-0.0203
Area & MMIAP	0.0511	-0.0122	0.0491	-0.0108	0.0238	-0.0003
AoB & AHS	0.0888	0.0925	0.0885	0.0891	0.0774	0.0847
AoB & MMIAP	0.0302	0.1944	0.0275	0.1949	0.0471	0.1898
AHS & MMIAP	-0.0373	0.0708	-0.0344	0.0708	-0.0112	0.0972

$^{(*)}$ PLR-SCE, Torus, Separable and Unilateral are same defined as the ones in Table 5.5.

5.5 Summary and Remark

In this chapter, we examine the house pricing problem in the Hong Kong real estate market. We aim to identify the main factors that affect the average price per feet^2 of a house from a group of variables. The nonparametric model, additive model and semi-parametric model are adopted for analysis, and the two-step estimations introduced in Chapter 2 to Chapter 4 are used. With the fitting results, we show that our proposed methods perform better than the traditional estimations in all cases. Specially, when we fail to select the key explanatory variables into the fitting, our methods can improve the estimation results greatly. Hence, in the case that data of some key explanatory variable are not available or the key explanatory variables are hard to be identified, our estimation methods can play an important role. Moreover, the results also show that AoB and MMIAP have minor contribution to the fitting, AHS may be highly correlated with MMIF, and a combination of Area, MMIAP and MMIF with Area being a linear explanatory variable and the others being non-parametric ones can determine the average price of a house well.

Chapter 6

Conclusion

In our thesis, we focus on the nonlinear regression estimation with spatial data. For the large scale variation of the model, we adopt the nonparametric, additive nonparametric and semi-parametric structure; for the small scale variation of the model, we assume the errors satisfy the torus, separable or unilateral SGAR model. The former is applied to figure the nonstationary mean (trend of factor) effect, while the later models the spatial dependency. The explanatory variables, assumed to be exogenous, are independent of the errors. By taking both exogenous influence and spatial dependency into account, our proposed models overmatch two types of potential competitors: models with linear structure and spatial correlated errors, and models with nonlinear structure and *i.i.d*. errors.

The local linear estimation technique is applied in all of the three models. For the additive model, the marginal integration approach is used. Moreover, throughout this thesis, we adopt the two-step estimation procedure considered in Martins-Filho & Yao (2009). This procedure firstly forms a new process with the same conditional mean

as the original one and *i.i.d*. errors, and then applies the traditional fitting to the new process. Martins-Filho & Yao (2009) considered such two-step estimation only in the time series anaylsis, and it is meaningful for us to introduce the method into the spatial area. Asymptotic properties of both first- and second-step estimators are considered in our thesis. For the first-step estimators of the unknown regression function, the convergence rate with all three types of errors is considered, and when errors satisfy the separable or unilateral SGAR model, the asymptotic normality is established. For the second-step estimators of the unknown regression function, the asymptotic normality with three types of error structure is established. In the semi-parametric model, we also establish the asymptotic normality of the first- and second-step estimators of the linear parameters.

Simulations are conducted to assess the performance of our fitting. The results show that our estimation works better than the traditional methods, given that spatially correlated errors exist. The improvement of our estimation is significant when the volatility of the errors is large. As an illustration of our approach, a case study of the housing price of Hong Kong is given. It is shown that our approach improves the estimation, especially when some key factor is absent in the modelling.

