NONPARAMETRIC ESTIMATION OF CONDITIONAL VARIANCE AND COVARIANCE FUNCTIONS

JIANG HUI

(B.Sc., WUHAN UNIVERSITY)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2016

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Jiang Hui

19th Jan 2016

Thesis Supervisor

Xia Yingcun Professor; Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546, Singapore.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Professor Xia Yingcun, for his patient guidance, continuous support and encouragement. I am really honored to have him as my supervisor. He always impresses me with his wealth of knowledge during our discussions. I truly appreciate all his contributions of time, enlightening ideas and invaluable advices. What I learned from him will benefit me for my whole life.

Meanwhile, I would like to give my sincere thanks to Dr. Patrick W. Saart for his kindly help and constructive suggestions for the second chapter. I am also thankful to other faculty members and support staffs of the Department of Statistics and Applied Probability for their effort and support during my study in NUS.

I also feel grateful to my dear friends for their assistance, accompany and encouragement. Last but not least, I would like to take this opportunity to thank my family. Special appreciations are given to my parents for their unconditional love, considerable understanding and continuous support.

Contents

Summary

In recent times, in order to study the relationship between large numbers of variables based on some given information, estimation of the conditional covariance matrix has received much attention in many areas. Even though several statistical models and methods have been introduced in literature, those models still have limited capability to describe different patterns of dependence in the data. In this thesis, we study the estimation problem of conditional covariance matrix from two aspects. First, to study the correlation structure for a portfolio of financial assets, we explore the effect of the exogenous variable on pairwise correlations by utilizing a reduced rank model. Therefore, we could identify the functional driving factors based on smoothing techniques and tools in factor analysis, but without additional model specification. Simulation studies and an empirical analysis are conducted to demonstrate the validity of our approach. The second problem considered is how to efficiently estimate conditional variance functions. Instead of estimating the mean functions at the first stage, we propose a novel approach by combining the techniques in kernel smoothing and difference-based method, which outperforms two existing approaches in most cases. Furthermore, we provide detailed theoretical justifications in Chapter 2 and Chapter 3 respectively, including consistency and asymptotic normality of our proposed estimators.

List of Tables

List of Figures

CHAPTER 1

Introduction

Modelling variances and covariance matrices are common statistical problems in different fields, such as graphical modelling, machine learning and financial econometrics. Various popular approaches have been well understood and applied for estimating variances. Additionally, for estimating covariance matrices, substantive available methods have been proposed to illustrate the structure of these matrices. Recently, due to the nature of some collected data sets, describing and forecasting dynamic variances and covariance matrices has attracted considerable attention. In particular, nonparametric and semiparametric models have been utilized and extensively studied for estimating conditional variance functions. Although there are not many nonparametric or semiparametric models for conditional covariance matrices, several useful estimation strategies have been developed and widely applied in practice. In the following, we will discuss some existing approaches to these estimation problems.

1.1 Conditional variance estimation

Many scientific studies require information about the volatility of a random variable (or covolatility of two random variables) given past information or the values of other variables, which is frequently measured by conditional variance (or conditional covariance) in statistical analysis. Estimation of conditional variance function or covariance function is important in a variety of statistical applications. Recently, nonparametric and semiparametric regression models have also been employed for this estimation problem.

Variance function estimation via nonparametric heteroskedastic regression models is an active area in statistical analysis. Let $Y \in \mathbb{R}$ be a scalar response, $\mathrm{X} \in \mathbb{R}^p$ be p-dimensional covariate. The following nonparametric model is usually considered in literature,

$$Y = m(\mathrm{X}) + \sigma(\mathrm{X})\varepsilon, \tag{1.1}$$

where $m(x)$ is an unknown mean function, $\sigma^2(x)$ is an unknown variance function, and ε is the error term satisfying $E(\varepsilon|\mathrm{X}) = 0$, $\mathbb{V}\mathrm{ar}(\varepsilon^2|\mathrm{X}) = 1$. Given samples $\{(\mathrm{X}_i, Y_i) : i = 1, \cdots, n\}$, we are particularly interested in the conditional variance of Y given $\mathrm{X} = \mathrm{x}$, denoted by

$$\sigma^2(\mathrm{x}) = \mathbb{V}\mathrm{ar}(Y|\mathrm{X} = \mathrm{x}) = E[\{Y - m(\mathrm{X})\}^2|\mathrm{X} = \mathrm{x}]$$

where $m(\mathrm{x}) = E(Y|\mathrm{X} = \mathrm{x})$. This problem has received much attention for the univariate case $p = 1$. In this situation, to estimate $\sigma^2(\mathrm{x})$, a natural idea is to firstly estimate the unknown conditional mean function $m(\mathrm{x})$. Yao and Tong (1994) proposed a direct estimator by separately estimating $E(Y^2|\mathrm{X})$ and $E(Y|\mathrm{X})$, and then obtained the variance estimator, but such an estimator could introduce a large bias. Even though Härdle and Tsybakov (1997) improved the estimator by using common bandwidth and kernel, this method seems still not completely adaptive to the unknown conditional mean function $m(\mathrm{x})$. Based on the squared residuals obtained after a preliminary estimation of $m(\mathrm{x})$, Hall and Carroll (1989) considered kernel estimators of $\sigma^2(\mathrm{x})$ and investigated the effect of the smoothness of $m(\mathrm{x})$ on the convergence rate for variance function estimator. Additionally, Ruppert, Wand, Holst and Hössjer (1997) and Fan and Yao (1998) investigated this problem by means of local polynomial smoothing on the estimated squared residuals, and proved that such residual-based estimator could be adaptive to the unknown $m(\mathrm{x})$ under some regularity conditions. However, when $m(\mathrm{x})$ is not smooth enough or has high fluctuation, the estimation of $m(\mathrm{x})$ is not so efficient, and neither the estimation of $\sigma^2(\mathrm{x})$. See the discussion of Wang, Brown, Cai and Levine (2008) and the examples therein.

There is much literature also dealing with this estimation problem when $p > 1$. However, direct extension of such nonparametric estimation procedures to the multivariate case may not be feasible, because it may encounter the well-known "curse of dimensionality". Nevertheless, several effective estimators may be constructed by restricting the functional form of the unknown mean and variance functions. Some useful classes of models have been successfully applied in high dimensional cases, including, but not limited to, additive models and single-index models. In addition, without estimating mean functions or computing link functions, difference-based approaches have been studied for variance function estimation problem. The details of difference-based methods will be further discussed in Chapter 3.

1.2 Conditional covariance matrix estimation

To our knowledge, estimating the covariance matrix (or its inverse) could be difficult due to two main obstacles: (a) the positive-definiteness constraint and (b) the high dimensionality problem. For modelling the constant covariance matrices, a number of alternatives have been extensively studied in literature (Pourahmadi (1999), Bickel and Levina (2008a), Bickel and Levina (2008b), Levina, Rothman and Zhu (2008), Lam and Fan (2009) and Cai and Liu (2011)), and most of them focus on the sparse estimation procedures to achieve parsimonious structure. In real situations, the constant assumption imposed on the covariance matrix may be violated, thus developing desirable covariance matrix estimators in this case could be much more difficult in terms of efficiency, generality and computational cost.

The estimation problems about time varying conditional covariance matrix are most commonly discussed. Consider a multivariate process $\{\mathrm{X}_t\}$, $\mathrm{X}_t = (X_{1,t}, \cdots, X_{p,t})^\top$ with mean zero, and allow the information set at time $t-1$ represented by $\mathcal{F}_{t-1}$, then the conditional covariance matrix assumed to follow a time-varying structure is defined as

$$\mathbb{C}\mathrm{ov}(\mathrm{X}_t|\mathcal{F}_{t-1}) = \mathbf{H}_t. \tag{1.2}$$

Numerous statistical models for conditional covariance matrix $\mathbf{H}_t$ have been developed in the literature.

1.2.1 Estimation through multivariate GARCH models

Bollerslev, Engle and Wooldridge (1988) introduced the VEC model, a direct generalization of the univariate GARCH model by using the vectorization operation. For instance, the VEC$(1,1)$ model is given by

$$\mathrm{vec}(\mathbf{H}_t) = \mathrm{vec}(\mathbf{\Omega}) + \mathbf{A}\mathrm{vec}(\mathrm{X}_{t-1}\mathrm{X}_{t-1}^\top) + \mathbf{B}\mathrm{vec}(\mathbf{H}_{t-1}), \tag{1.3}$$

where A and B are $p^2 \times p^2$ matrices, and $\mathrm{vec}(\cdot)$ denotes the vectorization of $p \times p$ matrix, but some strong constraints should be imposed on the parameters to ensure the positive definiteness of the conditional covariance matrix. In order that the positivity of $\mathbf{H}_t$ could be easily guaranteed under weak conditions, Engle and Kroner (1995) presented the BEKK representation for the dynamic covariance matrix. Accordingly, the BEKK$(1,1,m)$ model is expressed as

$$\mathbf{H}_t = \mathbf{C} + \sum_{j=1}^{m} \mathbf{A}_j \mathrm{X}_{t-1}\mathrm{X}_{t-1}^\top \mathbf{A}_j^\top + \sum_{j=1}^{m} \mathbf{B}_j \mathbf{H}_{t-1} \mathbf{B}_j^\top, \tag{1.4}$$

where $\mathbf{C}$, $\mathbf{A}_j$, and $\mathbf{B}_j$ are $p \times p$ matrices, and $\mathbf{C}$ is positive definite. However, estimating through these models is intractable due to the so-called curse of dimensionality. It is thus not surprising that other less parameterized model structure should be suggested in order to circumvent this difficulty.

Conditional Correlation Models

One strand of literature focused on modelling of the conditional covariance matrix $\mathbf{H}_t$ through the following formula,

$$\mathbf{H}_t = \mathbf{D}_t^{\frac{1}{2}} \mathbf{R}_t \mathbf{D}_t^{\frac{1}{2}}, \tag{1.5}$$

where $\mathbf{R}_t$ represents the conditional correlation matrix, and $\mathbf{D}_t = \text{diag}(h_{1,t}, \cdots, h_{p,t})$ is a diagonal matrix with conditional variances as diagonal elements. Therefore, most researchers have focused on the estimation of conditional correlations due to the decomposition (1.5). Conventionally, researchers have modelled correlation as an unchanged and unconditional variable. After many years, with some empirical evidence provided to demonstrate an opposing view, they have gradually realized that correlation indeed varies through time. The awareness of the time-variability of correlation has propelled a continually growing amount of work on developing various conditional correlation models. Because of the empirical evidence on the autocorrelation structure of correlations, researchers have devoted their effort to explore whether existing conditional variance methods on the basis of past information, i.e. the so-called GARCH models, could be generalized with the aim of modeling conditional correlation. An example of the earlier generation models of this sort is the Constant Conditional Correlation GARCH model of Bollerslev (1990), in which the conditional correlation matrix $\mathbf{R}_t$ in (1.5) is replaced by a constant matrix $\mathbf{R}$. In addition, there were two other alternative dynamic conditional correlation GARCH models, which were discussed in Tse and Tsui (2002) and Engle (2002), namely the VC-GARCH and the DCC-GARCH models, respectively. More explicitly, the conditional correlation matrix $\mathbf{R}_t$ in VC-GARCH model is defined as

$$\mathbf{R}_t = (1 - \theta_1 - \theta_2)\mathbf{R} + \theta_1 \mathbf{\Psi}_{t-1} + \theta_2 \mathbf{R}_{t-1}, \tag{1.6}$$

where $\mathbf{R}$ is a symmetric $p \times p$ positive definite parameter matrix with all diagonal elements being 1, θ_1 and θ_2 are non-negative parameters satisfying $\theta_1 + \theta_2 < 1$, and $\mathbf{\Psi}_{t-1}$ is the $p \times p$ correlation matrix of X_{t-1} for the preceding M $\mathrm{X}_{t-1}, \cdots, \mathrm{X}_{t-M}$. Analogous to VC-GARCH model, Engle (2002) studied the following dynamic matrix process $\mathbf{Q}_t = (q_{kl,t})$ given by

$$\mathbf{Q}_t = (1 - \theta_1 - \theta_2)\mathbf{S} + \theta_1 \mathrm{X}_{t-1}\mathrm{X}_{t-1}^\top + \theta_2 \mathbf{Q}_{t-1},$$

where $\mathbf{S}$ is the $p \times p$ unconditional correlation matrix of X_t, and θ_1, θ_2 are non-negative scalar parameters satisfying $\theta_1 + \theta_2 < 1$. Then the correlation matrix $\mathbf{R}_t$ in the DCC-GARCH model is specified as

$$\mathbf{R}_t = \mathrm{diag}(q_{11,t}^{-\frac{1}{2}}, \cdots, q_{pp,t}^{-\frac{1}{2}})\ \mathbf{Q}_t\ \mathrm{diag}(q_{11,t}^{-\frac{1}{2}}, \cdots, q_{pp,t}^{-\frac{1}{2}}). \tag{1.7}$$

Accordingly, expression (1.5) and the above models indicate that the conditional correlations play a significant role in estimating conditional volatilities and covariances. The conditional correlations of model (1.6) and (1.7) are assumed to obey the same dynamics, but these conditions do not hold in general. Additionally, those assumptions on the correlation matrix simplifies the estimation procedure and largely reduces the number of parameters, but they have limited capability to describe heteroskedasticity for more general cases.

Factor Models

To explore the structure of the covariance matrix, another commonly used approach is relevant to the factor models, which could provide a parsimonious representation and allow the semi-positive definiteness maintained. In the beginning, we could briefly introduce the definition of the factor model. Consider random variables $X_1, \cdots, X_p$, and the model assumes that they could be expressed as

$$X_k = a_{k1}F_1 + \cdots + a_{kq}F_q + \eta_k, \quad k = 1, \cdots, p, \tag{1.8}$$

where $q \ll p$, $F_j, j = 1, ..., q$ are called factors, and $\eta_k, k = 1, ..., p$ are random disturbances with $E(\eta_k \eta_l) = 0$ for $k \neq l$ and uncorrelated with F_j's, $E(\eta_k) = 0$, $E(\eta_k F_j) = 0$. Note that the factors $F_1, \cdots, F_q$ are the same for each X_k, and factors themselves could be correlated.

For those approaches designed for estimating dynamic covariance matrices based on factor models, they assume that the co-movements are mainly driven by a few underlying variables. The first category consists of factor models with factors following GARCH-type processes. Engle, Ng and Rothschild (1990) recommended the K-factor GARCH model, which allows the factors following GARCH processes and the time-varying part of conditional covariance matrix having reduced rank K, but leaves the constant part unrestricted. Vrontos et al. (2003) presented a full-factor GARCH model by assuming a triangular structure of the parameter matrix and obtained conditionally uncorrelated factors. Additionally, some orthogonal models have been described in the literature. In the orthogonal GARCH (O-GARCH) model of Alexander and Chibumba (1997), unconditionally uncorrelated principal components were obtained by utilizing the orthogonal matrix of eigenvectors derived from the sample covariance matrix as the linear transformation matrix. The orthogonality condition was relaxed to square and invertible in the generalized orthogonal GARCH (GO-GARCH) model of van der Weide (2002). Unfortunately, some restrictions are imposed on the dynamic specifications of these models. Recently, Fan, Wang and Yao (2008) proposed an alternative for modelling multivariate volatilities by use of conditionally uncorrelated components (CUCs), which are similar to the independent components in ICA. The second category consists of latent factor models, in which the common factors could not be expressed as functions of past data. Diebold and Nerlove (1989) reported that the commonalities in the conditional variance movements are well captured by latent factor ARCH models. Usually, those factors are supposed to be conditionally uncorrelated. Recently, Sentana and Fiorentini (2001), Doz and Renault (2004), and Fiorentini, Sentana and Shephard (2004) discussed identification, representation and estimation of those models, as well as inference methods.

For the purposes of improving the performance of estimation algorithms, researchers have made some efforts to alleviate computational burden. For example, van der Weide (2002) considered a multi-step approach by identifying a portion of the invertible link matrix by means of principal component analysis (PCA) at first, and estimating the second part of the invertible matrix as well as the remaining parameters in a second step. For the estimation of the CUC model proposed by Fan, Wang and Yao (2008), a high dimensional optimization problem has been converted to a series of simpler subproblems.

1.2.2 Nonparametric estimation methods

Note that the aforementioned methods are based on parametric model representations and mainly focused on dynamic covariance matrix dependent on past information. Recently, nonparametric models and methods have been employed to address the estimation problem of conditional covariance matrix. In this category, there are several approaches available for estimating dynamic covariance matrices. For example, Wu and Pourahmadi (2003) considered large covariance matrix estimation problem in longitudinal data analysis and designed nonparametric estimators of covariance matrices by adopting the two-step estimation strategy presented in Fan and Zhang (2000). By regressing each variable on its predecessors, the resulting covariance matrix estimator could be guaranteed to be positive definite. Recently, Yin et al. (2010) proposed a consistent kernel estimator for the conditional covariance matrix. In order to achieve a sparse structure of conditional variance matrices, Chen and Leng (2015) made use of entrywise thresholding based on the preliminary nonparametric covariance matrix estimator.

1.3 Research objectives and organizations

From the above summary, it could be observed that some research gaps still exist.

- Most existing approaches require pre-specified structure of the conditional covariance matrix (such as the multivariate GARCH models), which may not describe covariance matrices of various types. Besides, even though several nonparametric models and methods for modelling conditional covariance matrix have been presented, they have limited capability to examine how an exogenous variable could affect the behaviour of elements of covariance matrices, and lack the ability to identify the main drivers.

- For the estimation problem of conditional variance function, estimating unknown conditional mean is necessary for most existing nonparametric variance estimators. However, as mentioned before, those estimators seem to be undesirable when the mean function is heavily oscillating. In addition, even though difference-based methods could evade this issue, they demand a complex construction of the difference sequences especially for multivariate cases.

Therefore, we aim to make some contributions to fill these gaps in the following chapters. In Chapter 2, we study the estimation of conditional covariance matrix, mainly focusing on the pairwise conditional correlations. Specifically, we introduce a reduced rank model for pairwise correlation coefficients conditional on an exogenous variable to extract the features of dependence. Furthermore, our proposed model is applied to investigate the well-known asymmetric effect of market return on stock returns correlations. In Chapter 3, we propose a novel approach to model conditional variance (covariance) function, by combining the strength of kernel smoothing and difference-based methods. Numerical studies based on comparisons between our estimator and its competitors are conducted. The results suggest that our estimation strategy could outperform other two estimators in most situations. Chapter 4 contains a overall conclusion of this thesis and discussions on open questions of future research.