Appendix

A.1 Matlab Code for the estimation of SGAR models

```
function [I_B,P,Q,r2]=Estimate_torus_IB(data,n1,n2)
% Estimating the coefficients of the torus SGAR model
Sdata(n1,n2)=0;
for i=1:n1
  for j=1:n2
    Sdata(i,j)=data((i-1)*n2+j);
  end
end
Y=data;
for i=1:n1
  for j=1:n2
    l=0;
    if i==1&&j==1
      X(1,(i-1)*n2+j)=Sdata(i+n1-1,j);
      X(2,(i-1)*n2+j)=Sdata(i,j+n2-1);
      X(3,(i-1)*n2+j)=Sdata(i+1,j);
      X(4,(i-1)*n2+j)=Sdata(i,j+1);
      l=2;
    end
    if i==1&&j==n2
      X(1,(i-1)*n2+j)=Sdata(i+n1-1,j);
      X(2,(i-1)*n2+j)=Sdata(i,j-1);
      X(3,(i-1)*n2+j)=Sdata(i+1,j);
      X(4,(i-1)*n2+j)=Sdata(i,j-n2+1);
      l=2;
    end
    if i==n1&&j==1
      X(1,(i-1)*n2+j)=Sdata(i-1,j);
      X(2,(i-1)*n2+j)=Sdata(i,j+n2-1);
      X(3,(i-1)*n2+j)=Sdata(i-n1+1,j);
      X(4,(i-1)*n2+j)=Sdata(i,j+1);
      l=2;
    end
    if i==n1&&j==n2
      X(1,(i-1)*n2+j)=Sdata(i-1,j);
      X(2,(i-1)*n2+j)=Sdata(i,j-1);
      X(3,(i-1)*n2+j)=Sdata(i-n1+1,j);
      X(4,(i-1)*n2+j)=Sdata(i,j-n2+1);
      l=2;
    end
    if l~=2
      if i==1
```

```
            X(1,(i-1)*n2+j)=Sdata(i+n1-1,j);
            X(2,(i-1)*n2+j)=Sdata(i,j-1);
            X(3,(i-1)*n2+j)=Sdata(i+1,j);
            X(4,(i-1)*n2+j)=Sdata(i,j+1);
            l=1;
          end
          if j==1
            X(1,(i-1)*n2+j)=Sdata(i-1,j);
            X(2,(i-1)*n2+j)=Sdata(i,j+n2-1);
            X(3,(i-1)*n2+j)=Sdata(i+1,j);
            X(4,(i-1)*n2+j)=Sdata(i,j+1);
            l=1;
          end
          if j==n2
            X(1,(i-1)*n2+j)=Sdata(i-1,j);
            X(2,(i-1)*n2+j)=Sdata(i,j-1);
            X(3,(i-1)*n2+j)=Sdata(i+1,j);
            X(4,(i-1)*n2+j)=Sdata(i,j-n2+1);
            l=1;
          end
          if i==n1
            X(1,(i-1)*n2+j)=Sdata(i-1,j);
            X(2,(i-1)*n2+j)=Sdata(i,j-1);
            X(3,(i-1)*n2+j)=Sdata(i-n1+1,j);
            X(4,(i-1)*n2+j)=Sdata(i,j+1);
            l=1;
          end
        end
        if l==0
          X(1,(i-1)*n2+j)=Sdata(i-1,j);
          X(2,(i-1)*n2+j)=Sdata(i,j-1);
          X(3,(i-1)*n2+j)=Sdata(i+1,j);
          X(4,(i-1)*n2+j)=Sdata(i,j+1);
        end
      end
    end
    [theta,r2]=torus_SAR_ML_f(n1,n2,X,Y);
    P=theta(1);
    Q=theta(2);
    n=n1*n2;
    B=zeros(n,n);
    for i=1:n
      l=0;
      if i==1
        B(i,n+i-n2)=P;
        B(i,i+n2-1)=Q;
        B(i,i+n2)=P;
        B(i,i+1)=Q;
        l=2;
      end
      if i==n2
        B(i,n+i-n2)=P;
        B(i,i-1)=Q;
        B(i,i+n2)=P;
        B(i,i-n2+1)=Q;
        l=2;
```

```
    end
    if i==n-n2+1
      B(i,i-n2)=P;
      B(i,i+n2-1)=Q;
      B(i,i-n+n2)=P;
      B(i,i+1)=Q;
      l=2;
    end
    if i==n
      B(i,i-n2)=P;
      B(i,i-1)=Q;
      B(i,i-n+n2)=P;
      B(i,i-n2+1)=Q;
      l=2;
    end
    if l~=2
      if i<n2
        B(i,n+i-n2)=P;
        B(i,i-1)=Q;
        B(i,i+n2)=P;
        B(i,i+1)=Q;
        l=1;
      end
      if mod(i,n2)==1
        B(i,i-n2)=P;
        B(i,i+n2-1)=Q;
        B(i,i+n2)=P;
        B(i,i+1)=Q;
        l=1;
      end
      if i>n-n2
        B(i,i-n2)=P;
        B(i,i-1)=Q;
        B(i,i-n+n2)=P;
        B(i,i+1)=Q;
        l=1;
      end
      if mod(i,n2)==0
        B(i,i-n2)=P;
        B(i,i-1)=Q;
        B(i,i+n2)=P;
        B(i,i-n2+1)=Q;
        l=1;
      end
    end
    if l==0
      B(i,i-n2)=P;
      B(i,i-1)=Q;
      B(i,i+n2)=P;
      B(i,i+1)=Q;
    end
  end
  I_B=diag(ones(n,1))-B;

  function [theta,roul2]=torus_SAR_ML_f(n1,n2,Z,Y)
  % Sub-function of 'Estimate_torus_IB'.
```

```
n=n1*n2;
N=[0 -1;-1 0;0 1;1 0]';
[sz1,sz2]=size(N);
C(sz2,n)=0;
S(sz2,n)=0;
s(:,n)=[0 0]';
k=1;
for i=1:n1
  for j=1:n2
    s(:,k)=[i j]';
    k=k+1;
  end
end
for k=1:n
  for i=1:sz2
    C(i,k)=cos(2*pi*s(1,k)*N(2,i)/n1+2*pi*s(2,k)*N(1,i)/n2);
    S(i,k)=sin(2*pi*s(1,k)*N(2,i)/n1+2*pi*s(2,k)*N(1,i)/n2);
  end
end
V=0;R=0;
for k=1:n
  V=V+C(:,k);
  R=R+(S(:,k)*(S(:,k))'-C(:,k)*(C(:,k))');
end
X=0;U=0;
for k=1:n
  X=X+Z(:,k)*(Z(:,k))';
  U=U+Z(:,k)*Y(k);
end
Theta=X^(-1)*U;
l=0;
while l==0
  roul2=0;
  for k=1:n
    roul2=roul2+(Y(k)-Theta'*Z(:,k))^2;
  end
  roul2=roul2/n;
  theta=(R-X/roul2)^(-1)*(V-U/roul2);
  if sum(abs(theta-Theta))<0.00000001
    l=1;
  else
    Theta=theta;
    Roul2=roul2;
  end
end

function [I_B,P,Q,r2]=Estimate_bilateral_IB(E,n1,n2,p,q)
% Estimating the coefficients of the separable SGAR model
n=n1*n2;
I1=diag(ones(1,n1));I2=diag(ones(1,n2));
U1(n1,n1)=0;U2(n2,n2)=0;
for i=2:n1
  U1(i,i-1)=1;
end
for i=2:n2
  U2(i,i-1)=1;
```

```
end
IB=@(C)(kron(I1,I2)-C(1)*kron(U1+U1',I2)-C(2)*kron(I1,U2+U2')-
  C(1)*C(2)*kron(U1+U1',U2+U2'));
for i=1:n1
  R1(i,1)=2*cos(pi*i/(n1+1));
end
for i=1:n2
  R2(i,1)=2*cos(pi*i/(n2+1));
end
L=@(C)(E'*(kron(I1,I2)-C(1)*kron(U1+U1',I2)-C(2)*kron(I1,U2+U2')-
  C(1)*C(2)*kron(U1+U1',U2+U2'))'*(kron(I1,I2)-C(1)*kron(U1+U1',I2)-
  C(2)*kron(I1,U2+U2')-C(1)*C(2)*kron(U1+U1',U2+U2'))*E*
  (prod(prod(ones(n1,n2)-C(1)*R1*ones(1,n2)-C(2)*ones(n1,1)*R2'-
  C(1)*C(2)*R1*R2')))^(-2/n));
temp=fminsearch(L,[p,q]);
P=temp(1);Q=temp(2);
I_B=IB([P,Q]);
r2=E'*I_B'*I_B*E/n;

function [I_B,P,Q,r2]=Estimate_unilateral_IB(E,n1,n2)
% Estimating the coefficients of the unilateral SGAR model
E=E-mean(E);
for i=1:n1
  ES(i,1:n2)=E((i-1)*n2+1:i*n2);
end
C00=mean(E.^2);
C10=mean(mean(ES(1:n1-1,:).*ES(2:n1,:)));
C01=mean(mean(ES(:,1:n2-1).*ES(:,2:n2)));
C11=mean(mean(ES(2:n1,1:n2-1).*ES(1:n1-1,2:n2)));
P=(C10*C00-C01*C11)/(C00^2-C11^2);
Q=(C10-P*C00)/C11;
r2=(1-P^2-Q^2)*C00-2*P*Q*C11;
U1=zeros(n1,n1);U2=zeros(n2,n2);
I1=diag(ones(n1,1));I2=diag(ones(n2,1));
for i=2:n1
  U1(i,i-1)=1;
end
for i=2:n2
  U2(i,i-1)=1;
end
I_B=kron(I1,I2)-P*kron(U1,I2)-Q*kron(I1,U2);
```