CHAPTER 2

A Reduced Rank Model for Conditional Correlations

2.1 Introduction

In the financial world, often financial market participants must manage a large number of financial assets simultaneously. The obvious examples are equity investors who often face risks which could affect assets in portfolio in various ways and must therefore find a solution to hedge against these risks. In practice, this may be achieved by means of diversification across several stock markets and/or asset classes, for instance. However, constructing an efficient portfolio to benefit the most from diversification is not a straightforward matter since it requires knowledge about comovements and associations, i.e. correlations, of the assets in question. In addition, such knowledge about the correlations is required in a wide range of financial applications, especially asset capital allocation, risk management, pricing models and option pricing among others. We will develop in this chapter an alternative method that is capable of explaining what drives correlations between financial assets and how. The new method, which we will refer to hereafter as the *reduced rank model for conditional correlation coefficients*, is designed for studying pairwise conditional correlation structure of financial returns in a functional context of *a semiparametric factor model*.

In chapter 1, several popular statistical models of GARCH type about conditional correlations have been reviewed. A well-known example is the Constant Conditional Correlation GARCH model of Bollerslev (1990). In addition, there are other alternative dynamic conditional correlation GARCH models, which have been discussed, for example, in Tse and Tsui (2002), Engle (2002) and Aielli (2013) namely the VC-GARCH, DCC-GARCH and the cDCC models, respectively. Although introduced with some general specifications and do not suffer from the curse of dimensionality problem, these models have quite limited capability. In particular, these models are not able to explain the roles market variables, such as return or volatility, play in driving changes in the behavior of correlations between stock returns, which are of particular interest to financial analysts (see, for example, Ang and Chen (2002) and Amira, Taamouti and Tsafack (2011)).

As an alternative, Pelletier (2006) proposed a model with a regime-switching correlation framework, which presented that the correlations remain constant in each regime while the change between the states was controlled by transition probabilities. Silvennoinen and Teräsvirta (2015) introduced an alternative model which they referred to as the Smooth Transition Conditional Correlation GARCH (STCC-GARCH) model. The STCC-GARCH model enables the conditional correlations to change between two states smoothly as a function of a transition variable. Hence, these models are associated to some extent with a pre-specified model structure of the covariance (e.g. the GARCH-type evolution or regime-switching GARCH model, etc). This leads to an important limitation which resides in the fact that the number of parameters required explodes with the dimension of the model (see e.g. Kring et al. (2007) and Santos and Moura (2014)).

Since the ability to model co-movements for portfolios with a large number of assets and the changes in their behavior are essential in many areas of financial management, existence of the above-mentioned drawbacks suggests that directly modeling the assets by a multivariate GARCH model might not be feasible. Instead, an asset manager must consider some form of factor-model techniques so as to reduce the overall dimension of the time series modeling problem. The use of factors to reduce the dimensionality of multivariate GARCH models was proposed in a seminal paper by Engle, Ng and Rothschild (1990), and further developed by, among the others, Vrontos et al. (2003) and Lanne and Saikkonen (2007). More recently Sheppard and Xu (2014) introduced the so-called Factor-HEAVY (F-HEAVY) model utilizing high frequency data, which has a deep root into the GARCH modeling of conditional volatility. Nonetheless, the purpose of most existing factor-based models, including the F-HEAVY, is to study the way in which covariance matrix changes, while these changes are driven by the past information generated by the time series themselves. As the results, the focus of the studies in multivariate factor GARCH is on predictive models, rather than on nonparametric measurement of past volatility and correlations. On the contrary, the semiparametric factor model introduced in this chapter enables examination of what exogenous forces and how they drive the changes in the correlations of returns. We focus on exploring the asymmetric effect of the exogenous variable on pairwise correlations and identifying the main drivers of the asymmetry in pairwise correlations in a similar spirit to Ang and Chen (2002) and Amira, Taamouti and Tsafack (2011). The importance of the factor-approach is to summarize the common patterns in the pairwise correlations. It will soon be clear that the method developed in this paper sits well within the well-known functional data analysis framework and hence inherits the ability to deal with high-dimensional time series problems. Furthermore, it is based on nonprametric smoothing and thus model free, which makes it less likely to suffer modeling mis-specification compared to the existing methods.

The new technique begins with the empirical estimation of the pairwise correlation coefficients of the returns conditional on a particular variable that is of empirical interest, the selection of which is determined by the research problem under consideration. For the sake of clarity, one can think the above conditional variable as playing a similar role in our model to the transition variable in the STCC-GARCH model of Silvennoinen and Teräsvirta (2015). Since the (pairwise) conditional correlation coefficients are derived based on unknown conditional mean and conditional variance, their estimators must be constructed using empirical estimates. Under the assumption that the conditional correlation coefficient functions share a finite number of common factors, we explore a method of common functional factor analysis along the line of the existing techniques of principal component analysis. To this end, we establish estimators of both the orthogonal functional factors and the corresponding loading coefficients. The theoretical analysis in this chapter concentrates on the derivation of consistency and the asymptotic distribution of these estimators that are needed in order to perform statistical inference in the analysis.

Moreover, the empirical investigation of this chapter focuses on estimation and analysis of the conditional correlation coefficients for returns of a portfolio of assets, which consists of thirty major American companies included in the Dow Jones Industrial Average, i.e. Dow30 hereafter. A full list of companies included in the Dow30 can be found in various websites, for example `www.money.cnn.com`. From the empirical point of view, the questions of how and what drives the observed time-varying correlation structure in financial markets relate directly to the selection of the conditional variable used in the estimation of the newly developed reduced rank model. In the current chapter, we examine suggestions from two popular school of thoughts, which focuses on market volatility and market return as the driving factors, respectively. Interestingly, we are able to establish the empirical evidence in support of the well-known asymmetric-effect of market return on the conditional correlations of the stock returns only when the possible leverage-effect on the market has been taken into consideration. The volatility effect of market return seems to lead to high correlations of the stock returns during the bull market, so that the asymmetric-effect of market return is not evidenced. Nonetheless, once the leverage-effect in the market is disentangled and the volatility effect is removed, correlations of the stock returns drop significantly in the bull market. In turn, this leads to the apparent asymmetric-effect of the market return.

This chapter is organized as follows. Section 2.2 discusses the basic construction of our new method, including model assumptions, identification and estimation procedures. Section 2.3 presents the main asymptotic results of the chapter, which focus on the consistency and asymptotic distribution of all the nonparametric estimators involved. These results are convincingly demonstrated by Monte Carlo simulations in Section 2.4. We then perform empirical analysis in Section 2.5, while all technical proofs are given in Section 2.7.

2.2 Conditional correlation coefficients

Note that, in the current section and the next, the conditioning variable, denoted by U, plays a similar role in our model to the so-called transition variable in the STCC-GARCH model of Silvennoinen and Teräsvirta (2015). In practice, the choice of U can be selected in accordance to the empirical question under investigation. However, since the purpose here is to introduce the model in the general context, we will illustrate and discuss this process in more specific details in Section 2.5. In this section, we first present the basic construction of our new method, reduced rank model for conditional correlation coefficients, which includes model assumption and identification. Then, we discuss the model's practical operation, which covers the estimation procedures and suggested methods of selecting the number of common factors.

2.2.1 Definitions

In the current chapter, we first focus on the study of pairwise conditional correlations. Suppose r_1 and r_2 are returns of two stocks with $E(r_1) = E(r_2) = 0$, so that the unconditional correlation coefficient is defined as

$$\rho_{1,2} = \frac{E(r_1 r_2)}{\sqrt{Er_1^2 Er_2^2}}, \tag{2.1}$$

where $E(r_1 r_2)$ measures the co-movement of r_1 and r_2. We have by conditioning upon U,

$$E(r_1 r_2 | U) = \mu_1(U)\mu_2(U) + E\{(r_1 - \mu_1(U))(r_2 - \mu_2(U))|U\}, \tag{2.2}$$

where $\mu_k(U) = E(r_k|U), k = 1, 2$. In other words, the co-movement between r_1 and r_2 is determined by U based on (i) the effect on the means of r_1 and r_2, and (ii) the effect through the conditional covariance after the effect due to the conditional mean is removed.

Expression (2.2) suggests that we need to consider these two effects separately. After standardization, we may define the correlation due to the effect passing through the conditional means as

$$\phi_{1,2}(U) = \frac{E(r_1|U)E(r_2|U)}{\sqrt{E(r_1^2|U)E(r_2^2|U)}}, \tag{2.3}$$

where $|\phi_{1,2}(U)| \leq 1$ due to the Cauchy-Schwartz inequality. The quantity in (2.3) measures the co-movement in the conditional mean and hence it is referred to as the "conditional mean correlation". Similarly, we may define the correlation due to the effect passing through the conditional covariance as

$$\varrho_{1,2}(U) = \frac{E\{(r_1 - \mu_1(U))(r_2 - \mu_2(U))|U\}}{\sqrt{E((r_1 - \mu_1(U))^2|U)E((r_2 - \mu_2(U))^2|U)}}. \tag{2.4}$$

In (2.4), $\varrho_{1,2}(U)$ is the effect of U on the cross correlation between $r_1 - \mu_1(U)$ and $r_2 - \mu_2(U)$ with the effect on the mean being removed and is therefore referred to as the "conditional correlation coefficient".

Ang and Chen (2002) introduced a measure of conditional correlation, which was defined as $Corr(r_1, r_2|c_1 \leq U \leq c_2)$. However, this definition can cause confusion. In this chapter, we discuss the conditional correlation by considering $c_1 \to c_2$, i.e. $Corr(r_1, r_2|U)$. As an example, we consider the capital asset pricing model (CAPM) in financial analysis, which states that

$$r_k = \alpha_k + \beta_k U + e_k, \quad k = 1, ..., m, \tag{2.5}$$

where U is the market return with $\text{Var}(U) = \sigma_U^2$, and

$$E(e_k|U) = 0, \quad \text{Cov}(e_k, e_\ell|U) = \begin{cases} \sigma_k^2, & \text{if } \ell = k, \\ 0, & \text{otherwise.} \end{cases}$$

We have the unconditional correlation

$$\rho_{k,\ell} = \frac{\beta_k \beta_\ell \sigma_U^2}{(\beta_k^2 \sigma_U^2 + \sigma_k^2)^{1/2} (\beta_\ell^2 \sigma_U^2 + \sigma_\ell^2)^{1/2}},$$

but the conditional correlation

$$\varrho_{k,\ell}(U) = 0.$$

However, if the noises share some common innovations, for example, if

$$e_k = \rho_{k1}(U)\epsilon_1 + \rho_{k2}(U)\epsilon_2, \quad k = 1, ..., m$$

then

$$\varrho_{k,\ell}(U) = \frac{\rho_{k1}(U)\rho_{\ell 1}(U) + \rho_{k2}(U)\rho_{\ell 2}(U)}{\{\rho_{k1}^2(U) + \rho_{k2}^2(U)\}^{1/2}\{\rho_{\ell 1}^2(U) + \rho_{\ell 2}^2(U)\}^{1/2}}.$$

It is thus important to note that the conditional correlation coefficient defined above is not caused by the common factors in the conditional mean.

2.2.2 Model assumption and identification

Suppose there are m assets to be considered and the return of the k-th asset is written as

$$r_k = \mu_k(U) + \sigma_k(U)\varepsilon_k, \quad k = 1, ..., m, \tag{2.6}$$

where $E(\varepsilon_k^2|U) \equiv 1$ almost surely. When U is selected as the market return, it is not difficult to see that the CAPM model described in (2.5) can be taken as a special case. When dealing with sample correlations, it should be taken into account that the return of a given stock should be standardized before being used for estimation of the correlation. Hence, it is useful for the estimation purpose to consider the model

$$(r_k - \mu_k(U))^2 = \sigma_k^2(U) + \sigma_k^2(U)\xi_k, \quad k = 1, ..., m,$$

where $\xi_{k,t} = \varepsilon_{k,t}^2 - 1$, as done in Fan and Yao (1998), for example.

For the co-movement of $\varepsilon_k, k = 1, ..., m$, we assume that the conditional correlation coefficient functions share $p \leq m$ common functional factors based on

$$E(\varepsilon_k \varepsilon_\ell | U) \equiv \varrho_{k,\ell}(U) = a_{k\ell} + G_{k\ell}(U) = a_{k\ell} + b_{k\ell}^{[1]} F_1(U) + ... + b_{k\ell}^{[p]} F_p(U), \tag{2.7}$$

where as usual it is assumed that

$$E\{F_j(U)\} = 0, \quad E\{F_{j_1}(U) F_{j_2}(U)\} = 0, \quad j, j_1, j_2 = 1, ..., p, \; j_1 \neq j_2, \tag{2.8}$$

and

$$Var(F_1) \geq ... \geq Var(F_p)$$

for identification purpose. In our analysis, we incorporate uncorrelated measurement errors to reflect additive measurement errors, so that the model we consider is

$$\varepsilon_k \varepsilon_\ell \;=\; \varrho_{k,\ell}(U) + \epsilon_{k,\ell} = a_{k\ell} + b_{k\ell}^{[1]} F_1(U) + ... + b_{k\ell}^{[p]} F_p(U) + \epsilon_{k,\ell}, \tag{2.9}$$

where $\epsilon_{k,\ell}$ are conditionally uncorrelated with each other for all $1 \leq k < \ell \leq m$, i.e.

$$E\{\epsilon_{k_1,\ell_1} \epsilon_{k_2,\ell_2} | U\} = 0, \quad \text{if} \;\; \{k_1, \ell_1\} \neq \{k_2, \ell_2\}.$$

With observations at $\{(r_{k,t}, U_t) : t = 1, ...n, k = 1, ..., m\}$, where t and k denote respectively the t-th time point and the k-th asset, our model of interest is thus (2.6) with

$$\varepsilon_{k,t} \varepsilon_{\ell,t} \;=\; a_{k\ell} + b_{k\ell}^{[1]} F_1(U_t) + ... + b_{k\ell}^{[p]} F_p(U_t) + \epsilon_{k,\ell,t}, \tag{2.10}$$

which we will refer to hereafter as the "reduced rank model".

The reduced rank model differs from most existing models since here the common functional factors $F_1(U), ..., F_p(U)$ and corresponding coefficients are all unobservable. A similar model was considered in studies on semi-parametric comparison of regression curves. A few well-known examples are Härdle and Marron (1990) and Munk and Dette (1998), who studied the comparison of two functions, and James, Hastie and Sugar (2000), who used a similar model but under a random effect setting. In addition, the semiparametric panel data model was also investigated by Boneva, Linton and Vogt (2015). They examined the common component structure of nonparametric functions, however, their dependent variables are observable. Under our model framework, $\varepsilon_{k,t}, k = 1, ..., m$, are latent variables and are designed to be estimated nonparametrically based on a GARCH framework. Naturally, the estimation error at the first stage will be inherited, which may increase the difficulty in identifying common factors and estimating corresponding loadings.

In the remaining of this section, we discuss in details the theoretical construction of our method. To do so, let us denote the vector of individual conditional correlation coefficient functions by $\varrho(u) = (\varrho_{1,2}(u), ..., \varrho_{1,m}(u), \varrho_{2,3}(u), ..., \varrho_{2,m}(u), ..., \varrho_{m-1,m}(u))^\top$. In addition, let $\mathrm{G}(U) = (G_{12}(U), \cdots, G_{1m}(U), G_{23}(U), \cdots, G_{2m}(U), \cdots, G_{m-1,m}(U))^\top$ and $a = (a_{12}, ..., a_{1m}, a_{23}, ..., a_{2m}, ..., a_{m-1,m})^\top$, then write $\varrho(u) = a + \mathrm{G}(u)$ and $\mathrm{G}(U) = \mathbf{B}\mathrm{F}(U)$, where

$$\mathbf{B} = (b_1, ..., b_p) \quad \text{and} \quad \mathrm{F}(U) = (F_1(U), \cdots, F_p(U))^\top, \tag{2.11}$$

where $b_k = (b_{12}^{[k]}, \cdots, b_{1m}^{[k]}, b_{23}^{[k]}, \cdots, b_{2m}^{[k]}, \cdots, b_{m-1,m}^{[k]})^\top, k = 1, ..., p$.

Consequently, with observations at $\{U_t : t = 1, ..., n\}$, we define the $m(m-1)/2 \times n$ matrices

$$\mathbf{G} = (\mathrm{G}(U_1), \cdots, \mathrm{G}(U_n)),\ \ \mathbf{F} = (\mathrm{F}(U_1), \cdots, \mathrm{F}(U_n)),\ \ \boldsymbol{\varrho} = (\varrho(U_1), \cdots, \varrho(U_n))$$

and write

$$\mathbf{G} = \mathbf{BF} \quad \text{and} \quad \boldsymbol{\varrho} = a1_n^\top + \mathbf{G},$$

where 1_n is a column vector of length n with all elements being 1. For ease of exposition, hereafter we let $M = m(m-1)/2$.

From (2.7), since it is reasonable to assume that the information of the pairwise conditional correlation coefficients could be fully captured by the p uncorrelated functional factors, our plan is to apply a similar technique used in principal component analysis to our problem. Let us denote the covariance matrix of $\mathrm{G}(U)$ by

$$\boldsymbol{\Lambda} = \mathrm{Cov}(\mathrm{G}(U)) = E\{\mathrm{G}(U)\mathrm{G}^\top(U)\}. \tag{2.12}$$

An immediate idea is to employ the eigenvalue-eigenvector decomposition. For simplicity, we assume that eigenvalues $\lambda_1, ..., \lambda_M$ of $\boldsymbol{\Lambda}$ satisfy $\lambda_1 > \cdots > \lambda_p > 0$ and $\lambda_{p+1} = \cdots = \lambda_M = 0$ and let $\mathrm{V}_1, ..., \mathrm{V}_M$ denote the corresponding orthonormal eigenvectors. Then $\boldsymbol{\Lambda}$ can be factorized as

$$\boldsymbol{\Lambda} = \mathbf{VDV}^\top = \mathbf{V}_1^*\mathbf{D}^*\mathbf{V}_1^{*\top}, \tag{2.13}$$

where $\mathbf{D} = \mathrm{diag}(\lambda_1, ..., \lambda_M)$ is a $M \times M$ diagonal matrix, $\mathbf{V} = (\mathrm{V}_1, ..., \mathrm{V}_M) = (\mathbf{V}_1^*, \mathbf{V}_2^*)$ is a $M \times M$ matrix, $\mathbf{D}^* = \mathrm{diag}(\lambda_1, ..., \lambda_p)$, $\mathbf{V}_1^* = (\mathrm{V}_1, ..., \mathrm{V}_p)$, and

$\mathbf{V}_2^* = (\mathrm{V}_{p+1}, ..., \mathrm{V}_M)$.

On the one hand, we have the eigenvalue-eigenvector decomposition stated in (2.13). But on the other hand, we indicated previously that $\mathrm{G}(U) = \mathbf{B}\mathrm{F}(U)$, so that

$$\mathbf{\Lambda} = \mathbf{B}E\{\mathrm{F}(U)\mathrm{F}^\top(U)\}\mathbf{B}^\top. \tag{2.14}$$

In order to proceed, we assume $E\{\mathrm{F}(U)\mathrm{F}^\top(U)\} = \mathbf{D}^*$, which is equivalent to suggesting that $\mathrm{F}(U) = \mathbf{V}_1^{*\top}\mathrm{G}(U)$. Another way of illustrating this point is to consider the matrix $E\{\mathrm{G}(U)\mathrm{F}^\top(U)\}$, which is

$$E\{\mathrm{G}(U)\mathrm{G}^\top(U)\mathbf{V}_1^*\} = E\{\mathbf{B}\mathrm{F}(U)\mathrm{F}^\top(U)\},$$

$$\mathbf{\Lambda}\mathbf{V}_1^* = \mathbf{B}\mathbf{D}^*,$$

$$\mathbf{V}_1^*\mathbf{D}^* = \mathbf{B}\mathbf{D}^*,$$

since $\mathbf{D}^* = \mathrm{diag}(\lambda_1, ..., \lambda_p)$, thus

$$\mathbf{B} = \mathbf{V}_1^* \;\text{ or }\; b_j = \mathrm{V}_j. \tag{2.15}$$

Expressions (2.15) will be essential when we introduce the estimation procedure in the next section.

2.2.3 Estimator of conditional correlation coefficients

Let $\hat{\mu}_k(u)$, $\hat{\mu}_\ell(u)$, $\hat{\sigma}_k^2(u)$ and $\hat{\sigma}_\ell^2(u)$ be local linear estimators of $\mu_k(u)$, $\mu_\ell(u)$, $\sigma_k^2(u)$ and $\sigma_\ell^2(u)$, respectively. Furthermore, note that ε_k and ε_ℓ are unobservable in practice, but can be estimated by $\hat{\varepsilon}_{k,t} = (r_{k,t} - \hat{\mu}_k(U_t))/\hat{\sigma}_k(U_t)$

and $\hat{\varepsilon}_{\ell,t} = (r_{\ell,t} - \hat{\mu}_\ell(U_t))/\hat{\sigma}_\ell(U_t)$. We can then write

$$\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} = \varrho_{k,\ell}(U_t) + \epsilon_{k,\ell,t} + \hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} - \varepsilon_{k,t}\varepsilon_{\ell,t}.$$

By applying local linear method, an alternative estimator of $\varrho_{k,\ell}(u)$ can be constructed as

$$\hat{\varrho}_{k,\ell}(u) = \frac{\sum_{t=1}^n W_{n,h}(U_t - u)\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t}}{\sum_{t=1}^n W_{n,h}(U_t - u)}, \tag{2.16}$$

where $W_{n,h}(U_t - u) = s_{n,h,2}K_h(U_t - u) - s_{n,h,1}K_h(U_t - u)(U_t - u)$, $K(\cdot)$ is a kernel function, $K_h(U_t - u) = K\left(\frac{U_t - u}{h}\right)/h$, and $s_{n,h,r} = \sum_{t=1}^n K_h(U_t - u)(U_t - u)^r$ for $r = 0, 1, 2$. Moreover, by letting

$$\varrho^*_{k,\ell}(u) = \frac{\sum_{t=1}^n W_{n,h}(U_t - u)\varepsilon_{k,t}\varepsilon_{\ell,t}}{\sum_{t=1}^n W_{n,h}(U_t - u)},$$

then we are able to write

$$\hat{\varrho}_{k,\ell}(u) = \varrho^*_{k,\ell}(u) + \frac{\sum_{t=1}^n W_{n,h}(U_t - u)(\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} - \varepsilon_{k,t}\varepsilon_{\ell,t})}{\sum_{t=1}^n W_{n,h}(U_t - u)}. \tag{2.17}$$

We will present in Section 2.3 the asymptotic properties of $\hat{\varrho}_{k,\ell}(u)$.

2.2.4 Estimators of common functional factors and coefficients

The basic construction of the model discussed in Section 2.2.2 suggests that we can make use of the eigenvalue-eigenvector decomposition to estimate the common functional factors and loading coefficients. To do so, we must first obtain an empirical estimate of the covariance matrix $\mathbf{\Lambda}$, which we will take the following approximation in the current chapter

$$\mathbf{\Lambda}_{\mathbf{G}} = \frac{1}{n}\mathbf{G}\mathbf{G}^{\top}.$$

Once the empirical estimate of the conditional correlation coefficients are obtained, then we may estimate $a_{k\ell}$ by

$$\hat{a}_{k\ell} = n^{-1}\sum_{t=1}^{n}\hat{\varrho}_{k,\ell}(U_t).$$

We then estimate each function $G_{k\ell}(u)$ separately by

$$\hat{G}_{k\ell}(u) = \frac{\sum_{t=1}^{n}(\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} - \hat{a}_{k\ell})W_{n,h}(U_t - u)}{\sum_{t=1}^{n}W_{n,h}(U_t - u)}, \tag{2.18}$$

so that we may form $\hat{\mathrm{G}}(U) = (\hat{G}_{12}(U), \cdots, \hat{G}_{1m}(U), \hat{G}_{23}(U), \cdots, \hat{G}_{2m}(U), \cdots, \hat{G}_{m-1,m}(U))^{\top}$. With observations at $\{U_t : t = 1, ..., n\}$, the $M \times n$ matrix $\mathbf{G}$ can be estimated by $\hat{\mathbf{G}} = (\hat{\mathrm{G}}(U_1), \cdots, \hat{\mathrm{G}}(U_n))$. Accordingly, an estimate of $\mathbf{\Lambda}_{\mathbf{G}}$ can be constructed as

$$\mathbf{\Lambda}_{\hat{\mathbf{G}}} = \frac{1}{n}\hat{\mathbf{G}}\hat{\mathbf{G}}^{\top}.$$

Secondly, we obtain the empirical estimates of the eigenvalues $\lambda_1, ..., \lambda_M$ and the corresponding orthonormal eigenvectors $\mathrm{V}_1, ..., \mathrm{V}_M$ of $\mathbf{\Lambda}$. The asymptotic results presented in Section 2.3 suggest that we can do so through computing the eigenvalues and the corresponding orthonormal eigenvectors of $\mathbf{\Lambda}_{\hat{\mathbf{G}}}$, which are defined in this chapter as $\hat{\lambda}_1, ..., \hat{\lambda}_M$ and $\hat{\mathrm{V}}_1, ..., \hat{\mathrm{V}}_M$, respectively. Recall that our goal is to obtain

$$\hat{\mathbf{B}} = (\hat{b}_1, ..., \hat{b}_p) \quad \text{and} \quad \hat{\mathrm{F}}(U) = (\hat{F}_1(U), \cdots, \hat{F}_p(U))^\top, \tag{2.19}$$

where $\hat{b}_k = (\hat{b}_{12}^{[k]} \cdots \hat{b}_{1m}^{[k]}, \hat{b}_{23}^{[k]}, \cdots \hat{b}_{2m}^{[k]}, \cdots, \hat{b}_{m-1,m}^{[k]})^\top, k = 1, ..., m$. They are the estimates of $\mathbf{B}$ and $\mathrm{F}(U)$ as defined in (2.19), respectively. The first p component functions can be obtained by $\hat{F}_j(u) = \hat{\mathrm{V}}_j^\top \hat{\mathrm{G}}(u)$ for $j = 1, ..., p$. Finally, based on (2.15) we can directly estimate b_j by $\hat{b}_j = \hat{\mathrm{V}}_j$.

Next, we present the estimators of the common functional factors and loading coefficients under the assumption that there exist a number of common factors $p \leq m$ such that $\lambda_1 > \cdots > \lambda_p > 0$, $\lambda_{p+1} = \cdots = \lambda_M = 0$. However, this quantity is unknown in practice. Furthermore, previous experience of functional principal component analysis shows that statistical inference is more difficult for higher-order principal components. Estimation of the new reduced rank model does share a similar difficulty and so selecting the number of common factors is also an important model selection problem.

To this end, Li, Wang and Carroll (2013) introduced a number of information criteria, which are useful in selecting the number of principal components within the context of functional data analysis. In principle, these criteria should also be useful for selecting the common factors in our context. Inspired by Bai and Ng (2002), we consider the following class of information criteria:

$$IC(p) \quad = \quad \log[\hat{\sigma}_{[p]}^2] + p g_{M,n}, \tag{2.20}$$

where

$$\hat{\sigma}^2_{[p]} = \frac{1}{nM}\sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{\ell\neq k}^{m}\left(\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} - \hat{a}_{k\ell} - \hat{b}^{[1]}_{k\ell}\hat{F}_1(U_t) - \cdots - \hat{b}^{[p]}_{k\ell}\hat{F}_p(U_t)\right)^2$$

is defined similarly to the estimated variance in Bai and Ng (2002) and

$$g_{M,n} = \left(\frac{M+n}{nM}\right)\log\left(\frac{nM}{M+n}\right),$$

is a penalty function. Finally, we select the number of components as

$$\hat{p} = \min_p IC(p).$$

2.3 Asymptotics

We first present the asymptotic properties of the estimator for $\varrho_{k,\ell}(u)$. For the estimator $\hat{\varrho}_{k,\ell}(u)$ defined by (2.16), the following asymptotic results are provided.

Theorem 2.1. Suppose that the regularity conditions (C1)-(C6) in the Appendix hold, then for particular k and ℓ, as $n \to \infty$, we have

$$(nh)^{1/2}\{\hat{\varrho}_{k,\ell}(u) - \varrho_{k,\ell}(u) - \frac{1}{2}w_2^K B_{\hat{\varrho}_{k,\ell}}(u)h^2\} \to N(0, f_U^{-1}(u)\omega_{2,k,\ell}(u)), \quad (2.21)$$

where

$$B_{\hat{\varrho}_{k,\ell}}(u) = \varrho''_{k,\ell}(u) - \frac{\varrho_{k,\ell}(u)(\sigma_k^2(u))''}{2\sigma_k^2(u)} - \frac{\varrho_{k,\ell}(u)(\sigma_\ell^2(u))''}{2\sigma_\ell^2(u)},$$

$$\omega_{2,k,\ell}(u) = \nu_K^2\zeta_\epsilon^{k,\ell}(u) + \frac{1}{4}\nu_{K*K}^2\varrho_{k,\ell}^2(u)\zeta_\xi^{k,\ell}(u) - \varrho_{k,\ell}(u)\nu_{K,K*K}\zeta_{\epsilon,\xi}^{k,\ell}(u),$$

with

$$\zeta_{\epsilon}^{k,\ell}(u) = E\{\epsilon_{k,\ell,t}^2 | U_t = u\}, \zeta_{\xi}^{k,\ell}(u) = E\{(\xi_{k,t} + \xi_{\ell,t})^2 | U_t = u\},$$

$$\zeta_{\epsilon,\xi}^{k,\ell}(u) = E\{\epsilon_{k,\ell,t}(\xi_{k,t} + \xi_{\ell,t}) | U_t = u\}.$$

Next, we present asymptotic results for estimators of $\hat{F}_j(u)$ and $\hat{b}_{k\ell}^{[j]}$. Let

$$\tilde{\epsilon}_t = (\epsilon_{1,2,t}, ..., \epsilon_{1,m,t}, \epsilon_{2,3,t}, ..., \epsilon_{2,m,t}, ..., \epsilon_{m-1,m,t})^\top,$$

$$\tilde{\xi}_t = (\xi_{1,t} + \xi_{2,t}, ..., \xi_{1,t} + \xi_{m,t}, \xi_{2,t} + \xi_{3,t}, ..., \xi_{2,t} + \xi_{m,t}, ..., \xi_{m-1,t} + \xi_{m,t})^\top,$$

and $\boldsymbol{\epsilon} = (\tilde{\epsilon}_1, ..., \tilde{\epsilon}_n)$, $\boldsymbol{\xi} = (\tilde{\xi}_1, ..., \tilde{\xi}_n)$.

Theorem 2.2. Suppose that the eigenvalues of $\mathbf{\Lambda}$ satisfy $\lambda_1 > ... > \lambda_p > 0$, $\lambda_{p+1} = \cdots = \lambda_M = 0$. Let $\mathbf{I}$ be the identity matrix of size M, and $(\lambda_j \mathbf{I} - \mathbf{\Lambda})^+$ be the Moore-Penrose inverse of $\lambda_j \mathbf{I} - \mathbf{\Lambda}$. Under conditions (C1)-(C6), as $n \to \infty$, for $j = 1, ..., p$,

$$\sqrt{n}\left(\hat{\lambda}_j - \lambda_j - (\frac{1}{2} w_2^K h^2) E\Big\{2F_j(U)F_j''(U) - b_j^\top F_j(U)(\varrho(U) \circ \sigma(U))\Big\}\right) \xrightarrow{d} N(0, \sigma_{\lambda_j}^2), \quad (2.22)$$

where $\circ$ denotes the hadamard product of two matrices having the same dimensions, and

$$\begin{aligned} \sigma_{\lambda_j}^2 &= E\{I_{j,1}^2\} + 2\sum_{s=1}^{\infty} E\{I_{j,1} I_{j,s+1}\} \\ &= E\Big\{F_j^2(U_1) b_j^\top \mathrm{Cov}\big(2\tilde{\epsilon}_1 - \varrho(U_1) \circ \tilde{\xi}_1 \big| U_1\big) b_j\Big\} + E\{F_j^4(U_1)\} - \lambda_j^2 \\ &+ 2\sum_{s=1}^{\infty} E\Big\{F_j(U_1)F_j(U_{s+1}) b_j^\top \mathrm{Cov}\big(2\tilde{\epsilon}_1 - \varrho(U_1) \circ \tilde{\xi}_1, 2\tilde{\epsilon}_{s+1} - \varrho(U_{s+1}) \circ \tilde{\xi}_{s+1} \big| U_1, U_{s+1}\big) b_j\Big\} \end{aligned}$$

$$+2\sum_{s=1}^{\infty}E\{(F_j^2(U_1)-\lambda_j)(F_j^2(U_{s+1})-\lambda_j)\},$$

with

$$I_{j,t} \;=\; 2b_j^\top\tilde{\epsilon}_tF_j(U_t)-b_j^\top(\varrho(U_t)\circ\tilde{\xi}_t)F_j(U_t)+F_j^2(U_t)-EF_j^2(U).$$

Moreover, for the corresponding estimated eigenvectors $\hat{\mathrm{V}}_1,...,\hat{\mathrm{V}}_p$, under conditions (C1)-(C6), as $n\to\infty$, for $j=1,...,p$,

$$\sqrt{n}\Big(\hat{\mathrm{V}}_j-\mathrm{V}_j-(\frac{1}{2}w_2^Kh^2)E\mathrm{W}_{j,1}\Big)\xrightarrow{d}\mathbf{N}_M(0,\mathbf{\Sigma}_{\mathrm{V}_j}), \tag{2.23}$$

where

$$E\mathrm{W}_{j,1}=E\Big\{(\lambda_j\mathbf{I}-\mathbf{\Lambda})^+\Big[\sum_{i=1}^{p}\mathrm{V}_iF_i(U_t)F_j''(U_t)+\sum_{i=1}^{p}\mathrm{V}_iF_j(U_t)F_i''(U_t)$$
$$-\frac{1}{2}\sum_{i=1}^{p}\mathrm{V}_iF_i(U_t)\mathrm{V}_j^\top(\varrho(U_t)\circ\sigma(U_t))-\frac{1}{2}F_j(U_t)(\varrho(U_t)\circ\sigma(U_t))\Big]\Big\},$$

$$\mathbf{\Sigma}_{\mathrm{V}_j}=\mathrm{Cov}(\mathrm{H}_{j,1})+2\sum_{s=1}^{\infty}\mathrm{Cov}(\mathrm{H}_{j,1},\mathrm{H}_{j,s+1})$$
$$=(\lambda_j\mathbf{I}-\mathbf{\Lambda})^+\Bigg[\Big(\sum_{i=1}^{p}\mathrm{V}_i\mathrm{V}_j^\top F_i(U_1)+F_j(U_1)\mathbf{I}\Big)\mathrm{Cov}\big(\tilde{\epsilon}_1-\frac{1}{2}\varrho(U_1)\circ\tilde{\xi}_1|U_1\big)$$
$$\Big(\sum_{i=1}^{p}\mathrm{V}_j\mathrm{V}_i^\top F_i(U_1)+F_j(U_1)\mathbf{I}\Big)+2\sum_{s=1}^{\infty}\mathrm{Cov}\Big(\big(\sum_{i=1}^{p}\mathrm{V}_i\mathrm{V}_j^\top F_i(U_1)+F_j(U_1)\mathbf{I}\big)\big(\tilde{\epsilon}_1-\frac{1}{2}\varrho(U_1)\circ\tilde{\xi}_1\big),$$
$$\big(\sum_{i=1}^{p}\mathrm{V}_i\mathrm{V}_j^\top F_i(U_{s+1})+F_j(U_{s+1})\mathbf{I}\big)\big(\tilde{\epsilon}_{s+1}-\frac{1}{2}\varrho(U_{s+1})\circ\tilde{\xi}_{s+1}\big)\Big)$$
$$+\Big(\sum_{i=1}^{p}\mathrm{V}_iF_i(U_1)F_j(U_1)\Big)\Big(\sum_{i=1}^{p}\mathrm{V}_i^\top F_i(U_1)F_j(U_1)\Big)$$
$$+2\sum_{s=1}^{\infty}\mathrm{Cov}\Big(\sum_{i=1}^{p}\mathrm{V}_iF_i(U_1)F_j(U_1),\sum_{i=1}^{p}\mathrm{V}_iF_i(U_{s+1})F_j(U_{s+1})\Big)\Bigg](\lambda_j\mathbf{I}-\mathbf{\Lambda})^+,$$

with

$$\mathrm{W}_{j,t}=(\lambda_j\mathbf{I}-\mathbf{\Lambda})^+\Big[\sum_{i=1}^{p}\mathrm{V}_iF_i(U_t)F_j''(U_t)+\sum_{i=1}^{p}\mathrm{V}_iF_j(U_t)F_i''(U_t)$$

$$-\frac{1}{2}\sum_{i=1}^{p} \mathrm{V}_i F_i(U_t)\mathrm{V}_j^\top(\varrho(U_t)\circ\sigma(U_t)) - \frac{1}{2}F_j(U_t)(\varrho(U_t)\circ\sigma(U_t))\Big],$$

$$\mathrm{H}_{j,t} = (\lambda_j\mathbf{I}-\mathbf{\Lambda})^{+}\Big[\Big(\sum_{i=1}^{p}\mathrm{V}_i\mathrm{V}_j^\top F_i(U_t) + F_j(U_t)\mathbf{I}\Big)\big(\tilde{\epsilon}_t - \frac{1}{2}\varrho(U_t)\circ\tilde{\xi}_t\big)$$
$$+\sum_{i=1}^{p}\mathrm{V}_i F_i(U_t)F_j(U_t)\Big].$$

Because $b_j = \mathrm{V}_j$ by (2.15), and $b_j = (b_{12}^{[j]}, ..., b_{1m}^{[j]}, b_{23}^{[j]}, ..., b_{2m}^{[j]}, ..., b_{m-1,m}^{[j]})$, the asymptotic results for the estimated coefficients vector $\hat{b}_j$ is equivalent to results for $\hat{\mathrm{V}}_j$. In this case, the following corollary could be obtained directly from the above theorem.