A.2 Matlab Code for LLR-SCE

```
function [Beta_O,Beta_M,P,Q]=LLR_SCE_f(Y,X,n1,n2,d,h1,h2,MD,p,q)
% Main function for the LLR-SCE fitting considered in Chp 2.
n=n1*n2;
Beta_O=LLR_f(Y,X,n,d,h1);
E=Y-Beta_O(:,1);
switch MD
  case {1}
    [I_B,P,Q,r2]=Estimate_torus_IB(E,n1,n2);
  case {2}
    [I_B,P,Q,r2]=Estimate_bilateral_IB(E,n1,n2,p,q);
```

```
    case {3}
      [I_B,P,Q,r2]=Estimate_unilateral_IB(E,n1,n2);
end
Z=I_B*Y+(diag(ones(1,n))-I_B)*Beta_O(:,1);
Beta_M=LLR_f(Z,X,n,d,h2);

function Beta=LLR_f(Y,X,n,d,h)
% Local linear fitting.
Beta=zeros(n,d+1);
for l=1:n
   x=X(l,:);
   Xh=[ones(n,1) (X-ones(n,1)*x)./(ones(n,1)*h)];
   K=ones(1,n);
   for i=1:n
      for j=1:d
        K(i)=K(i)*(2*pi*h(j)^2)^(-1/2)*exp(-1/2*((X(i,j)-
          x(j))/h(j))^2);
      end
   end
   W=diag(K);
   Beta_temp=(Xh'*W*Xh)^(-1)*(Xh'*W*Y);
   Beta(l,1)=Beta_temp(1);
   Beta(l,2:d+1)=Beta_temp(2:d+1)./h';
end

function h=BS_LLR_f(Y,X,n,d)
% Bandwidth selection for local linear fitting.
A=(max(X)-min(X))/40; % foot-step of bandwidth
B=(max(X)-min(X))/2;
A0=2*A;   % initial value of the bandwidth
C=(B-A0)./A;
l(d)=0;
j=1;
while l(d)<=C(d)
   for i=1:d
      h(i)=A0(i)+l(i)*A(i);
   end
   Cv(j,1)=CV_LLR_f(Y,X,n,d,h);
   Cv(j,2:d+1)=h;
   j=j+1;
   l(1)=l(1)+1; k=1;
   if l(1)>C(1)
      for i=2:d-1
         if l(i)==C(i)
            k=i;
         else
            break
         end
      end
      if k+1<=d
         l(k+1)=l(k+1)+1;
         l(1:k)=zeros(1,k);
      end
   end
end
MIN=min(Cv(:,1));
```

```
for i=1:j-1
  if Cv(i,1)==MIN
    h=Cv(i,2:d+1);
    break
  end
end

function MSE=CV_LLR_f(Y,X,n,d,h)
% Sub-function of 'BS_LLR_f'.
MSE=0;
for i=1:n
  Yp=CV_LLR_predict_f(Y,X,n,i,d,h);
  MSE=MSE+(Y(i)-Yp)^2;
end
MSE=MSE/n;

function [h1,h2]=BS_LLR_SCE_iteration_f(Y,X,d,h1,n1,n2,MD,p,q)
% Bandwidth selection for LLR-SCE.
% Some initial h1 should be given.
A=(max(X)-min(X))/40; % foot-step of bandwidth
B=(max(X)-min(X))/2;
A0=2*A; % initial value of bandwidth
C=(B-A0)./A;
j=1;H1=0;H2=0;G='just go';
while G=='just go'
  l=0;l(d)=0;J=j;
  while l(d)<=C(d)
    for i=1:d
      h2(i)=A0(i)+l(i)*A(i);
    end
    Cv(j,1)=DCV_LLR_SCE_f(Y,X,n1,n2,d,h1,h2,MD,p,q);
    Cv(j,2:d+1)=h1;
    Cv(j,d+2:2*d+1)=h2;
    j=j+1;
    l(1)=l(1)+1;k=1;
    if l(1)>C(1)
      for i=2:d-1
        if l(i)==C(i)
          k=i;
        else
          break
        end
      end
      if k+1<=d
        l(k+1)=l(k+1)+1;
        l(1:k)=zeros(1,k);
      end
    end
  end
  MIN=min(Cv(J:j-1,1));
  for i=J:j-1
    if Cv(i,1)==MIN
      h2=Cv(i,d+2:2*d+1);
      break
    end
  end
```

```
    if H2==h2
      FCv=Cv(J:j-1,:);
      break
    else
      H2=h2;
    end
    l=0;l(d)=0;J=j;
    while l(d)<=C(d)
      for i=1:d
        h1(i)=A0(i)+l(i)*A(i);
      end
      Cv(j,1)=DCV_LLR_SCE_f(Y,X,n1,n2,d,h1,h2,MD,p,q);
      Cv(j,2:d+1)=h1;
      Cv(j,d+2:2*d+1)=h2;
      j=j+1;
      l(1)=l(1)+1;k=1;
      if l(1)>C(1)
        for i=2:d-1
          if l(i)==C(i)
            k=i;
          else
            break
          end
        end
        if k+1<=d
          l(k+1)=l(k+1)+1;
          l(1:k)=zeros(1,k);
        end
      end
    end
    MIN=min(Cv(J:j-1,1));
    for i=J:j-1
      if Cv(i,1)==MIN
        h1=Cv(i,2:d+1);
        break
      end
    end
    if H1==h1
      FCv=Cv(J:j-1,:);
      break
    else
      H1=h1;
    end
end

function MSE=DCV_LLR_SCE_f(Y,X,n1,n2,d,h1,h2,MD,p,q)
% Sub-function of 'BS_LLR_SCE_iteration_f'.
n=n1*n2;
for i=1:n
  Y1(i,1)=CV_LLR_predict_f(Y,X,n,i,d,h1);
end
E1=Y-Y1;
switch MD
  case {1}
    [I_B,P,Q,r2]=Estimate_torus_IB(E1,n1,n2);
  case {2}
```

```
    [I_B,P,Q,r2]=Estimate_bilateral_IB(E1,n1,n2,p,q);
  case {3}
    [I_B,P,Q,r2]=Estimate_unilateral_IB(E1,n1,n2);
end
Z=Y1+I_B*E1;
for i=1:n
  Y2(i,1)=CV_LLR_predict_f(Z,X,n,i,d,h2);
end
E2=Y-Y2;
switch MD
  case {1}
    [I_B,P,Q,r2]=Estimate_torus_IB(E2,n1,n2);
  case {2}
    [I_B,P,Q,r2]=Estimate_bilateral_IB(E2,n1,n2,p,q);
  case {3}
    [I_B,P,Q,r2]=Estimate_unilateral_IB(E2,n1,n2);
end
temp=I_B'*I_B;
tD=0;tD(n,1)=0;
for i=1:n
  tD(i,1)=temp(i,i);
end
C_m=diag(ones(1,n))-temp/mean(tD);
E3=mean(E2)+C_m*(E2-mean(E2));
MSE=mean((Y-Y2-E3).^2);

function Beta=CV_LLR_predict_f(Y,X,n,i,d,h)
% Sub-function of 'CV_LLR_f' and 'DCV_LLR_SCE_f'.
U=0;V=0;
for j=1:n
  if j~=i
    K=1;
    for k=1:d
      K=K*((2*pi*h(k)^2)^(-1/2)*exp(-1/2*((X(j,k)-X(i,k))/h(k))^2));
    end
    Xh=[1 (X(j,:)-X(i,:))./h];
    U=U+K*Xh'*Xh;
    V=V+Y(j)*Xh'*K;
  end
end
Beta=[1 zeros(1,d)]*U^(-1)*V;
```