Corollary 2.1. Suppose that all assumptions in Section 2.7 are fulfilled, then for a particular estimated vector $\hat{b}_j$, as $n\to\infty$, for $j = 1, ..., p$,

$$\sqrt{n}\Big(\hat{b}_j - b_j - (\frac{1}{2}w_2^K h^2)E\mathrm{W}_{j,1}\Big) \xrightarrow{d} \mathbf{N}_M(0, \mathbf{\Sigma}_{\mathrm{V}_j}), \tag{2.24}$$

where $E\mathrm{W}_{j,1}$ and $\mathbf{\Sigma}_{\mathrm{V}_j}$ are the same as which have been given in Theorem 2.2.

Theorem 2.3. Assume that conditions (C1)-(C6) in Section 2.7 hold, and section 2.2.4 shows that $F_j(u) = \mathrm{V}_j^\top \mathrm{G}(u)$, $\hat{F}_j(u) = \hat{\mathrm{V}}_j^\top\hat{\mathrm{G}}(u)$, as $n\to\infty$, we have

$$\sqrt{nh}\left(\hat{F}_j(u) - F_j(u) - (\frac{1}{2}w_2^K h^2)EA_1(u)\right) \xrightarrow{d} N(0, \sigma_{F_j}^2), \tag{2.25}$$

where

$$EA_1(u) = [F_j''(u) - EF_j''(U)] - \frac{1}{2}\mathrm{V}_j^\top\Big[\Big(\varrho(u)\circ\sigma(u)\Big) - E\Big(\varrho(U)\circ\sigma(U)\Big)\Big]$$
$$+E\Big[F_i(U)F_j''(U)\mathrm{V}_i^\top + F_j(U)F_i''(U)\mathrm{V}_i^\top - \frac{1}{2}\mathrm{V}_j^\top(\varrho(U)\circ\sigma(U))F_i(U)\mathrm{V}_i^\top$$

$$-\frac{1}{2}(\varrho^\top(U)\circ\sigma^\top(U))F_j(U)\Big](\lambda_j-\mathbf{\Lambda})^+\mathbf{V}_1^*\mathrm{F}(u),$$

$$\sigma_{F_j}^2=\frac{V_j^\top\Big[\nu_K^2\mathrm{Var}(\tilde\epsilon_1)+\frac{1}{4}\nu_{K*K}^2\Big((\varrho(u)\varrho^\top(u))\circ\mathrm{Var}(\tilde\xi_1)\Big)-\nu_{K,K*K}E\{\tilde\epsilon_1(\varrho^\top(u)\circ\tilde\xi_1^\top)\}\Big]V_j}{f_U(u)}.$$

Finally, we present the asymptotic consistency of $\hat p$, which is selected as the minimizer of the above-introduced information criterion, to the true number of common factors. Assume that the true value of p is p_0. For $p\le p_0$, denote $\mathbf{V}_{1,[p]}^*=(\mathrm{V}_1,...,\mathrm{V}_p)$, $\mathbf{V}_{1,[p+1:p_0]}^*=(\mathrm{V}_{p+1},...,\mathrm{V}_{p_0})$, $\mathbf{D}_{[p]}^*=diag(\lambda_1,...,\lambda_p)$, and $\mathbf{D}_{[p+1:p_0]}^*=diag(\lambda_{p+1},...,\lambda_{p_0})$.

Theorem 2.4. Let $\hat p$ be the minimizer of the information criterion defined in (2.20) among $0\le p\le p_{max}$ with $p_{max}>p_0$ being a fixed search limit, and the regularity conditions (C1)-(C6) hold. If the penalty function g_n satisfies (i) $g_{m,n}\xrightarrow{P}0$, (ii) $g_{m,n}/\Big(h^2+\big(\frac{\log n}{nh}\big)^{\frac{1}{2}}\Big)\xrightarrow{P}\infty$ as $n\to\infty$. Then, $\lim_{n\to\infty}P(\hat p=p_0)=1$.

2.4 Simulation Studies

The main focus of this section is to present simulation studies that examine the finite sample performance of the newly proposed framework. In particular, our objective is to examine the finite sample performance of (i) the local linear estimator for the conditional co-movement of returns, (ii) the newly proposed estimators for the common factors, (iii) the information criterion for selecting the number of the common factors, and to conduct a robustness analysis of the finite sample performance under some specific features, which are common in finance. To achieve these objectives, our studies are conducted based on simulated data from a known data generating pro-

Table 2.1: Finite Sample Performance of the Estimation Procedure

		$m = 15$			$m = 30$		
v	ASE	100	300	600	100	300	600
2	ASE_{F_1}	0.1187	0.0467	0.0321	0.1046	0.0419	0.0279
	ASE_{F_2}	0.1145	0.0592	0.0347	0.1011	0.0496	0.0329
	ASE_C	0.0057	0.0021	0.0009	0.0065	0.0018	0.0008
3	ASE_{F_1}	0.1834	0.0877	0.0552	0.1709	0.0838	0.0529
	ASE_{F_2}	0.1070	0.0603	0.0318	0.1104	0.0472	0.0321
	ASE_C	0.0052	0.0019	0.0009	0.0063	0.0018	0.0009

cess. Specifically, we assume that the return process follows

$$r_k = a_k + b_k\mu(U) + c_{k0}\epsilon_0 + c_{k1}\epsilon_1 f_1(U) + c_{k2}\epsilon_2 f_2(U), \;\; k = 1, ..., m, \tag{2.26}$$

where $a_k, b_k, c_{k0}, c_{k1}, c_{k2}$ are constant coefficients and $\epsilon_0, \epsilon_1, \epsilon_2$ are random renovations with zero mean. For the model in (2.26), it is clear that $E(r_k|U) = a_k + b_k\mu(U)$. In all the simulation studies in this section, we define $\mu(U) = U$ with $U \sim$ Uniform$(0, 1)$, while the required parameters in (2.26) are generated from independent normal distributions, specifically $a_k, b_k, c_{k0}, c_{k1}, c_{k2} \sim$ Normal$(0, 0.2)$. In order to demonstrate the robustness of our method, we consider two illustrative scenarios as follows:

Scenario 1: Let $\epsilon_0, \epsilon_1, \epsilon_2 \overset{\text{IID}}{\sim}$ Normal(0,1). In addition, let

$$f_1(U) = \sqrt{1 + \cos(v\pi U)} \text{ and } f_2(U) = \sqrt{1 + \sin(2\pi U)}.$$

The above specifications suggest that we have

$$Cov(r_k, r_\ell|U) = Corr(r_k, r_\ell|U) \equiv C_{k\ell}(U) = \alpha_{k\ell} + \beta_{k\ell}F_1(U) + \gamma_{k\ell}F_2(U), \tag{2.27}$$

where $\alpha_{k\ell} = c_{k0}c_{\ell 0} + c_{k1}c_{\ell 1} + c_{k2}c_{\ell 2}$, $\beta_{k\ell} = c_{k1}c_{\ell 1}$ and $\gamma_{k\ell} = c_{k2}c_{\ell 2}$. In the other words, $C_{k\ell}(U)$ involves two common factors defined by

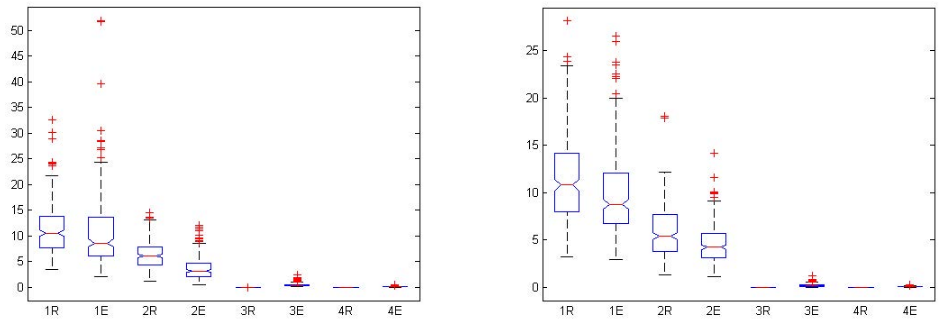
$$F_1(U) = \cos(v\pi U) \text{ and } F_2(U) = \sin(2\pi U). \tag{2.28}$$

In the simulation study that follows, we set the value of parameter v in (2.28) as either 2 or 3. Note that the latter introduces a rougher first common factor compared to the former and hence the resulting conditional correlation functions are less smooth as the results. These functions can be considered as representing structural breaks in the conditional co-movements of returns.

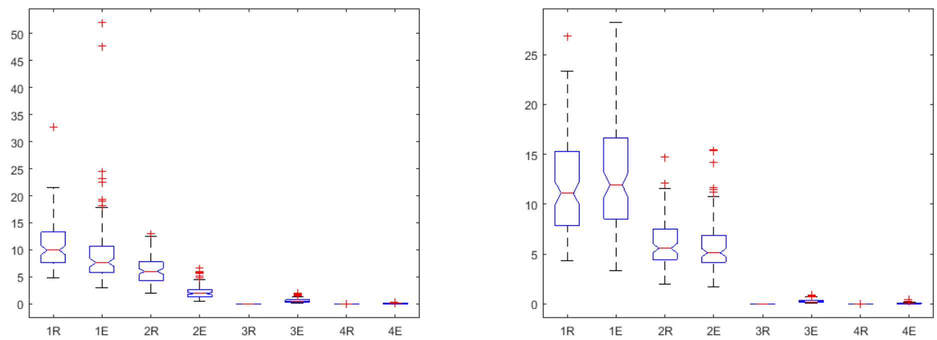
Scenario 2: Let $f_1(U)$ and $f_2(U)$ be defined as in Scenario 1, where $v = 2$, but let $\epsilon_0, \epsilon_1, \epsilon_2 \overset{\text{IID}}{\sim} t_\nu$. Such specifications suggest that we have instead the $C_{k\ell}(U)$ with parameters $\alpha_{k\ell} = c_{k0}c_{\ell 0}\sigma_\epsilon^2 + c_{k1}c_{\ell 1}\sigma_\epsilon^2 + c_{k2}c_{\ell 2}\sigma_\epsilon^2$, $\beta_{k\ell} = c_{k1}c_{\ell 1}\sigma_\epsilon^2$ and $\gamma_{k\ell} = c_{k2}c_{\ell 2}\sigma_\epsilon^2$, where $\sigma_\epsilon^2 = \nu/(\nu - 2)$ is the unconditional variance of ϵ_j, for $j = 1, 2, 3$. In the simulation study that follows, we set the parameter ν to 20, 15, 10 or 5. In the probability theory, it is well-known that the Student's t distribution has heavier tails than those of the normal distribution. Hence, from the finance point of view, Scenario 2 simulate return processes with a heavy-tailed behavior. The first three parameter values, namely 20, 15, and 10, reflect the range of values we obtain by fitting the Student's t-distribution with the MLE to the empirically estimated standardized returns of the Dow30, which is denoted in Section 2.5 by $\hat{\varepsilon}_{k,t}$. To this end, it seems to be the case that multiple estimation and smoothing steps, which are required, lead to confidence intervals that includes point-estimates which are relatively close to normality (see Section 2.5.1 for details). In addition, $\nu = 5$ is included as a benchmark.

Figure 2.1: Boxplots for eigenvalues calculated based on $C_{k\ell}(\cdot)$ and $\hat{C}_{k\ell}(\cdot)$ at $m = 30$.

(a) Scenario 1 for $v = 2$ with $n = 100$ and 600 (left and right panel, respectively)



(b) Scenario 1 for $v = 3$ with $n = 100$ and 600 (left and right panel, respectively)



We will concentrate first on the simulation work done based on the Scenario 1. For the first set of simulation results in Tables 2.1, 2.2 and Figure 2.1, we set the number of observations on the time series dimension as $n = 100, 300, 600$ or 1000. We would also like to investigate the importance of the number of assets in the portfolio on the finite sample performance

Table 2.2: Finite Sample Performance of the Information Criteria

v	m	n	$\widehat{p}=0$	$\widehat{p}=1$	$\widehat{p}=2$	$\widehat{p}=3$	$\widehat{p}=4$
2	15	100	0.2560	0.5960	0.1480	0.0000	0.0000
		300	0.0640	0.3760	0.5600	0.0000	0.0000
		600	0.0160	0.2520	0.7320	0.0000	0.0000
		1000	0.0080	0.0800	0.9120	0.0000	0.0000
	30	100	0.2360	0.5760	0.1680	0.0000	0.0000
		300	0.0440	0.3760	0.5800	0.0000	0.0000
		600	0.0000	0.0000	0.9960	0.0040	0.0000
		1000	0.0000	0.0000	1.0000	0.0000	0.0000
3	15	100	0.2480	0.4520	0.3000	0.0000	0.0000
		300	0.0360	0.1280	0.8360	0.0000	0.0000
		600	0.0120	0.0800	0.9080	0.0000	0.0000
		1000	0.0560	0.0160	0.9280	0.0000	0.0000
	30	100	0.2400	0.4360	0.3240	0.0000	0.0000
		300	0.0200	0.1000	0.8800	0.0000	0.0000
		600	0.0120	0.0680	0.9200	0.0000	0.0000
		1000	0.0000	0.0000	1.0000	0.0000	0.0000

and therefore set the parameter m as either 15 or 30. The number of simulation replications is 250. We focus first on the finite-sample performance of the local linear estimator for the conditional co-movement and the proposed estimators for the common factors. The relevant simulation results are summarized in Table 2.1 and Figure 2.1. In the table, the short abbreviation "ASE" stands for the "average squared errors". For $j = 1, 2$,

$$ASE_{F_j} = \frac{1}{n}\sum_{t=1}^{n}\{\hat{F}_j(U_t) - F_j(U_t)\}^2$$

and

$$ASE_C = \frac{1}{nM}\sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{\ell\neq k}^{m}\{\hat{C}_{k\ell}(U_t) - C_{k\ell}(U_t)\}^2$$

Table 2.3: Finite Sample Performance with Non-normal Renovations at $m = 30$

	$\nu = 20$		$\nu = 10$		$\nu = 5$	
ASE	100	600	100	600	100	600
ASE_{F_1}	0.1377	0.0283	0.1591	0.0404	0.3558	0.2777
ASE_{F_2}	0.1564	0.0382	0.1985	0.0642	0.4443	0.2719
ASE_C	0.0072	0.0012	0.0116	0.0017	0.0314	0.0054

measure the finite-sample performance of the proposed estimator for the jth common factor and for the estimator of the conditional co-movement of returns for any one simulation replication, respectively. For given values of m and n, Table 2.1 reports the averages of ASE_{F_j} over the simulation replications. In all cases, the estimation errors have a strong tendency to converge to zero as the number of observations increases. An interesting point to make is the fact that increasing the number of asset from $m = 15$ to $m = 30$ is able to slightly improve the overall finite sample performance. In addition, the short abbreviations "R" and "E" (for example, as in "1R" and "1E") in Figure 2.1 indicate that the eigenvalues are computed based on $C_{k\ell}(\cdot)$ (as defined in (2.27)) and $\hat{C}_{k\ell}(\cdot)$, respectively. Since there are two common factors, i.e. $p_0 = 2$, in our model example, 3R and 4R in Figure 2.1 are appropriately equal to zero. From the figures, it is apparent that the estimation of the eigenvalues performs well, especially since 3E and 4E in the figures are virtually zero across all simulation replications and since the distributions of the estimates tend to follow closely those of the true eigenvalues. Therefore, we have convincing evidence that the proposed estimation procedure for the common factors perform well especially for the number of observations of above 500, i.e. about two-year of sample for daily return data.

The important factor contributing to this success is the ability of our method to accurately estimate the conditional co-movement of the simulated returns. In Table 2.1, this is demonstrated by the small magnitude and the tendency of the averaged ASE_C to converge to zero. Let us also point out that specifying the conditional variable U as in Section 4.1 of the current chapter contains a special case, which is consistent to taking $\tau = \frac{t}{n} \in (0, 1)$, for $t = 1, \ldots, n$. When such a special case is considered our experimental design is of a similar nature to that of Engle (2002), which was also used in Aslanidis and Casas (2013), CS hereafter, to illustrate the finite sample performance of the local-linear estimator introduced in their paper. Note that for this special case the CS estimator is merely a simplified version of the local linear estimator introduced in the current chapter. On the one hand, this suggests that satisfactory simulation results in this section can be interpreted as the ability of our method to nonparametrically model the conditional covariance matrix of returns under misspecification. On the other hand, it also means that the finite-sample superiority of the nonparametric estimator found in CS over the DCC and cDCC models should also hold for the local linear estimator introduced in the current chapter.

Our attention is now shifted to the finite sample performance of the above-introduced information criterion for selecting the number of common factors. Note that the error terms, which are required in the calculation, are estimated based on $\varepsilon_k = (r_k - a_k - b_k U)/\sqrt{\sigma_k^2(U)}$, where $\sigma_k^2(U) = c_{k0}^2 + c_{k1}^2 f_1^2(U) + c_{k2}^2 f_2^2(U)$. The empirical distribution of the selected number of components summarized in Table 2.2 is obtained by setting $p_{max} = 4$ with $p_0 = 2$, which should be obvious from the specification of (2.27). In Table 2.2, it is clear that lower numbers of common factors than p_0 are often wrongly

Table 2.4: Finite Sample Performance with Non-normal Renovations at $m = 30$

ν	n	$\widehat{p} = 0$	$\widehat{p} = 1$	$\widehat{p} = 2$	$\widehat{p} = 3$	$\widehat{p} = 4$
20	100	0.2600	0.6200	0.1200	0.0000	0.0000
	600	0.0040	0.0080	0.9800	0.0000	0.0000
10	100	0.1050	0.7100	0.1850	0.0000	0.0000
	600	0.0000	0.0150	0.9850	0.0000	0.0000
5	100	0.1950	0.6100	0.1950	0.0000	0.0000
	600	0.0000	0.0750	0.9250	0.0000	0.0000

selected when $n = 100$. However, the results improve substantially as we increase the number of observations to $n = 300$. Further improvement is made when $n = 600$ and 1000 where the right number of common factors is selected up to 100% of the replications for $m = 30$.

We will now concentrate on the simulation work done based on the Scenario 2. Since the importance of the size of portfolio has already examined previously, it is sufficient for our purpose to set the number of observations, n, to either 100 or 600 with $m = 30$. The simulation results are presented in Tables 2.3 and 2.4. In Table 2.3, the effects of the deviation from the normality assumption by within the range found in our empirical data, i.e. ν is between 20 to 10, seem to be minimal. Significance changes in the results only become apparent by a reduction of the degree of freedom to $\nu = 5$, i.e. a level by which data transformation might be required for an application where empirical support for the Student's t-distribution and the degree of freedom can be established. Nonetheless, such a negative changes are not apparent in Table 2.4, which show the finite sample performance of the information criteria. The information criteria seems to have performed consistently well across the degree of freedom in question.

2.5 Effects of market variables on the correlation structure

The empirical study in this section focuses on estimation and analysis of the conditional correlation coefficients for returns of a portfolio of the Dow30 for the observation period between 1 July 1990 to 31 July 2014. Important questions that will be the subject of main interest are how and what drives the observed time-varying correlation structure of the Dow30 portfolio. In the literature, while there is a broad agreement that the correlation structure in financial markets is not constant over time, an outstanding issue of concern is on the driving factor (or factors) behind the observed time variation. Generally, there are two school of thoughts, who are contradictorily in favor of the market volatility and the market return, respectively. The following paragraphs provide a brief review of these arguments in turn.

A number of previous studies have found that the cross-correlations estimated during volatile periods are significantly larger compared to those computed during calm periods. Using multivariate GARCH models, Longin and Solnik (1995) reported that cross correlations among international markets tended to grow especially in periods of high volatility. Similarly, Ramchand and Susmel (1998) examined the relation between variance and correlation under a conditional time and state varying structure, and found that the correlations are much larger when U.S. market is in a high variance condition. Furthermore, Chesnay and Jondeau (2001) applied a multivariate Markov-switching model, where the correlation matrix are varied across regimes, to explore the relationship between stock market turbulence and international correlation, and detected the significant increase of correlation during the turbulent periods. In addition, there were also other studies based on the Markov-switching models who have also found that correlation was generally higher in high-volatility regime (see, for example, Ang and Bekaert (2002)).

On the other hand, Longin and Solnik (2001) established a pattern of asymmetric dependence using extreme value theory, which implied that international stock markets were more highly correlated during extraordinary market downturns than during extraordinary market upturns. Later, Ang and Chen (2002) developed a statistic for testing the asymmetries in conditional correlations based on exceedance correlation and established evidence in support of Longin and Solnik (2001). Another branch of relevant work attempted to connect the variability of correlations of stock returns to the overall economic condition, which was represented by a proxy of market return. Erb, Harvey and Viskanta (1994), for example, suggested that correlations were time-varying and dependent on the economic circumstances. More importantly, they found a strong tendency for correlation to rise during periods of recession.

It is noteworthy that these schools of thought often consider the market return and volatility as two separate and competing entities. Hence, in order to perform the empirical analysis of interest, we may select the conditional variable, U, as either a measure of the market return or that of the market volatility. However, in the literature it has long been discussed the observed tendency of an asset's volatility to be negatively associated with the asset's return, i.e. what is commonly referred to as the "leverage effect".

Furthermore, it has also been demonstrated that the leverage effect is basically asymmetric, i.e. declines in stock prices are accompanied by larger increases in volatility than the decrease in volatility that accompanies rising stock markets. Hence, it is also the main interest of the research in this section to also examine if and how the presence of the leverage effect affects our investigation on the driving factor behind the observed time variation of stocks correlations. For the sake of clarity, we will present first in Section 2.5.1 relevant methodological details and estimation results, while a through discussion on the financial implications and interpretation will be given in the Section 2.5.2.

2.5.1 Relevant methodological details and estimation results

Let us begin with the following empirical details: (i) The data used, which consist of the daily close prices (adjusted for dividends and splits) of the Dow30 components and S&P500, and the Chicago Board Options Exchange Market Volatility Index (VIX) between 1 July 1990 to 31 July 2014, are retrieved from Yahoo Finance. (ii) Usually, the closing prices are transformed into returns by taking natural logarithms and differencing. These leads, therefore, to $m = 30$ with $M = 30 \times (30 - 1)/2 = 435$ conditional correlation coefficients and $n = 6068$ number of observations. (iii) The market volatility is represented in our study by the VIX, which is a popular measure of the implied volatility of S&P 500 index options. (iv) The market return is represented in our study by the return of S&P500. Furthermore, it is assumed that the return follows an AR(1)+GARCH(1,1) process. Intuitively, this assumption implies that the leverage-effect may influence the market return through both volatility and its persistence that leads to temporally dependence of the market return, i.e. autocorrelation. As the results, the leverage-effect for the market can be excluded by first modeling the conditional mean and volatility using the AR(1)+GARCH(1,1) model, then devolatilizing the raw market return using the resulting conditional variances. Hereafter, let us refer to the resulting process as the *devolatilized* market return such that raw S&P500 return counterpart is referred to as the *non-devolatilized* market return. (v) We also apply a similar devolatilization to the Dow30 returns.

For the sake of clarity, let us also collect a list of methodological remarks here: (vi) The estimation procedure employed can be summarized as the following steps.

Step 1: For a given selection of U, (either as the nondevolatilized/devolatilized market return or the VIX for market volatility) the first step in our estimation procedure is to obtain the local linear estimates of $\mu_k(U)$, $\mu_\ell(U)$, $\sigma_k^2(U)$, and $\sigma_\ell^2(U)$.

Step 2: These are then used in the calculation of the estimates for the conditional correlation functions, i.e. $\hat{\varrho}_{k,\ell}(u)$ in (2.16).

Step 3: The asymptotic results in Section 2.3 suggest that we can calculate $\hat{G}_{k,\ell}(u)$ as $\hat{\varrho}_{k,\ell}(u) - \hat{a}_{k\ell}$, then construct $\hat{\mathbf{G}}$ in order to obtain the covariance matrix $\mathbf{\Lambda}_{\hat{\mathbf{G}}} = \frac{1}{n}\hat{\mathbf{G}}\hat{\mathbf{G}}^\top$.

Step 4: We are then able to calculate $\mathbf{V}_{1p}^*$ for each value of $p \leq m$, so that the common factor analysis can be conducted based on the $IC(p)$ criterion defined in (2.20).

Step 5: Once the number of common factors is selected, we are then able to obtain the empirical estimate of the common factor based on $\hat{F}_1(u) =$

$\hat{V}_1^\top \hat{G}(u)$.

The 99% point-wise confidence bands are computed based on the asymptotic variance formula, $\sigma^2_{F_1}$, which was defined in Theorem 2.3. This calculation requires the use of $\hat{V}_1$, which is calculated under the condition $\| V_1 \| = 1$, where $\hat{\varepsilon}_{k,t} = (r_{k,t} - \hat{\mu}_k(U_t))/\hat{\sigma}_k(U_t)$, $\hat{\varepsilon}_{\ell,t} = (r_{\ell,t} - \hat{\mu}_\ell(U_t))/\hat{\sigma}_\ell(U_t)$ and $\hat{\xi}_{k,t} = \hat{\varepsilon}^2_{k,t} - 1$. Step 6: To compute the involved nonparametric estimators , we employ the normal kernel function given by $K(x) = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}$ with $h = std(U)/n^{0.2}$. The above choice of kernel function leads to $\nu^2_K = 1/2\sqrt{\pi}$, $\nu^2_{K*K} = 1/2\sqrt{2\pi}$ and $\nu_{K,K*K} = 1/\sqrt{3}\sqrt{2\pi}$. (vii) The methods and associated results introduced in the current chapter are readily applicable to higher-frequency financial data. For example, we should be able to employ, as conveniently in our empirical analysis, the intraday return at the one-minute (or five-minute, ten-minute, etc) sampling frequency. Nonetheless, it is important to note that the main motivation of the current study is on the identification and estimation of the asymmetry of the overall cross-correlations. This differs significantly from other studies that motivate the use of higher frequency-financial data such as Sheppard and Xu (2014). (viii) We have also tried different specifications on the conditional mean and conditional variance equations. However, the functional-based nature of the method and use of the smooth technique mean that they do not bring about significant changes to the results. (ix) We have also attempted to incorporate the asymmetry in the leverage effect into our analysis. This was done by modeling the volatility based on the GJR-GARCH model of Glosten, Jagannathan and Runkle (1993). Although the asymmetric effect of market variables were felt more strongly in magnitude, the differences in the results were not statistically significant. (x) Comparing to, for example, the cDCC model, where

$O(m^3)$ (alternatively $O(m^2)$ or smaller) calculations are required for the full likelihood function (for the composite likelihood function) because of computation of the inverse matrix and constant matrix, our proposed factor approach is based on nonparametric model structure in which m conditional variance functions are estimated at first stage, then $m(m-1)/2$ conditional correlations are estimated nonparametrically. In addition, the eigenvectors of a $m(m-1)/2$ by $m(m-1)/2$ matrix need to be computed to obtain the common functional factors.

The first picture in each panel in Figure 2.2 displays empirical estimates of 435 correlation functions of the Dow30 components conditioned on a given selection of U, i.e. U_{Dv}, U_{Nv} and U_V which denote the devolatilized, nondevolatilized market return and the market volatility, respectively. Although the correlation functions in each of these pictures seem to have its own pattern, overall they tend to share some essential common features. Let us take the first picture of panel (a), which represents the case for U_{Dv}, as an example. In most cases, large negative or positive return on the S&P500 index implies high correlations, i.e. a convex v-shaped conditional correlation function. The common feature is even more apparent in the first picture of panel (c), which represents the case of U_V, where we witness (almost linearly) positive correlation functions with low degree of variation.

Table 2.5: Information Criterion for Common Factor Analysis

U	$IC(\hat{p}=1)$	$IC(\hat{p}=2)$	$IC(\hat{p}=3)$	$IC(\hat{p}=4)$	$IC(\hat{p}=5)$	$IC(\hat{p}=6)$
U_{Dv}	0.7893	0.8435	0.8981	0.9521	1.0063	1.0598
U_{Nv}	1.0833	1.1370	1.1910	1.2450	1.2988	1.3527
U_V	0.4641	0.5194	0.5730	0.6271	0.6800	0.7350

Figure 2.2: Empirical estimates of $\varrho_{k,\ell}(U)$ based on $\hat{\varrho}_{k,\ell}(U)$ and $\hat{\varrho}^{[j]}_{k,\ell}(U)$ for $j = 1$.

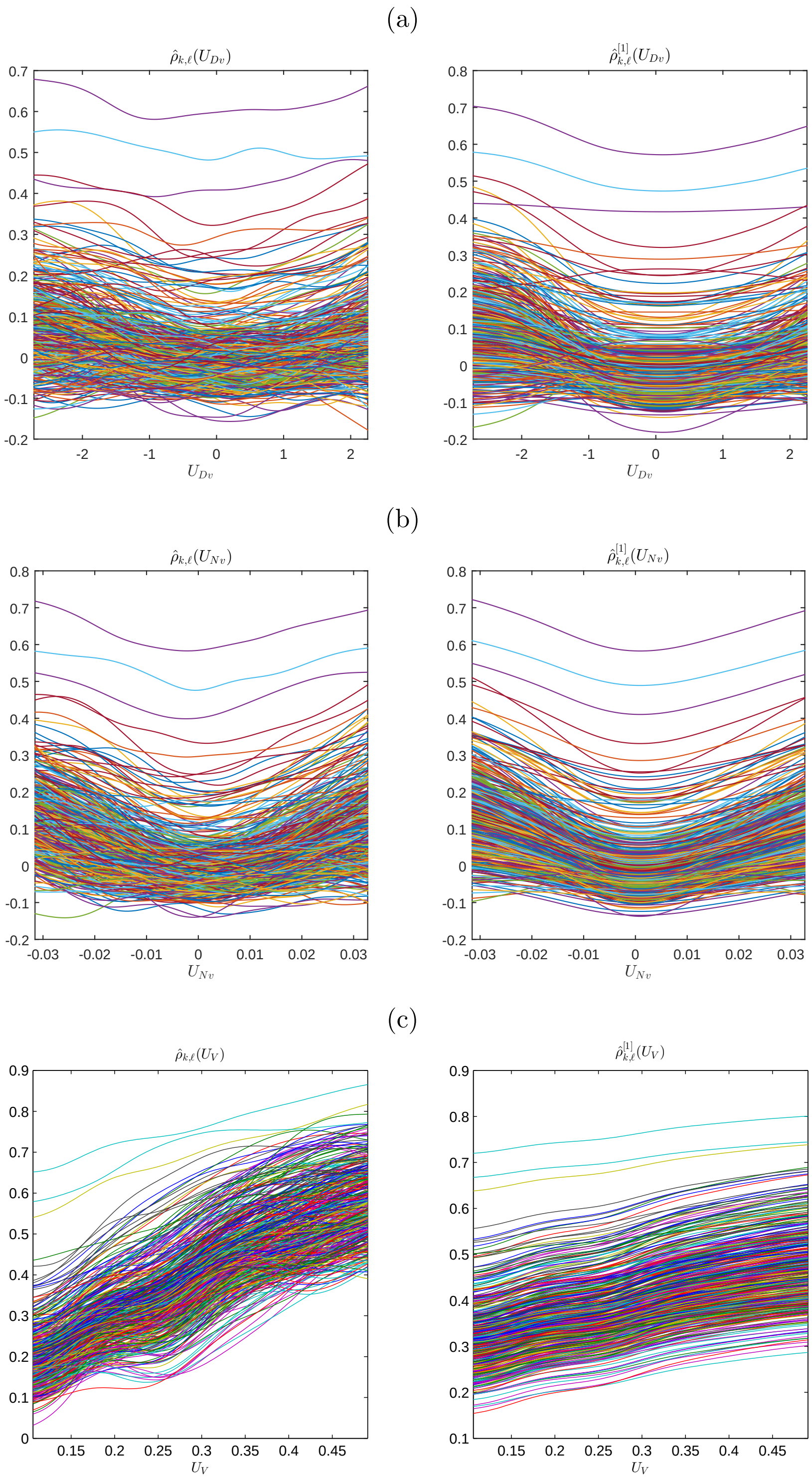
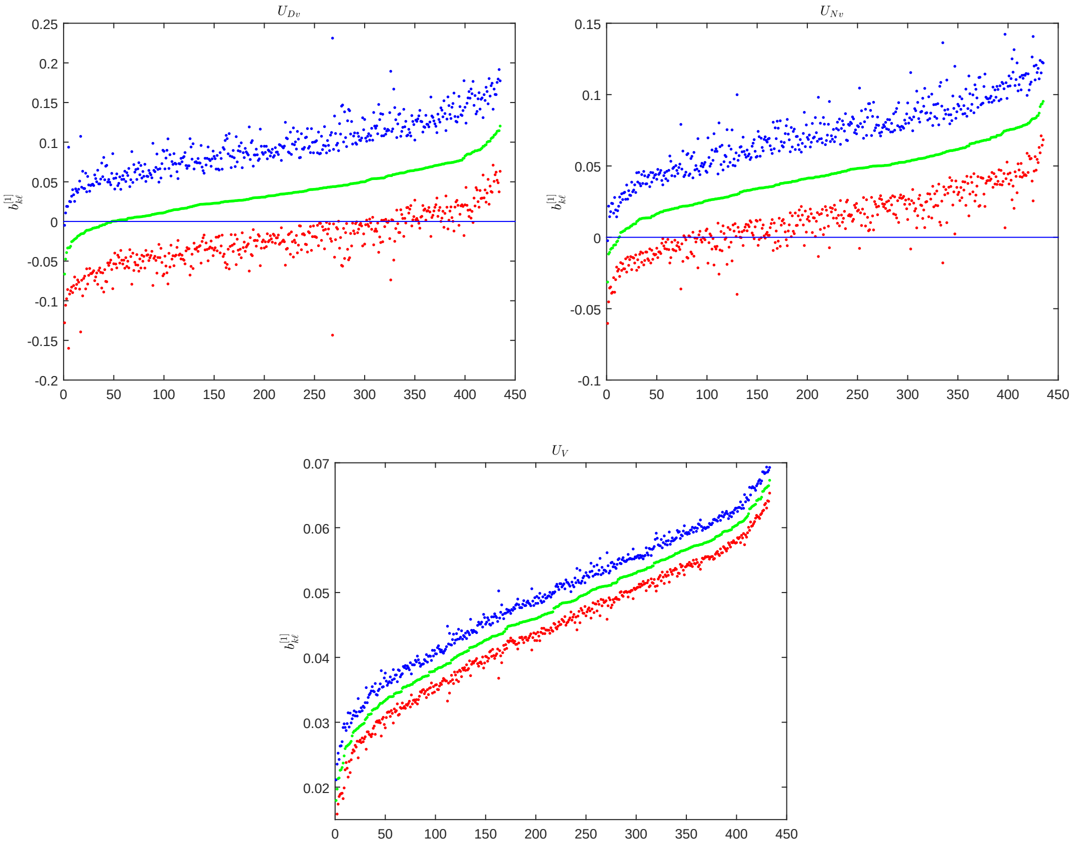
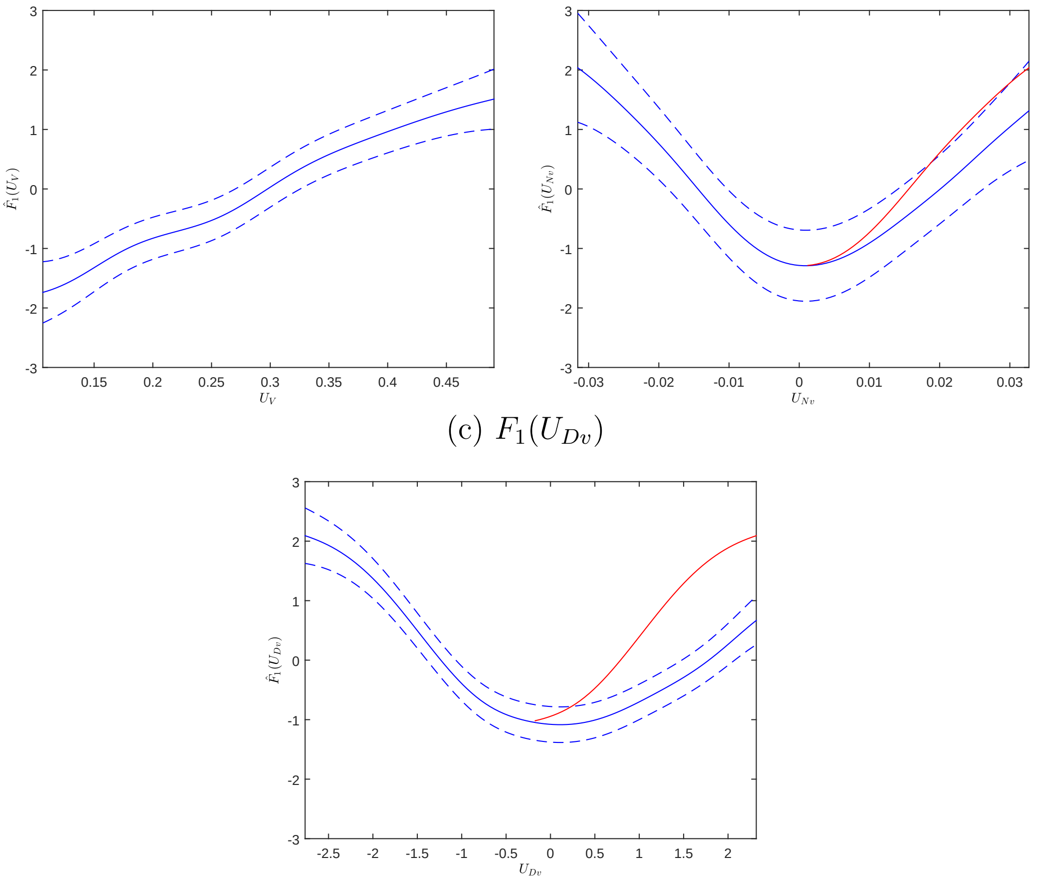


Figure 2.3: $\hat{b}_{k\ell}^{[1]}$ presented in ascending order and the 90%.