A.3 Matlab Code for ADD-SCE

```
function [SM1,mu1,M1,SM2,mu2,M2,P,Q]=Additive_SCE_f(Y,
  X,n1,n2,b,d,h1,h2,MD,p,q)
% Main function of the ADD-SCE fitting considered in Chp 3.
n=n1*n2;
[SM1,mu1,M1]=Additive_f(Y,X,n,d,b,h1);
E=Y-SM1;
switch MD
  case {1}
    [I_B,P,Q,r2]=Estimate_torus_IB(E,n1,n2);
  case {2}
```

```
      [I_B,P,Q,r2]=Estimate_bilateral_IB(E,n1,n2,p,q);
   case {3}
      [I_B,P,Q,r2]=Estimate_unilateral_IB(E,n1,n2);
end
Z=I_B*Y+(diag(ones(1,n))-I_B)*SM1;
[SM2,mu2,M2]=Additive_f(Z,X,n,d,b,h2);

function [SM,mu,M]=Additive_f(Y,X,n,d,b,h)
% Fitting of additive models with marginal integration.
mu=mean(Y);
D=sum(b);
for i=1:d
   low_bound=sum(b(1:i-1))+1;
   up_bound=sum(b(1:i));
   for j=1:n
     MX=zeros(n,1);
     for k=1:n
        x=X(k,:); x(low_bound:up_bound)=X(j,low_bound:up_bound);
        Xh1=(X-ones(n,1)*x)./(ones(n,1)*h);
        Xh=[ones(n,1) Xh1];
        K=prod(((2*pi)^(-1/2)*exp(-Xh1.^2/2)./(ones(n,1)*h))');
        W=diag(K);
        MX(k,1)=[1 zeros(1,D)]*(Xh'*W*Xh)^(-1)*(Xh'*W*Y);
     end
     M(j,i)=mean(MX)-mu;
   end
end
SM=mu+(sum(M'))';

function h=BS_Additive_f(Y,X,n,b,d)
% Bandwidth selection of the additive models.
D=sum(b);
A=(max(X)-min(X))/40;  % foot-step of bandwidth
B=(max(X)-min(X))/2;
A0=2*A;   % A0 is the initial value of the bandwidth
C=(B(1)-A0(1))/A(1);
for i=1:C
   h=A0+(i-1)*A;
   Cv(i,1)=CV_Additive_f(Y,X,n,b,d,h,D);
   Cv(i,2:D+1)=h;
end
MIN=min(Cv(:,1));
for j=1:C
   if Cv(j,1)==MIN
     h=Cv(j,2:D+1);
     break
   end
end

function MSE=CV_Additive_f(Y,X,n,b,d,h,D)
% Sub-function of 'BS_Additive_f'.
MSE=0;
for l=1:n
   Yp=CV_Additive_predict_f(Y,X,n,l,b,d,h,D);
   MSE=MSE+(Y(l)-Yp)^2;
end
```

```
MSE=MSE/n;

function Yp=CV_Additive_predict_f(Y,X,n,l,b,d,h,D)
% Sub-function of 'CV_Additive_f'.
Yn=Y([1:l-1 l+1:n],1);
Xn=X([1:l-1 l+1:n],:);
mu=mean(Yn);
M=0;
for i=1:d
   low_bound=sum(b(1:i-1))+1;
   up_bound=sum(b(1:i));
   MX=zeros(n-1,1);
   for k=1:n-1
      x=Xn(k,:); x(low_bound:up_bound)=X(l,low_bound:up_bound);
      Xh1=(Xn-ones(n-1,1)*x)./(ones(n-1,1)*h);
      Xh=[ones(n-1,1) Xh1];
      K=prod(((2*pi)^(-1/2)*exp(-Xh1.^2/2)./(ones(n-1,1)*h))');
      W=diag(K);
      MX(k,1)=[1 zeros(1,D)]*(Xh'*W*Xh)^(-1)*(Xh'*W*Yn);
   end
   M(i)=mean(MX)-mu;
end
Yp=mu+sum(M);
```