Next we perform the common factor analysis based on the information criterion presented in (2.20). The relevant $IC(\hat{p})$ values are shown in Table 2.5. For each of the rows, minimization of these values suggests that a single common factor, $p = 1$, should be selected for all cases. The second pictures in panels (a) to (c) of Figure 2.2 present the empirical estimates of the conditional correlation coefficient functions calculated according to the suggestion made by the information criterion that there exists only one common factor, i.e. $\varrho_{k,\ell}(U) = a_{k\ell} + G_{k\ell}(U) = a_{k\ell} + b_{k\ell}^{[1]} F_1(U)$. Hereafter, let us denote these estimates by $\hat{\varrho}_{k,\ell}^{[1]}(U) = \hat{a}_{k\ell} + \hat{b}_{k\ell}^{[1]} \hat{F}_1(U)$, where the upper-subscript [1] indicates an involvement of a single common factor. In all cases of U, the graphs

Figure 2.4: Empirical estimate of common factors



(a) $F_1(U_V)$ (b) $F_1(U_{Nv})$ (c) $F_1(U_{Dv})$

seem to provide graphical evidence in support of a single common factor, i.e. a conclusion reached due to the fact that the shape of $\hat{\varrho}^{[1]}_{k,\ell}(U)$ closely follows that of $\hat{\varrho}_{k,\ell}(U)$. As the results, the financial discussion in the next section will focus heavily on $F_1(U)$. For the sake of completion, we present in panels (a), (b) and (c) of Figures 2.4 empirical estimates of $F_1(U)$ computed based on U_V, U_{Nv} and U_{Dv} respectively. The 99% point-wise confidence bands were calculated as discussed in Step 4. The red-solid curve in each of the figures will be discussed in detailed in the next section.

We will now focus on the coefficients $b^{[j]}_{k\ell}$. In a sense, $b^{[j]}_{k\ell}$ should quantify the contribution of the j-th common factor on the $k\ell$ conditional correlation function, i.e. a role which is usually played by the so-called functional principal component scores in the functional data analysis literature. This is not the case in our model, however, due to the necessity of the assumption $\mathbf{B} = \mathbf{V}_1^*$, which is stated in (2.15). Nonetheless, since a single common factor was selected, the shape of $\hat{\varrho}_{k,\ell}^{[1]}(U)$ depends on $b_{k\ell}^{[1]}$ and so it is important that we perform inferences for $\hat{b}_{k\ell}^{[1]}$. To do so we calculate the standard errors and consequently the 90% confidence intervals of $\hat{b}_{k\ell}^{[1]}$. Figure 2.3 presents $\hat{b}_{k\ell}^{[1]}$ in ascending order together with the associated 90% confidence intervals for the cases of U_{Dv}, U_{Nv} and U_V (see panels (a), (b) and (c), respectively). In panel (a), the fact that most of the $\hat{b}_{k\ell}^{[1]}$ presented are positive further suggests that the shape of the common factor is well taken by the (pairwise) conditional correlation functions under consideration. In addition, a similar conclusion can also be obtained in panels (b) and (c) but with stronger statistical significance. Observe, however, that the confidence bands in (c) seem to be smaller than those in panels (a) and (b). This is due mostly to the empirical estimate of $\mathbf{\Sigma}_{V_1}$, which is quite small compared to those for cases of the market returns. Such a result was influenced by $\hat{\varrho}_{k,\ell}(U_V)$, which we witnessed in Figure 2.2(c) that they were (almost linearly) positive correlation functions with relatively low degree of variation. In this case, higher correlation leads to larger value of the largest eigenvalue, but also the eigenvector with lower variance. In addition, the first common factor explains up to 97% of the total variations compared to only 70% and 77% in panels (a) and (b), respectively.

2.5.2 Financial implications and interpretations

In this section, we will discuss first important implications of the above results about the effects of the market variables on correlation structure of the Dow30 portfolio. We will then focus more specifically on the asymmetric effect of market return.

Let us begin with the kind of effect that market volatility has on the correlations of the returns of the Dow30. Here, the VIX is used as a proxy for the market volatility. The estimation result in Figure 2.4(a) suggests that correlation significantly increases during volatile periods. This finding is in agreement with the conclusion made by many existing studies (some studies of which are mentioned in the paragraph just above Section 2.5.1). We consider next the empirical estimates of the common factors presented in Figures 2.4(b) and (c) which are associated with the nondevolatilized and devoatilized market returns, respectively. In these cases, the first common factor provides a strong evidence against the constant-correlation hypothesis, which was championed by a number of earlier studies (see Kaplanis (1988), for example).

An important question often investigated in the literature is whether co-movements in the returns are stronger during general market recession than they are during boom periods (see Andersen et al. (2001), and Chesnay and Jondeau (2001), for example). In order to shed some light on this issue, we draw in Figure 2.4(b) a solid red-line, which represents the exact replication of the blue estimate that runs across the negative region of U_{Nv}. The fact that the solid red-line lays almost everywhere in between the pairwise confidence bands provides an empirical evidence (at least at the 1% significance level) against such an asymmetry. In the next step, we perform a similar analysis to the above, but this time based on U_{Dv}, i.e. the devolatilized market return and the result is reported in Figure 2.4(c). We find that the correlations decrease quite significantly in the positive region of the market return compared to those presented in Figure 2.4(b). The fact that the solid red-line lays almost everywhere outside the pairwise confidence bands provides an empirical evidence in support of the asymmetric effect of market return on the conditional correlations of the stock returns. Such a finding can be interpreted as follows. Once the leverage-effect in the market is disentangled and the volatility effect is removed, correlations of the stock returns drop significantly during the bull while remaining unchanged in the bear market. In effect, the tailing off in the correlations leads to the apparent asymmetric-effect of the market return, which is clearly apparent in Figure 2.4(c).

The above discussion considered two extreme cases, where the conditional variable is either the devolatilized, U_{Dv}, or nondevolatilized, U_{Nv}. For the sake of comparison, we also consider a case by which devolatilization is done based on AR(0)+GARCH(1,1). This practice reflects the point we have made that the leverage-effect does not only influence market return through volatility, but also through volatility persistence, which leads to temporally dependence of return, i.e. autocorrelation. However, we have found the result to be closely similar to that in Figure 2.4(c) and so it is not reported.

2.6 Conclusions

In this chapter, we first derived and provided theoretical discussion (specifically the uniform consistency and asymptotic distribution) of an alternative local-linear-smoothing estimator for the (pairwise) conditional correlation coefficients of asset returns. By treating the resulting conditional correlation coefficients as functional data, we developed a new method to study the correlation structure for a portfolio of financial assets. The new method was developed along the line of tools in principal component analysis, which consist of selecting the number and estimation of the common factors together with the corresponding loadings. More importantly, it was based on nonparametric smoothing and thus model free, which makes it less likely to suffer modeling mis-specification compared to the existing methods. We provided detailed theoretical discussion, in particular the consistency and asymptotic distribution, of the information criterion and the nonparametric estimators involved under some regularity conditions. As illustrated in our empirical analysis, the new technique was capable of describing the movement of the local (pairwise) correlations of financial returns conditional upon a particular measure of interest. We studied the effects of a set of market variables, e.g. return and volatility, on the correlation structure of asset returns for a portfolio which consists of the Dow30 components. Under our model setting, we were able to identify the common functional factor that influenced the behavior of cross conditional correlations of the returns. The common factor estimation showed evidence of the well-known asymmetric effect of market return on stock returns correlations. However, through the calculation of the relevant asymptotic pointwise confidence bands, we found that the asymmetric effect was statistically significant only when the leverage effect on the market was taken into consideration. The volatility effect on market return seemed to lead to higher correlations of the stock returns during the bull market, so that the asymmetric effect was not evidenced.

2.7 Theoretical justification

To make statistical inference, we need to find the asymptotic distribution of the estimators, including those for $\mu_k(u), \sigma_k^2(u)$, $\varrho_{k,\ell}(u)$, $F_j(u)$ and $a_{k\ell}, b_{k\ell}^{[j]}$, $k = 1, ..., m, 1 \leq k < \ell \leq m$, $j = 1, ..., p$. The assumptions needed for our analysis are listed below, and the proofs of theorems are provided.

(C1) Let $f_U(\cdot)$ denote the marginal density of U_t, and $f_s(\cdot, \cdot)$ denote the joint density of (U_t, U_{t+s}). Suppose that $f(\cdot)$ has a bounded support, such as $[c, d]$, $f_U(u) > 0$, and $|f_U(u) - f_U(u')| \leq \Delta_1 |u - u'|$ for all given points $u, u' \in [c, d]$ and some $\Delta_1 > 0$. Meanwhile, $f_s(u_0, u_s) > 0$ for $u_0, u_s \in [c, d]$. Further, $\sup_{u \in [c,d]} f_U(u) \leq L_0 < \infty$, $\sup_{u_0, u_s \in [c,d]} f_s(u_0, u_s) \leq L_1 < \infty$.

(C2) $E|r_{k,t}|^{4(1+\delta)} \leq L_2 < \infty$, $E|\epsilon_{k,\ell,t}|^{4(1+\delta)} \leq L_2 < \infty$, for $k, \ell = 1, ..., m$, $t = 1, ..., n$, and some $\delta > 0$. Meanwhile,

$$\sup_{u_0 \in [c,d]} E[|r_{k,t}|^{4(1+\delta)} | U_t = u_0] \leq L_2 < \infty,$$
$$\sup_{u_0 \in [c,d]} E[|\epsilon_{k,\ell,t}|^{4(1+\delta)} | U_t = u_0] \leq L_2 < \infty,$$
$$\sup_{u_0, u_s \in [c,d]} E[|\epsilon_{k,\ell,t}| | U_t = u_0, U_{t+s} = u_s] \leq L_2 < \infty,$$
$$\sup_{u_0, u_s \in [c,d]} E[|\epsilon_{k,\ell,t}\epsilon_{k,\ell,t+s}| | U_t = u_0, U_{t+s} = u_s] \leq L_2 < \infty,$$

for all $s \in \mathbb{Z}$ and some sufficiently large L_2. Moreover, for particular k_1, k_2 and ℓ_1, ℓ_2,

$$E\{\epsilon_{k_1,\ell_1,t}\epsilon_{k_2,\ell_2,t}|U_t = u_0\} = 0, \quad \text{if } \{k_1, \ell_1\} \neq \{k_2, \ell_2\},$$

$$E\{\epsilon_{k_1,\ell_1,t}\epsilon_{k_2,\ell_2,t+s}|U_t = u_0, U_{t+s} = u_s\} = 0, \quad \text{if } \{k_1, \ell_1\} \neq \{k_2, \ell_2\}.$$

(C3) The time series $\{(r_{1,t}, r_{2,t}, \cdots, r_{m,t}, U_t) : t = 1, ..., n\}$ are strictly stationary and strong mixing with mixing coefficient $\alpha(N) \leq CN^{-\beta}$ for some $C > 0$ and $\beta > 2 + \frac{2}{\delta}$ for the same δ as in (C2). Furthermore, suppose that $(r_{1,t}, r_{2,t}, \cdots, r_{m,t}, U_t)$ has the same distribution with $(r_1, r_2, \cdots, r_m, , U)$.

(C4) (i) $\mu_k(u), \sigma_k^2(u), k = 1, ..., m$ are differentiable, and $\mu_k''(u), \sigma_k^{2''}(u)$ are uniformly continuous.

(ii) $F_j(\cdot), j = 1, ..., p$ are differentiable, and $F_j''(\cdot), j = 1, ..., p$ are uniformly continuous. In addition, the coefficients $a_{k\ell}$, $b_{k\ell}^{[j]}$ are bounded by some constants $\bar{a}$, $\bar{b} < \infty$, i.e. $a_{k\ell} < \bar{a}$, $|b_{k\ell}^{[j]}| \leq \bar{b}$ for all $1 \leq k < \ell \leq m$ and $j = 1, ..., p$.

(C5) The continuous symmetric kernel function $K(\cdot)$ has the following properties:

(i) $\int |K(v)|dv < \infty$, $\int K^2(v)dv < \infty$, and $\int K(v)dv = 1$, $\int vK(v)dv = 0$, $\int v^2K(v)dv = w_2^K$, $\int K^2(v)dv = \nu_K^2$.

(ii) For some $0 < C_1 < \infty$ and $0 < \Delta_2 < \infty$, either $K(\cdot)$ is a bounded function with a bounded support on $\mathbb{R}$ (such as $[-C_1, C_1]$), satisfying the Lipschitz condition, i.e. $|K(v_1) - K(v_2)| \leq \Delta_2|v_1 - v_2|$, or $K(\cdot)$ is differentiable, when $v \to \infty$, $K(v)e^{c_0 v} \to 0$ ($c_0 > 0$).

(iii) Let $K * K(v) = \int K(x)K(x+v)dx$, and $\nu_{K,K*K} = \int K(v)K * K(v)dv$, $\nu^2_{K*K} = \int (K * K(v))^2 dv$.

(C6) As $n \to \infty$, $h \to 0$, such that $h = O(n^{-\frac{1}{5}})$.

At the beginning, we introduce the following lemma, which will serve as essential tools to derive asymptotic results for the estimators.

Lemma 2.1. Under the regularity conditions, for model $Y_t = m(U_t) + \sigma(U_t)\varepsilon_t$, $t = 1, ..., n$, where (U_t, Y_t) is a strictly stationary time series, and $E\{\varepsilon_t|U_t\} = 0$. Let $\hat{m}(u)$ be the local linear estimator of $m(u)$.

(i) We have uniformly

$$\hat{m}(u) = m(u) + \frac{1}{2}w_2^K m''(u)h^2 + \frac{1}{nf_U(u)}\sum_{t=1}^{n} K_h(U_t - u)\sigma(U_t)\varepsilon_t + \delta_n, \quad (2.29)$$

where $\delta_n = o_P(h^2 + \{\log n/(nh)\}^{1/2})$.

(ii)

$$\sup_{u\in[c,d]}\left|\frac{1}{n}\sum_{t=1}^{n}[K_h(U_t - u)Y_t - E\{K_h(U_t - u)Y_t\}]\right| = O_p(\{\log n/(nh)\}^{1/2}), \quad (2.30)$$

$$\sup_{u,v\in[c,d]}\left|\frac{1}{n}\sum_{t=1}^{n}[K_h(U_t - u)K_h(U_t - v)Y_t - E\{K_h(U_t - u)K_h(U_t - v)Y_t\}]\right| = O_p(\frac{1}{h}\{\log n/(nh)\}^{1/2}). \quad (2.31)$$

Proof of Lemma 2.1. The proof of (i) could be found in Fan and Gijbels (1996), Fan and Yao (2003), and (ii) follows immediately from the results provided by Mack and Silverman (1982)and Hansen (2008). □

Denote $K_h(U_t-u)$ by $K_{h,t}(u)$, and denote $K*K_h(U_t-u)$ by $K*K_{h,t}(u)$. By this Lemma, we have the following results.

(a) Estimator of $\mu_k(u)$

$$\hat{\mu}_k(u)-\mu_k(u)=\frac{1}{2}w_2^K\mu_k''(u)h^2+N_1(u)+\delta_n,$$

where

$$N_1(u)=\frac{1}{nf_U(u)}\sum_{t=1}^{n}K_{h,t}(u)\sigma_k(U_t)\varepsilon_{k,t}\xrightarrow{d}N\{0,(nhf_U(u))^{-1}\nu_K^2\sigma_k^2(u)\}.$$

(b) Estimator of $\sigma_k^2(u)$.

$$\hat{\sigma}_k^2(u)=\sigma_k^2(u)+\frac{1}{2}w_2^K(\sigma_k^2(u))''h^2+N_2(u)+\delta_n,$$

where

$$N_2(u)=\frac{1}{nf_U(u)}\sum_{t=1}^{n}K_{h,t}(u)\sigma_k^2(U_t)\xi_{k,t}\xrightarrow{d}N\{0,\frac{\nu_K^2\sigma_k^4(u)\sigma_k^{*2}(u)}{nhf_U(u)}\},$$

where $\xi_{k,t}=\varepsilon_{k,t}^2-1$ and $\sigma_k^{*2}(u)=E(\xi_k^2|U=u)$.