A.4 Matlab Code for PLR-SCE

```
function [Beta_O,Beta_M,Mu_M,L_M,P,Q]=Semi_LLR_SCE_f(Y,
   X1,X2,n1,n2,b,d,h1,h2,MD,p,q)
% Main function of the PLR-SCE fitting considered in Chp 4.
n=n1*n2;
[Mu_O,L_O,Beta_O]=Semi_LLR_f(Y,X1,X2,n,b,d,h1);
E=Y-Mu_O-X1*L_O-Beta_O(:,1);
switch MD
   case {1}
      [I_B,P,Q,r2]=Estimate_torus_IB(E,n1,n2);
   case {2}
      [I_B,P,Q,r2]=Estimate_bilateral_IB(E,n1,n2,p,q);
   case {3}
      [I_B,P,Q,r2]=Estimate_unilateral_IB(E,n1,n2);
end
Z=I_B*Y+(diag(ones(n,1))-I_B)*(Mu_O+X1*L_O+Beta_O(:,1));
[Mu_M,L_M,Beta_M]=Semi_LLR_f(Z,X1,X2,n,b,d,h2);

function [Mu1,L1,Beta1]=Semi_LLR_f(Y,X1,X2,n,b,d,h)
% Traditional fitting of partially linear models.
Yb=Y-mean(Y);
X1b=X1-mean(X1);
K=ones(n,n);
for j=1:n
   for i=1:n
      for k=1:d
         K(i,j)=K(i,j)*((2*pi*h(k)^2)^(-1/2)*exp(-1/2*((X2(i,k)-
            X2(j,k))/h(k))^2));
      end
```

```
    end
  end
  U(d+1,d+1,n)=0;
  for j=1:n
    Xh=[ones(n,1) (X2-ones(n,1)*X2(j,:))./(ones(n,1)*h)];
    W=diag(K(:,j));
    U(:,:,j)=Xh'*W*Xh;
  end
  V0(d+1,n)=0;
  for j=1:n
    V0(1,j)=sum(Yb.*K(:,j));
    for k=1:d
      V0(k+1,j)=sum(Yb.*((X2(:,k)-X2(j,k))/h(k)).*K(:,j));
    end
  end
  V(d+1,b,n)=0;
  for j=1:n
    for i=1:b
      V(1,i,j)=mean(X1b(:,i).*K(:,j));
      for k=1:d
        V(k+1,i,j)=mean(X1b(:,i).*((X2(:,k)-X2(j,k))/h(k)).*K(:,j));
      end
    end
  end
  for j=1:n
    D0(j,1)=[1 zeros(1,d)]*U(:,:,j)^(-1)*V0(:,j);
    D(j,:)=[1 zeros(1,d)]*U(:,:,j)^(-1)*V(:,:,j);
  end
  Yc=Yb-D0;
  Xc=X1b-D;
  L1=(Xc'*Xc)^(-1)*(Xc'*Yc);
  Yd=Yb-X1b*L1;
  for j=1:n
    Xh=[ones(n,1) (X2-ones(n,1)*X2(j,:))./(ones(n,1)*h)];
    W=diag(K(:,j));
    Beta_temp=(Xh'*W*Xh)^(-1)*(Xh'*W*Yd);
    Beta1(j,1)=Beta_temp(1);
    Beta1(j,2:d+1)=Beta_temp(2:d+1)./h';
  end
  Mu1=mean(Y)-mean(X1)*L1;

  function h=BS_Semi_LLR_f(Y,X1,X2,n,b,d)
  % Bandwidth selection for traditional partially linear models.
  A=(max(X2)-min(X2))/40;  % foot-step of bandwidth
  B=(max(X2)-min(X2))/2;
  A0=2*A;   % A0 is the initial value of the bandwidth
  C=(B-A0)./A;
  l(d)=0;
  j=1;
  while l(d)<=C(d)
    for i=1:d
      h(i)=A0(i)+l(i)*A(i);
    end
    Cv(j,1)=CV_Semi_LLR_f(Y,X1,X2,n,b,d,h);
    Cv(j,2:d+1)=h;
    j=j+1;
```

```
    l(1)=l(1)+1; k=1;
    if l(1)>C(1)
      for i=2:d-1
        if l(i)==C(i)
          k=i;
        else
          break
        end
      end
      if k+1<=d
        l(k+1)=l(k+1)+1;
        l(1:k)=zeros(1,k);
      end
    end
  end
  MIN=min(Cv(:,1));
  for i=1:j-1
    if Cv(i,1)==MIN
      h=Cv(i,2:d+1);
      break
    end
  end

  function MSE=CV_Semi_LLR_f(Y,X1,X2,n,b,d,h)
  % Sub-function of 'BS_Semi_LLR_f'
  Yb=Y-mean(Y);
  X1b=X1-mean(X1);
  MSE=0;
  for i=1:n
    Yp=CV_Semi_LLR_predict_f(Y,X1,Yb,X1b,X2,n,i,b,d,h);
    MSE=MSE+(Y(i)-Yp)^2;
  end
  MSE=MSE/n;

  function [h1,h2]=BS_Semi_LLR_SCE_iteration_f(Y,
    X1,X2,b,d,h1,n1,n2,MD,p,q)
  % Bandwidth selection for PLR-SCE.
  A=(max(X2)-min(X2))/40;  % foot-step of bandwidth
  B=(max(X2)-min(X2))/2;
  A0=2*A; % A0 is the initial value of the bandwidth.
  C=(B-A0)./A;
  j=1;H1=0;H2=0;G='just go';
  while G=='just go'
    l=0;l(d)=0;J=j;
    while l(d)<=C(d)
      for i=1:d
        h2(i)=A0(i)+l(i)*A(i);
      end
      Cv(j,1)=DCV_Semi_LLR_SCE_f(Y,X1,X2,n1,n2,b,d,h1,h2,MD,p,q);
      v(j,2:d+1)=h1;
      Cv(j,d+2:2*d+1)=h2;
      j=j+1;
      l(1)=l(1)+1;k=1;
      if l(1)>C(1)
        for i=2:d-1
          if l(i)==C(i)
```

```
            k=i;
          else
            break
          end
        end
        if k+1<=d
          l(k+1)=l(k+1)+1;
          l(1:k)=zeros(1,k);
        end
      end
    end
    MIN=min(Cv(J:j-1,1));
    for i=J:j-1
      if Cv(i,1)==MIN
        h2=Cv(i,d+2:2*d+1);
        break
      end
    end
    if H2==h2
      FCv=Cv(J:j-1,:);
      break
    else
      H2=h2;
    end
    l=0;l(d)=0;J=j;
    while l(d)<=C(d)
      for i=1:d
        h1(i)=A0(i)+l(i)*A(i);
      end
      Cv(j,1)=DCV_Semi_LLR_SCE_f(Y,X1,X2,n1,n2,b,d,h1,h2,MD,p,q);
      Cv(j,2:d+1)=h1;
      Cv(j,d+2:2*d+1)=h2;
      j=j+1;
      l(1)=l(1)+1;k=1;
      if l(1)>C(1)
        for i=2:d-1
          if l(i)==C(i)
            k=i;
          else
            break
          end
        end
        if k+1<=d
          l(k+1)=l(k+1)+1;
          l(1:k)=zeros(1,k);
        end
      end
    end
    MIN=min(Cv(J:j-1,1));
    for i=J:j-1
      if Cv(i,1)==MIN
        h1=Cv(i,2:d+1);
        break
      end
    end
    if H1==h1
```

```
    FCv=Cv(J:j-1,:);
    break
  else
    H1=h1;
  end
end

function MSE=DCV_Semi_LLR_SCE_f(Y,X1,X2,n1,n2,b,d,h1,h2,MD,p,q)
% Sub-function of 'BS_Semi_LLR_SCE_iteration_f'.
n=n1*n2;
Yb=Y-mean(Y);
X1b=X1-mean(X1);
for i=1:n
  Y1(i,1)=CV_Semi_LLR_predict_f(Y,X1,Yb,X1b,X2,n,i,b,d,h1);
end
E1=Y-Y1;
switch MD
  case {1}
    [I_B,P,Q,r2]=Estimate_torus_IB(E1,n1,n2);
  case {2}
    [I_B,P,Q,r2]=Estimate_bilateral_IB(E1,n1,n2,p,q);
  case {3}
    [I_B,P,Q,r2]=Estimate_unilateral_IB(E1,n1,n2);
end
Z=Y1+I_B*E1;
Zb=Z-mean(Z);
for i=1:n
  Y2(i,1)=CV_Semi_LLR_predict_f(Z,X1,Zb,X1b,X2,n,i,b,d,h2);
end
E2=Y-Y2;
switch MD
  case {1}
    [I_B,P,Q,r2]=Estimate_torus_IB(E1,n1,n2);
  case {2}
    [I_B,P,Q,r2]=Estimate_bilateral_IB(E1,n1,n2,p,q);
  case {3}
    [I_B,P,Q,r2]=Estimate_unilateral_IB(E1,n1,n2);
end
temp=I_B'*I_B;
tD=0;tD(n,1)=0;
for i=1:n
  tD(i,1)=temp(i,i);
end
C_m=diag(ones(1,n))-temp/mean(tD);
E3=mean(E2)+C_m*(E2-mean(E2));
MSE=mean((Y-Y2-E3).^2);

function Yp=CV_Semi_LLR_predict_f(Y,X1,Yb,X1b,X2,n,i,b,d,h)
% Sub-function of 'CV_Semi_LLR_f' and 'DCV_Semi_LLR_SCE_f'.
U=0;V0=0;V=zeros(d+1,b);V1=0;
for j=1:n
  if j~=i
    K=1;
    for k=1:d
      K=K*((2*pi*h(k)^2)^(-1/2)*exp(-1/2*((X2(j,k)-
        X2(i,k))/h(k))^2));
```

```
      end
      Xh=[1  (X2(j,:)-X2(i,:))./h];
      U=U+K*Xh'*Xh;
      V0=V0+Yb(j)*Xh'*K;
      for k=1:b
        V(:,k)=V(:,k)+X1b(j,k)*Xh'*K;
      end
    end
end
D0=[1  zeros(1,d)]*U^(-1)*V0;
D=[1  zeros(1,d)]*U^(-1)*V;
Yc=Yb-D0;
Xc=X1b-D;
L=(Xc'*Xc)^(-1)*(Xc'*Yc);
Mu=mean(Y)-L'*mean(X1);
Yd=Yb-L'*X1b;
for j=1:n
  if j~=i
    K=1;
    for k=1:d
      K=K*((2*pi*h(k)^2)^(-1/2)*exp(-1/2*((X2(j,k)-
        X2(i,k))/h(k))^2));
    end
    Xh=[1  (X2(j,:)-X2(i,:))./h];
    V1=V1+Yd(j)*Xh'*K;
  end
end
Beta=[1  zeros(1,d)]*U^(-1)*V1;
Yp=Mu+L'*X1(i,:)+Beta;
```