(c) Estimator of $\varrho_{k,\ell}(u)$: $\hat{\varrho}_{k,\ell}(u)$. By the definition of $\varrho_{k,\ell}^*(u)$ and (2.29),

$$\varrho_{k,\ell}^*(u)=\varrho_{k,\ell}(u)+\frac{1}{2}w_2^K\varrho_{k,\ell}''(u)h^2+\frac{1}{nf_U(u)}\sum_{t=1}^{n}K_{h,t}(u)\epsilon_{k,\ell,t}+\delta_n.$$

From (2.17),

$$\hat{\varrho}_{k,\ell}(u)=\varrho_{k,\ell}^*(u)+\frac{\sum_{t=1}^{n}W_{n,h}(U_t-u)(\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t}-\varepsilon_{k,t}\varepsilon_{\ell,t})}{\sum_{t=1}^{n}W_{n,h}(U_t-u)}.$$

Together with above results,

$$r_{k,t} - \hat{\mu}_k(U_t) = \sigma_k(U_t)\varepsilon_{k,t} - \frac{1}{2}\mu_k''(U_t)w_2^K h^2 - \frac{1}{nf_U(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\sigma_k(U_q)\varepsilon_{k,q} + \delta_n,$$

$$\frac{1}{\hat{\sigma}_k(U_t)} = \frac{1}{\sigma_k(U_t)}\Big[1 - \frac{\sigma_k^{2''}(U_t)w_2^K h^2}{4\sigma_k^2(U_t)} - \frac{1}{2nf_U(U_t)\sigma_k^2(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\sigma_k^2(U_q)\xi_{k,q} + \delta_n\Big],$$

hence,

$$\hat{\varepsilon}_{k,t} = \varepsilon_{k,t} - \frac{1}{2}w_2^K\left[\frac{\mu_k''(U_t)}{\sigma_k(U_t)} + \frac{\varepsilon_{k,t}\sigma_k^{2''}(U_t)}{2\sigma_k^2(U_t)}\right]h^2 - \frac{1}{n\sigma_k(U_t)f_U(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\sigma_k(U_q)\varepsilon_{k,q}$$
$$-\frac{\varepsilon_{k,t}}{2nf_U(U_t)\sigma_k^2(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\sigma_k^2(U_t)\xi_{k,q} + \delta_n,$$

similarly,

$$\hat{\varepsilon}_{\ell,t} = \varepsilon_{\ell,t} - \frac{1}{2}w_2^K\left[\frac{\mu_\ell''(U_t)}{\sigma_\ell(U_t)} + \frac{\varepsilon_{\ell,t}\sigma_\ell^{2''}(U_t)}{2\sigma_\ell^2(U_t)}\right]h^2 - \frac{1}{n\sigma_\ell(U_t)f_U(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\sigma_\ell(U_q)\varepsilon_{\ell,q}$$
$$-\frac{\varepsilon_{\ell,t}}{2nf_U(U_t)\sigma_\ell^2(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\sigma_\ell^2(U_t)\xi_{\ell,q} + \delta_n,$$

thus,

$$\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} - \varepsilon_{k,t}\varepsilon_{\ell,t} = -\frac{1}{2}w_2^K\Big[\frac{\mu_k''(U_t)\varepsilon_{\ell,t}}{\sigma_k(U_t)} + \frac{\mu_\ell''(U_t)\varepsilon_{k,t}}{\sigma_\ell(U_t)} + \left(\frac{\sigma_k^{2''}(U_t)}{2\sigma_k^2(U_t)} + \frac{\sigma_\ell^{2''}(U_t)}{2\sigma_\ell^2(U_t)}\right)\varepsilon_{k,t}\varepsilon_{\ell,t}\Big]h^2$$
$$-\frac{1}{nf_U(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\Big[\frac{\varepsilon_{\ell,t}\sigma_k(U_q)\varepsilon_{k,q}}{\sigma_k(U_t)} + \frac{\varepsilon_{k,t}\sigma_\ell(U_q)\varepsilon_{\ell,q}}{\sigma_\ell(U_t)}$$
$$+\Big(\frac{\sigma_k^2(U_q)\xi_{k,q}}{2\sigma_k^2(U_t)} + \frac{\sigma_\ell^2(U_q)\xi_{\ell,q}}{2\sigma_\ell^2(U_t)}\Big)\varepsilon_{k,t}\varepsilon_{\ell,t}\Big] + \delta_n,$$

by taking conditional expectation at $U_t = u$,

$$E(\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} - \varepsilon_{k,t}\varepsilon_{\ell,t}|U_t = u)$$
$$= -\frac{1}{2}w_2^K\left[\frac{\varrho_{k,\ell}(u)\sigma_k^{2''}(u)}{2\sigma_k^2(u)} + \frac{\varrho_{k,\ell}(u)\sigma_\ell^{2''}(u)}{2\sigma_\ell^2(u)}\right]h^2 - E\Big\{\frac{1}{nf_U(U_t)}\sum_{q=1}^{n} K_{h,q}(U_t)\Big[\frac{\varepsilon_{\ell,t}\sigma_k(U_q)\varepsilon_{k,q}}{\sigma_k(U_t)}$$
$$+\frac{\varepsilon_{k,t}\sigma_\ell(U_q)\varepsilon_{\ell,q}}{\sigma_\ell(U_t)} + \Big(\frac{\sigma_k^2(U_q)\xi_{k,q}}{2\sigma_k^2(U_t)} + \frac{\sigma_\ell^2(U_q)\xi_{\ell,q}}{2\sigma_\ell^2(U_t)}\Big)\varepsilon_{k,t}\varepsilon_{\ell,t}\Big]\Big|U_t = u\Big\} + \delta_n, \tag{2.32}$$

for the second part of (2.32) on the right hand side, we focus on the approximation of the first term $E\Big\{\frac{1}{nf_U(U_t)}\sum_{q=1}^n K_{h,q}(U_t)\frac{\varepsilon_{\ell,t}\sigma_k(U_q)\varepsilon_{k,q}}{\sigma_k(U_t)}\Big\}$, and the others could be approximated similarly. For example, for $q=t$,

$$E\left[\frac{1}{nf_U(U_t)}K_{h,q}(U_t)\frac{\varepsilon_{\ell,t}\sigma_k(U_q)\varepsilon_{k,q}}{\sigma_k(U_t)}\Big|U_t=u\right]=O(\frac{1}{nh}),$$

for $q\neq t$,

$$E\left[\frac{1}{nf_U(U_t)}K_{h,q}(U_t)\frac{\varepsilon_{\ell,t}\sigma_k(U_q)\varepsilon_{k,q}}{\sigma_k(U_t)}\Big|U_t=u\right]=0.$$

Therefore,

$$E(\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t}-\varepsilon_{k,t}\varepsilon_{\ell,t}|U_t=u) = -\frac{1}{2}w_2^K\left[\frac{\varrho_{k,\ell}(u)\sigma_k^{2''}(u)}{2\sigma_k^2(u)}+\frac{\varrho_{k,\ell}(u)\sigma_\ell^{2''}(u)}{2\sigma_\ell^2(u)}\right]h^2+\delta_n,$$

then the following result could be derived by applying (2.31), i.e.

$$\begin{aligned}\hat{\varrho}_{k,\ell}(u)-\varrho^*_{k,\ell}(u) &= -\frac{1}{2}w_2^K\left[\frac{\varrho_{k,\ell}(u)\sigma_k^{2''}(u)}{2\sigma_k^2(u)}+\frac{\varrho_{k,\ell}(u)\sigma_\ell^{2''}(u)}{2\sigma_\ell^2(u)}\right]h^2-\frac{\varrho_{k,\ell}(u)}{nf_U(u)}\\ &\quad\sum_{t=1}^n K*K_{h,t}(u)\left[\frac{\sigma_k^2(U_t)\xi_{k,t}}{2\sigma_k^2(u)}+\frac{\sigma_\ell^2(U_t)\xi_{\ell,t}}{2\sigma_\ell^2(u)}\right]+\delta_n,\end{aligned}$$

where $K*K(v)=\int K(x)K(x+v)dx$, and $K*K_{h,t}(u)=\frac{1}{h}K*K(\frac{u-U_t}{h})$.

Finally,

$$\begin{aligned}\hat{\varrho}_{k,\ell}(u)-\varrho_{k,\ell}(u) &= \hat{\varrho}_{k,\ell}(u)-\varrho^*_{k,\ell}(u)+\varrho^*_{k,\ell}(u)-\varrho_{k,\ell}(u)\\ &= \frac{1}{2}w_2^K B_{\hat{\varrho}_{k,\ell}}(u)h^2+N_{\hat{\varrho}}(u)+\delta_n,\end{aligned}\qquad(2.33)$$

where

$$B_{\hat{\varrho}_{k,\ell}}(u)=\varrho''_{k,\ell}(u)-\varrho_{k,\ell}(u)\left(\frac{\sigma_k^{2''}(u)}{2\sigma_k^2(u)}+\frac{\sigma_\ell^{2''}(u)}{2\sigma_\ell^2(u)}\right),$$

$$N_{\hat{\varrho}}(u) = \frac{1}{nf_U(u)} \sum_{t=1}^{n} \left[K_{h,t}(u)\epsilon_{k,\ell,t} - K * K_{h,t}(u)\varrho_{k,\ell}(u) \left(\frac{\sigma_k^2(U_t)\xi_{k,t}}{2\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_t)\xi_{\ell,t}}{2\sigma_\ell^2(u)} \right) \right].$$

Lemma 2.2. Suppose that all assumptions are fulfilled, then for particular k and ℓ, as $n \to \infty$, we have uniformly,

$$\begin{aligned} &\hat{\varrho}_{k,\ell}(u) - \varrho_{k,\ell}(u) \\ &= \frac{1}{2} w_2^K \left[\varrho''_{k,\ell}(u) - \varrho_{k,\ell}(u) \left(\frac{\sigma_k^{2''}(u)}{2\sigma_k^2(u)} + \frac{\sigma_\ell^{2''}(u)}{2\sigma_\ell^2(u)} \right) \right] h^2 + \frac{1}{nf_U(u)} \sum_{t=1}^{n} K_{h,t}(u)\epsilon_{k,\ell,t} \\ &- \frac{\varrho_{k,\ell}(u)}{nf_U(u)} \sum_{t=1}^{n} K * K_{h,t}(u) \left[\frac{\sigma_k^2(U_t)\xi_{k,t}}{2\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_t)\xi_{\ell,t}}{2\sigma_\ell^2(u)} \right] + O_p(\delta_n), \end{aligned}$$

where $K_{h,t} = K_h(U_t - u)$, $K * K_{h,t}(u) = \frac{1}{h} K * K(\frac{u-U_t}{h})$, and $\delta_n = o_P(h^2 + \{\log n/(nh)\}^{1/2})$.

Proof of Lemma 2.2. The proof of this lemma could be found from the derivation of (2.33). □

Proof of Theorem 2.1. By Lemma 2.2,

$$\hat{\varrho}_{k,\ell}(u) \quad = \quad \varrho_{k,\ell}(u) + \frac{1}{2} w_2^K B_{\hat{\varrho}_{k,\ell}}(u)h^2 + N_{\hat{\varrho}}(u) + \delta_n,$$

where

$$B_{\hat{\varrho}_{k,\ell}}(u) = \varrho''_{k,\ell}(u) - \varrho_{k,\ell}(u) \left(\frac{\sigma_k^{2''}(u)}{2\sigma_k^2(u)} + \frac{\sigma_\ell^{2''}(u)}{2\sigma_\ell^2(u)} \right),$$

$$N_{\hat{\varrho}}(u) = \frac{1}{nf_U(u)} \sum_{t=1}^{n} \left[K_{h,t}(u)\epsilon_{k,\ell,t} - K * K_{h,t}(u)\varrho_{k,\ell}(u) \left(\frac{\sigma_k^2(U_t)\xi_{k,t}}{2\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_t)\xi_{\ell,t}}{2\sigma_\ell^2(u)} \right) \right].$$

For simplicity, let

$$Z_{n,t}(u) = K_{h,t}(u)\epsilon_{k,\ell,t} - K * K_{h,t}(u)\varrho_{k,\ell}(u) \left(\frac{\sigma_k^2(U_t)\xi_{k,t}}{2\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_t)\xi_{\ell,t}}{2\sigma_\ell^2(u)} \right),$$

then

$$\hat{\varrho}_{k,\ell}(u) - \varrho_{k,\ell}(u) - \frac{1}{2}w_2^K B_{\hat{\varrho}_{k,\ell}}(u)h^2 = \frac{1}{nf_U(u)}\sum_{t=1}^{n} Z_{n,t}(u) + \delta_n.$$

Based on the above formula,

$$E\{N_{\hat{\varrho}}(u)\} = E\{\frac{1}{nf_U(u)}\sum_{t=1}^{n} Z_{n,t}(u)\},$$

$$\text{Var}\{N_{\hat{\varrho}}(u)\} = \frac{1}{nf_U^2(u)}\text{Var}\{Z_{n,1}(u)\} + \frac{2}{nf_U^2(u)}\sum_{s=1}^{n-1}(1-\frac{s}{n})\text{Cov}(Z_{n,1}(u), Z_{n,s+1}(u)).$$

According to the assumptions, $E\{\epsilon_{k,\ell,t}|U_t\} = 0$, $E\{\xi_{k,t}|U_t\} = 0$, $E\{\xi_{\ell,t}|U_t\} = 0$, then $E\{Z_{n,t}(u)|U_t\} = 0$, $E\{N_{\hat{\varrho}}(u)\} = 0$, and

$$\begin{aligned} E\{Z_{n,t}^2(u)|U_t\} &= K_{h,t}^2(u)E\{\epsilon_{k,\ell,t}^2|U_t\} + K*K_{h,t}^2(u)\varrho_{k,\ell}^2(u)\Big[\frac{\sigma_k^4(U_t)E\{\xi_{k,t}^2|U_t\}}{4\sigma_k^4(u)} \\ &+\frac{\sigma_\ell^4(U_t)E\{\xi_{\ell,t}^2|U_t\}}{4\sigma_\ell^4(u)} + \frac{\sigma_k^2(U_t)\sigma_\ell^2(U_t)E\{\xi_{k,t}\xi_{\ell,t}|U_t\}}{2\sigma_k^2(u)\sigma_\ell^2(u)}\Big] \\ -K_{h,t}(u)K*K_{h,t}(u)\varrho_{k,\ell}(u)&\Big[\frac{\sigma_k^2(U_t)E\{\xi_{k,t}\epsilon_{k,\ell,t}|U_t\}}{\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_t)E\{\xi_{\ell,t}\epsilon_{k,\ell,t}|U_t\}}{\sigma_\ell^2(u)}\Big], \end{aligned}$$

then

$$\begin{aligned} &E\{Z_{n,t}^2(u)\} \\ &= \frac{f_U(u)}{h}\Big[\nu_K^2\zeta_\epsilon^{k,\ell}(u) + \frac{1}{4}\nu_{K*K}^2\varrho_{k,\ell}^2(u)\zeta_\xi^{k,\ell}(u) - \varrho_{k,\ell}(u)\nu_{K,K*K}\zeta_{\epsilon,\xi}^{k,\ell}(u)\Big] + o(\frac{1}{h}), \end{aligned}$$

where

$$\zeta_\epsilon^{k,\ell}(u) = E\{\epsilon_{k,\ell,t}^2|U_t = u\}, \zeta_\xi^{k,\ell}(u) = E\{(\xi_{k,t}+\xi_{\ell,t})^2|U_t = u\},$$

$$\zeta_{\epsilon,\xi}^{k,\ell}(u) = E\{\epsilon_{k,\ell,t}(\xi_{k,t}+\xi_{\ell,t})|U_t = u\}.$$

Let $d_n \to \infty$ be a sequence of integers such that $hd_n \to 0$. Define

$$\mathcal{Z}_1 = \sum_{s=1}^{d_n-1} |\mathrm{Cov}(Z_{n,1}(u), Z_{n,s+1}(u))|, \quad \mathcal{Z}_2 = \sum_{s=d_n}^{n-1} |\mathrm{Cov}(Z_{n,1}(u), Z_{n,s+1}(u))|.$$

Conditioning on (U_1, U_{s+1}), and by (C2), (C4) and (C5),

$$\begin{aligned}
&|\mathrm{Cov}(Z_{n,1}(u), Z_{n,s+1}(u))| \\
&= E\Big\{ \frac{1}{h^2}\Big[K\left(\frac{U_1-u}{h}\right)\epsilon_{k,\ell,1} - K*K\left(\frac{U_1-u}{h}\right)\varrho_{k,\ell}(u)\left(\frac{\sigma_k^2(U_1)\xi_{k,1}}{2\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_1)\xi_{\ell,1}}{2\sigma_\ell^2(u)}\right)\Big] \\
&\quad \Big[K\left(\frac{U_{s+1}-u}{h}\right)\epsilon_{k,\ell,s+1} - K*K\left(\frac{U_{s+1}-u}{h}\right)\varrho_{k,\ell}(u) \\
&\qquad \left(\frac{\sigma_k^2(U_{s+1})\xi_{k,s+1}}{2\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_{s+1})\xi_{\ell,s+1}}{2\sigma_\ell^2(u)}\right)\Big]\Big\} \\
&\leq CL_2
\end{aligned}$$

for some generic constant $C > 0$. Then it follows that $\mathcal{Z}_1 \leq d_n CL_2$. We now consider the contribution of $\mathcal{Z}_2$. For this α-mixing process, by Davydov's lemma,

$$|\mathrm{Cov}(Z_{n,1}(u), Z_{n,s+1}(u))| = E|(Z_{n,1}(u)Z_{n,s+1}(u))| \leq 8[\alpha(s)]^{\frac{\delta}{1+\delta}}\{E|Z_{n,1}(u)|^{2(1+\delta)}\}^{\frac{1}{1+\delta}}.$$

By conditioning on U_1, and using (C2) and (C3),

$$\begin{aligned}
&E|Z_{n,1}(u)|^{2(1+\delta)} \\
&= E\left|K_{h,1}(u)\epsilon_{k,\ell,1} - K*K_{h,1}(u)\varrho_{k,\ell}(u)\left(\frac{\sigma_k^2(U_1)\xi_{k,1}}{2\sigma_k^2(u)} + \frac{\sigma_\ell^2(U_1)\xi_{\ell,1}}{2\sigma_\ell^2(u)}\right)\right|^{2(1+\delta)} \\
&\leq CL_2 h^{-2(1+\delta)+1}.
\end{aligned}$$

Hence, for $\frac{\delta}{1+\delta} < \gamma < 1$,

$$\mathcal{Z}_2 \leq \sum_{s=d_n}^{n-1} 8[\alpha(s)]^{\frac{\delta}{1+\delta}}\{E|\tilde{Z}_{n,1}(u)|^{2(1+\delta)}\}^{\frac{1}{1+\delta}} \leq (CL_2)^{\frac{1}{1+\delta}} 8(h^{-2(1+\delta)+1})^{\frac{1}{1+\delta}} \sum_{s=d_n}^{\infty} [s^{-\beta}]^{\frac{\delta}{1+\delta}}$$

$$\le Mh^{-2+\frac{1}{1+\delta}} \sum_{s=d_n}^{\infty} s^{-2} = Mh^{-2+\frac{1}{1+\delta}} d_n^{-\gamma} \sum_{s=d_n}^{\infty} s^{-2+\gamma} = o(1/h)$$

by taking $h^{-1+\frac{1}{1+\delta}} d_n^{-\gamma} = 1$. Together with the above results,

$$\sum_{s=1}^{n-1} \mathrm{Cov}(Z_{n,1}(u), Z_{n,s+1}(u)) = o(1/h),$$

Thus,

$$\begin{aligned}&\mathrm{Var}\{N_{\hat{\varrho}}(u)\}\\ =&\frac{1}{nhf_U(u)}\Big[\nu_K^2\zeta_\epsilon^{k,\ell}(u) + \frac{1}{4}\nu_{K*K}^2\varrho_{k,\ell}^2(u)\zeta_\xi^{k,\ell}(u) - \varrho_{k,\ell}(u)\nu_{K,K*K}\zeta_{\epsilon,\xi}^{k,\ell}(u)\Big] + o(\frac{1}{nh}).\end{aligned}$$