References

Ali, M. M. (1979). Analysis of stationary spatial-temporal processes: estimation and prediction. *Biometrika*, **66**, 513-518.

Basawa, I. V. (1995). Inference for a class of causal spatial models. *Athens Conference on Applied Probability and Time Series Analysis*, Athens.

Basawa, I. V., Brockwell, P. J. & Mandrekar, V. (1992). Inference for spatial time series. *Proc. Interface-90*, New York.

Basu, S. & Reinsel, G. C. (1994). Regression models with spatially correlated errors. *J. Amer. Statist. Assoc.* **89**, 88-99.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *J. Roy. Statist. Soc. Ser. B.* **36**, 192-236.

Burrough, P. A. (1983). Problems of superimposed effects in the statistical study of the spatial variation of soil. *Agricultural Water Management*, **6**, 123-143.

Cai, Z. & Fan, J. (2000). Average regression surface for dependent data. *Journal of Multivariate Analysis*, **75**, 112-142.

Chen, H. (1988). Convergence rates for parametric components in a partly linear

model. *Ann. Statist.* **16**, 136-146.

Chen, R. & Härdle, W. (1995). Estimation and variable selection in additive nonparametric regression models. Discussion paper, SFB 373, Humboldt-Universität zu Berlin.

Chen, H. & Shiau, J. H. (1991). A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning & Inference*, **27**, 187-201.

Chen, H. & Shiau, J. H. (1994). Data-driven efficient estimators for a partially linear model. *Ann. Statist.* **22**, 211-237.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829-836.

Cressie, N. A. C. (1993). Statistics for spatial data. Wiley, New York.

Cressie, N. A. C. & Chan, N. H. (1989). Spatial modeling of regional variables. *J. Amer. Statist. Assoc.* **84**, 393-401.

Engle, R. F., Granger, W. J., Rice, J. & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **80**, 310-319.

Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87**, 998-1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Ann. Statist.* **21**, 196-216.

Fan, J. & Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20**, 2008-2036.

Fan, J. & Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B.* **57**, 371-394.

Fan, J. & Gijbels, I. (1996). Local polynomial modelling and its applications. Chapman & Hall, London.

Fan, Y. & Li, Q. (2003). A kernel-based method for estimating additive partially linear models. *Statistica Sinica*, **13**, 739-762.

Francisco-Fernandez, M. & Jean Opsomer, J. D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *The Canadian Journal of Statistics,* **33**, 279-295.

Francisco-Fernandez, M. & Vilar-Fernandez, J. M. (2001). Local polynomial regression estimation with correlated errors. *Communication in Statistics. A.* **30**, 1271-1293.

Gao, J., Lu, Z. & Tjøstheim, D. (2006). Estimation in semiparametric spatial regression. *Ann. Statist.* **34**, 1395-1435.

Hallin, M., Lu, Z. & Tran, L. T. (2001). Density estimation for spatial linear processes. *Bernoulli* **7**, 657-668.

Hallin, M., Lu, Z. & Tran, L. T. (2004). Local linear spatial regression. *Ann. Statist.* **32**, 2469-2500.

Hamilton, S. A. (1997). Local linear estimation in partly linear models. *Journal of Multivariate Analysis,* **60**, 1-19.

Heckman, N. E. (1986). Spline smoothing in partly linear models. *J. Roy. Statist. Soc. Ser. B.* **48**, 244-248.

Kashyap, R. L. (1979). Univariate and multivariate random field models for images. *Proc. Workshop on image modelling,* Chicago, IL.

Kashyap, R. L. (1980). Random field models on torus lattaces for finite images. Proc. *5th International Conference on Pattern Recognition,* Miami Beach, FL.

Kashyap, R. L. & Chellappa, R. (1983). Estimation and choice of neighbors in spatial-interaction models of images. *IEEE Transactions on Information Theory,* **29**, 60-72.

Linton, O. & Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika,* **82**, 93-100.

Liu, X. (2001). Kernel smoothing for spatially correlated data. Ph. D. thesis, Department of Statistics, Iowa State University.

Lu, Z., Lundervold, A., Tjøstheim, D. & Yao, Q. (2007). Exploring spatial nonlinearity using additive approximation. *Bernoulli*, **13**, 447-472.

Lu, Z. & Gijbels, I. (2001). Asymptotics for partly linear regression with dependent samples and ARCH errors: consistency with rates. *Science in China*, **44**, 168-183.

Mammen, E., Linton, O. & Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**,

1443-1490.

Martin, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika,* **66**, 209-217.

Martin, R. J. (1982). Some aspects of experimental design and analysis when errors are correlated. *Biometrika,* **69**, 597-612.

Martins-Filho, C. & Yao, F. (2009). Nonparametric regression estimation with general parametric error covariance. *Journal of Multivariate Analysis,* **100**, 309-333.

Masry, E. & Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *The Scandinavian Journal of Statistics,* **24**, 165-179.

Matheron, G. (1962). Traite de Geostatistique Appliquee, Tome I. *Memoires du Bureau de Recherches Geologique et Minieres*, **14**. Editions Technip, Paris.

Matheron, G. (1963a). Traite de Geostatistique Appliquee, Tome II. Le Krigeage. *Memoires du Bureau de Recherches Geologiques et Minieres*, **24**. Editions Bureau de Recherche Geologiques et Minieres, Paris.

Matheron, G. (1963b). Principles of geostatistics. *Economic Geology*, **58**, 1246-1266.

Müller, H. G. (1988). Nonparametric regression analysis of longitudinal data. Springer, Berlin.

Opsomer, J. D. & Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25**, 186-211.

Ord, J. K. (1975). Estimation methods for models of spatial interaction. *J. Amer. Statist. Assoc.* **70**, 120-126.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics & Probability Letters*, **4**, 203-208.

Robinson, P. M. (1988). Root-*n*-consistent semiparametric regression. *Econometrica*, **56**, 931-954.

Robinson, P. M. & Vidal Sanz, J. (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis,* **97**, 1090-1120.

Ruppert, D. & Wand, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.* **22**, 1346-1370.

Russo, D. & Bresler, E. (1981). Soil hydraulic properties as stochastic processes: I. An analysis of field spatial variability. *Soil Science Society of America Journal*, **45**, 682-687.

Schabenberger, O. & Gotway, C. A. (2005). Statistical methods for spatial data analysis. Chapman & Hall/CRC.

Schick, A. (1994). Estimation of the autocorrelation coefficient in the presence of a regression trend. *Statistics & Probability Letters*, **21**, 371-380.

Schick, A. (1996). Efficient estimation in a semiparametric additive regression model with autoregressive errors. *Stochastic Processes and their Applications*, **61**, 339-361.

Schick, A. (1998). An adaptive estimator of the autocorrelation coefficient in regression models with autoregressive errors. *Journal of Time Series Analysis*, **19**, 575-589.

Severance-Lossin, E. & Sperlich, S. (1999). Estimation of derivatives for additive separable models. Technical report.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B.* **50**, 413-436.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5**, 595-645.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348-1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 685-705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 592-606.

Sun, X., You, J., Chen, G. & Zhou, X. (2002). Convergence rates of estimators in partial linear regression models with MA(∞) error process. *Communication in Statistics. A.* **31**, 2251-2273.

Tjøstheim, D. (1978). Statistical spatial series modelling. *Adv. Appl. Prob.* **10**, 130-154.

Tjøstheim, D. (1983). Statistical spatial series modelling II: some further results on unilateral lattice processes. *Adv. Appl. Prob.* **15**, 562-584.

Tsybakov, A. B. (1986). Robust reconstruction of functions by the local-approximation method. *Problems of Information Transmission*, **1986**, 133-146.

Vilar-Fernandez, J. M. & Francisco-Fernandez, M. (2002). Local polynomial regression smoothers with AR-error structure. *TEST,* **11**, 439-464.

Watson, G. S. (1972). Trend surface analysis and spatial correlation. *Geological Society of America, Special Paper*, **146**, 39-46.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika,* **41**, 434-449.