Therefore, the following asymptotic normality could be obtained accordingly,

$$(nh)^{1/2}\{\hat{\varrho}_{k,\ell}(u) - \varrho_{k,\ell}(u) - \frac{1}{2}w_2^K B_{\hat{\varrho}_{k,\ell}}(u)h^2\} \to N(0, f_U^{-1}(u)\omega_{2,k,\ell}(u)),$$

where

$$B_{\hat{\varrho}_{k,\ell}}(u) = \varrho_{k,\ell}''(u) - \varrho_{k,\ell}(u)\left(\frac{\sigma_k^{2''}(u)}{2\sigma_k^2(u)} + \frac{\sigma_\ell^{2''}(u)}{2\sigma_\ell^2(u)}\right),$$

$$\omega_{2,k,\ell}(u) = \nu_K^2\zeta_\epsilon^{k,\ell}(u) + \frac{1}{4}\nu_{K*K}^2\varrho_{k,\ell}^2(u)\zeta_\xi^{k,\ell}(u) - \varrho_{k,\ell}(u)\nu_{K,K*K}\zeta_{\epsilon,\xi}^{k,\ell}(u).$$

□

Proof of Theorem 2.2. From section 2.2.4, local linear method is applied to estimate $G_{k\ell}(u)$,

$$\hat{G}_{k\ell}(u) = \frac{\sum_{t=1}^n(\hat{\varepsilon}_{k,t}\hat{\varepsilon}_{\ell,t} - \hat{a}_{k\ell})W_{n,h}(U_t - u)}{\sum_{t=1}^n W_{n,h}(U_t - u)} = \hat{\varrho}_{k,\ell}(u) - \hat{a}_{k\ell},$$

By (2.33), together with the definition of $\hat{a}_{k\ell}$ as well as (2.7), for a particular $G_{k\ell}(u)$, under the regularity conditions, we could have uniformly for $u \in [c, d]$,

$$\begin{aligned}\hat{G}_{k\ell}(u) &= G_{k\ell}(u) + \hat{\varrho}_{k,\ell}(u) - \varrho_{k,\ell}(u) - \hat{a}_{k\ell} + a_{k\ell} \\ &= G_{k\ell}(u) + \frac{1}{2}w_2^K h^2 B_{\hat{\varrho}_{k,\ell}}(u) + N_{\hat{\varrho}}(u) - \frac{1}{2}w_2^K h^2 \Big(\frac{1}{n}\sum_{t=1}^{n} B_{\hat{\varrho}_{k,\ell}}(U_t)\Big) \\ &\quad -\frac{1}{n}\sum_{t=1}^{n} N_{\hat{\varrho}}(U_t) - \frac{1}{n}\sum_{t=1}^{n} G_{k\ell}(U_t) + \delta_n, \end{aligned} \tag{2.34}$$

where $\delta_n = o_P(h^2 + \{\log n/(nh)\}^{1/2})$.

Let $K_f(u) = (\frac{K_{h,1}(u)}{f_U(u)}, ..., \frac{K_{h,n}(u)}{f_U(u)})^\top$, $K * K_f(u) = (\frac{K*K_{h,1}(u)}{f_U(u)}, ..., \frac{K*K_{h,n}(u)}{f_U(u)})^\top$,

$$\varrho(u) = \begin{pmatrix} \varrho_{1,2}(u) \\ \vdots \\ \varrho_{1,m}(u) \\ \varrho_{2,3}(u) \\ \vdots \\ \varrho_{2,m}(u) \\ \vdots \\ \varrho_{m-1,m}(u)) \end{pmatrix}, \varrho''(u) = \begin{pmatrix} \varrho''_{1,2}(u) \\ \vdots \\ \varrho''_{1,m}(u) \\ \varrho''_{2,3}(u) \\ \vdots \\ \varrho''_{2,m}(u) \\ \vdots \\ \varrho''_{m-1,m}(u)) \end{pmatrix}, \sigma(u) = \begin{pmatrix} \frac{\sigma_1^{2''}(u)}{2\sigma_1^2(u)} + \frac{\sigma_2^{2''}(u)}{2\sigma_2^2(u)} \\ \vdots \\ \frac{\sigma_1^{2''}(u)}{2\sigma_1^2(u)} + \frac{\sigma_m^{2''}(u)}{2\sigma_m^2(u)} \\ \frac{\sigma_2^{2''}(u)}{2\sigma_2^2(u)} + \frac{\sigma_3^{2''}(u)}{2\sigma_3^2(u)} \\ \vdots \\ \frac{\sigma_2^{2''}(u)}{2\sigma_2^2(u)} + \frac{\sigma_m^{2''}(u)}{2\sigma_m^2(u)} \\ \vdots \\ \frac{\sigma_{m-1}^{2''}(u)}{2\sigma_{m-1}^2(u)} + \frac{\sigma_m^{2''}(u)}{2\sigma_m^2(u)} \end{pmatrix},$$

and

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{1,2,1} & \cdots & \epsilon_{1,2,n} \\ \vdots & & \vdots \\ \epsilon_{1,m,1} & \cdots & \epsilon_{1,m,n} \\ \epsilon_{2,3,1} & \cdots & \epsilon_{2,3,n} \\ \vdots & & \vdots \\ \epsilon_{2,m,1} & \cdots & \epsilon_{2,m,n} \\ \vdots & & \vdots \\ \epsilon_{m-1,m,1} & \cdots & \epsilon_{m-1,m,n} \end{pmatrix} = (\tilde{\epsilon}_1, ..., \tilde{\epsilon}_n),$$

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_{1,1}+\xi_{2,1} & \cdots & \xi_{1,n}+\xi_{2,n} \\ \vdots & & \vdots \\ \xi_{1,1}+\xi_{m,1} & \cdots & \xi_{1,n}+\xi_{m,n} \\ \xi_{2,1}+\xi_{3,1} & \cdots & \xi_{2,n}+\xi_{3,n} \\ \vdots & & \vdots \\ \xi_{2,1}+\xi_{m,1} & \cdots & \xi_{2,n}+\xi_{m,n} \\ \vdots & & \vdots \\ \xi_{m-1,1}+\xi_{m,1} & \cdots & \xi_{m-1,n}+\xi_{m,n} \end{pmatrix} = (\tilde{\xi}_1, .., \tilde{\xi}_n),$$

$$\begin{aligned}\boldsymbol{\sigma}_\xi(u) &= \begin{pmatrix} \frac{\sigma_1^2(U_1)\xi_{1,1}}{\sigma_1^2(u)}+\frac{\sigma_2^2(U_1)\xi_{2,1}}{\sigma_2^2(u)} & \cdots & \frac{\sigma_1^2(U_n)\xi_{1,n}}{\sigma_1^2(u)}+\frac{\sigma_2^2(U_n)\xi_{2,n}}{\sigma_2^2(u)} \\ \vdots & & \vdots \\ \frac{\sigma_1^2(U_1)\xi_{1,1}}{\sigma_1^2(u)}+\frac{\sigma_m^2(U_1)\xi_{m,1}}{\sigma_m^2(u)} & \cdots & \frac{\sigma_1^2(U_n)\xi_{1,n}}{\sigma_1^2(u)}+\frac{\sigma_m^2(U_n)\xi_{m,n}}{\sigma_m^2(u)} \\ \frac{\sigma_2^2(U_1)\xi_{2,1}}{\sigma_2^2(u)}+\frac{\sigma_3^2(U_1)\xi_{3,1}}{\sigma_3^2(u)} & \cdots & \frac{\sigma_2^2(U_n)\xi_{2,n}}{\sigma_2^2(u)}+\frac{\sigma_3^2(U_n)\xi_{3,n}}{\sigma_3^2(u)} \\ \vdots & & \vdots \\ \frac{\sigma_2^2(U_1)\xi_{2,1}}{\sigma_2^2(u)}+\frac{\sigma_m^2(U_1)\xi_{m,1}}{\sigma_m^2(u)} & \cdots & \frac{\sigma_2^2(U_n)\xi_{2,n}}{\sigma_2^2(u)}+\frac{\sigma_m^2(U_n)\xi_{m,n}}{\sigma_m^2(u)} \\ \vdots & & \vdots \\ \frac{\sigma_{m-1}^2(U_1)\xi_{2,1}}{\sigma_{m-1}^2(u)}+\frac{\sigma_m^2(U_1)\xi_{m,1}}{\sigma_m^2(u)} & \cdots & \frac{\sigma_{m-1}^2(U_n)\xi_{m-1,n}}{\sigma_{m-1}^2(u)}+\frac{\sigma_m^2(U_n)\xi_{m,n}}{\sigma_m^2(u)} \end{pmatrix} \\ &= (\tilde{\sigma}_{\xi,1}(u), ..., \tilde{\sigma}_{\xi,n}(u)),\end{aligned}$$

therefore,

$$\begin{aligned}\hat{\mathrm{G}}(u) &= \mathrm{G}(u) + \frac{1}{2}w_2^K h^2[\varrho''(u) - \frac{1}{2}\varrho(u)\circ\sigma(u)] + \frac{1}{n}\boldsymbol{\epsilon}K_f(u) \\ &-\frac{1}{2n}\varrho(u)\circ(\boldsymbol{\sigma}_\xi(u)K * K_f(u)) - \frac{1}{2}w_2^K h^2\Big[\frac{1}{n}\sum_{t=1}^n \varrho''(U_t) - \frac{1}{2n}\sum_{t=1}^n \varrho(U_t)\circ\sigma(U_t)\Big] \\ &-\frac{1}{n}\sum_{t=1}^n \mathrm{G}(U_t) - \frac{1}{n}\sum_{t=1}^n\Big[\frac{1}{n}\boldsymbol{\epsilon}K_f(U_t) - \frac{1}{2n}\varrho(U_t)\circ(\boldsymbol{\sigma}_\xi(U_t)K * K_f(U_t))\Big] + \delta_n,\end{aligned} \tag{2.35}$$

and

$$\begin{aligned}\hat{\mathbf{G}} = \mathbf{G} &+ \frac{1}{2}w_2^K h^2(\boldsymbol{\varrho}'' - \frac{1}{2}\boldsymbol{\varrho}\circ\boldsymbol{\sigma}) + \frac{1}{n}\boldsymbol{\epsilon}\mathbf{K}_f - \frac{1}{2n}\boldsymbol{\varrho}\circ(\boldsymbol{\sigma}_\xi\mathbf{K} * \mathbf{K}_f) - \frac{1}{n}\mathbf{G}1_n1_n^\top \\ &-\frac{1}{2}w_2^K h^2(\frac{1}{n}\boldsymbol{\varrho}''1_n1_n^\top - \frac{1}{2n}(\boldsymbol{\varrho}\circ\boldsymbol{\sigma})1_n1_n^\top) \\ &-\frac{1}{n}\Big[\frac{1}{n}\boldsymbol{\epsilon}\mathbf{K}_f1_n1_n^\top - \frac{1}{2n}\Big(\boldsymbol{\varrho}\circ(\boldsymbol{\sigma}_\xi\mathbf{K} * \mathbf{K}_f)\Big)1_n1_n^\top\Big] + \delta_n\end{aligned}$$

$$= \mathbf{G} + \tilde{\boldsymbol{\mathcal{E}}}_n,$$

where $\circ$ denotes the hadamard product of two matrices, $\boldsymbol{\varrho}$, $\boldsymbol{\varrho}''$, $\boldsymbol{\sigma}$ are $M \times n$ matrices, i.e. $\boldsymbol{\varrho} = (\varrho(U_1), ..., \varrho(U_n))$, $\boldsymbol{\varrho}'' = (\varrho''(U_1), ..., \varrho''(U_n))$, $\boldsymbol{\sigma} = (\sigma(U_1), ..., \sigma(U_n))$, $\mathbf{K}_f$ is a $n \times n$ matrix, $\mathbf{K} * \mathbf{K}_f$ is a $n^2 \times n$ matrix, and $\boldsymbol{\sigma}_\xi$ is a $M \times n^2$ matrix, i.e. $\mathbf{K}_f = (K_f(U_1), ..., K_f(U_n))$, $\mathbf{K} * \mathbf{K}_f = \text{diag}(K * K_f(U_1), ..., K * K_f(U_n))$, and $\boldsymbol{\sigma}_\xi = (\boldsymbol{\sigma}_\xi(U_1), ..., \boldsymbol{\sigma}_\xi(U_n))$.

Recall $\varrho_{k,\ell}(u) = a_{k\ell} + G_{k\ell}(u)$ by (2.7), then $\boldsymbol{\varrho} = a 1_M 1_n^\top + \mathbf{G}$, $\boldsymbol{\varrho}'' = \mathbf{G}''$, therefore,

$$\begin{aligned}
\mathbf{E}_n &= \boldsymbol{\Lambda}_{\hat{\mathbf{G}}} - \boldsymbol{\Lambda} \\
&= \frac{1}{n}\hat{\mathbf{G}}\hat{\mathbf{G}}^\top - E\{G(U)G(U)^\top\} = \frac{1}{n}(\mathbf{G} + \tilde{\boldsymbol{\mathcal{E}}}_n)(\mathbf{G} + \tilde{\boldsymbol{\mathcal{E}}}_n)^\top - E\{G(U)G(U)^\top\} \\
&= (\frac{1}{2}w_2^K h^2)\frac{1}{n}\Big[\mathbf{G}''\mathbf{G}^\top + \mathbf{G}\mathbf{G}''^\top - \frac{1}{2}(\boldsymbol{\varrho} \circ \boldsymbol{\sigma})\mathbf{G}^\top - \frac{1}{2}\mathbf{G}(\boldsymbol{\varrho}^\top \circ \boldsymbol{\sigma}^\top)\Big] \\
&\quad - \frac{1}{2n^2}\Big[\big(\boldsymbol{\varrho} \circ (\boldsymbol{\sigma}_\xi \mathbf{K} * \mathbf{K}_f)\big)\mathbf{G}^\top + \mathbf{G}\big(\boldsymbol{\varrho}^\top \circ (\mathbf{K} * \mathbf{K}_f^\top \boldsymbol{\sigma}_\xi^\top)\big)\Big] \\
&\quad + \frac{1}{n^2}\Big(\boldsymbol{\epsilon}\mathbf{K}_f\mathbf{G}^\top + \mathbf{G}\mathbf{K}_f^\top\boldsymbol{\epsilon}^\top\Big) + \frac{1}{n}\mathbf{G}\mathbf{G}^\top - E\{G(U)G(U)^\top\} + o_p(\frac{1}{\sqrt{n}}),
\end{aligned}$$

due to the fact that $h = O(n^{-\frac{1}{5}})$ as $n \to \infty$, and $EG_{k\ell}(U) = 0$, $\sum_{t=1}^n G_{k\ell}(U_t) = O_p(\sqrt{n})$,

$$\mathbf{G}1_n = O_p(\sqrt{n}), \quad \mathbf{G}''1_n = O_p(n).$$

Note that under condition (C5), $K(\cdot)$ is a bounded function with a bounded support, satisfying the Lipschitz condition, then $K * K(\cdot)$ is also bounded with bounded support, and Lipschitz continuous. Note that by (C1), (C2) and (C4), we have $E|G_{k\ell}(U)/f(U)|^{2+\delta} < \infty$, $E\left|\frac{\varrho_{k_1,\ell_1}(U)G_{k_2\ell_2}(U)}{2\sigma_{k_1}^2(U)f_U(U)}\right|^{2+\delta} < \infty$, $E\left|\frac{\varrho_{k_1,\ell_1}(U)G_{k_2\ell_2}(U)}{2\sigma_{\ell_1}^2(U)f_U(U)}\right|^{2+\delta} < \infty$, for particular k, ℓ and k_1, ℓ_1, k_2, ℓ_2, thus the following equations hold uniformly for $u \in [c, d]$,

$$\left|\frac{1}{n}\sum_{q=1}^{n} G_{k\ell}(U_q)\frac{K_h(u-U_q)}{f(U_q)} - G_{k\ell}(u)\right| = O_p(h^2 + \left(\frac{logn}{nh}\right)^{\frac{1}{2}}),$$

$$\Big|\frac{1}{n}\sum_{q=1}^{n} \frac{\varrho_{k_1,\ell_1}(U_q)G_{k_2,\ell_2}(U_q)}{2\sigma_{k_1}^2(U_q)f_U(U_q)} K * K_h(u-U_q) - \frac{\varrho_{k_1,\ell_1}(u)G_{k_2,\ell_2}(u)}{2\sigma_{k_1}^2(u)}\Big|$$
$$= O_p(h^2 + \left(\frac{\log n}{nh}\right)^{\frac{1}{2}}),$$

$$\Big|\frac{1}{n}\sum_{q=1}^{n} \frac{\varrho_{k_1,\ell_1}(U_q)G_{k_2,\ell_2}(U_q)}{2\sigma_{\ell_1}^2(U_q)f_U(U_q)} K * K_h(u-U_q) - \frac{\varrho_{k_1,\ell_1}(u)G_{k_2,\ell_2}(u)}{2\sigma_{\ell_1}^2(u)}\Big|$$
$$= O_p(h^2 + \left(\frac{\log n}{nh}\right)^{\frac{1}{2}}),$$

then the following term could be approximated accordingly,

$$\frac{1}{n^2}\mathbf{G}\mathbf{K}_f^\top \boldsymbol{\epsilon}^\top = \frac{1}{n}\mathbf{G}\boldsymbol{\epsilon}^\top + o_p(\frac{1}{\sqrt{n}}),$$
$$\frac{1}{2n^2}\mathbf{G}(\boldsymbol{\varrho}^\top \circ (\mathbf{K} * \mathbf{K}_f^\top \boldsymbol{\sigma}_\xi^\top)) = \frac{1}{2n}\mathbf{G}(\boldsymbol{\varrho}^\top \circ \boldsymbol{\xi}^\top) + o_p(\frac{1}{\sqrt{n}}).$$

Therefore,

$$\mathbf{E}_n = \boldsymbol{\Lambda}_{\hat{\mathbf{G}}} - \boldsymbol{\Lambda} = (\frac{1}{2}w_2^K h^2)\mathbf{W}_n + \mathbf{H}_{n1} + \mathbf{H}_{n1}^\top + \mathbf{H}_{n2}$$
$$+ o_p(\frac{1}{\sqrt{n}})1_{\frac{m(m-1)}{2}}1_{\frac{m(m-1)}{2}}^\top, \tag{2.36}$$

where

$$\begin{aligned}
\mathbf{W}_n &= \frac{1}{n}[\mathbf{G}\mathbf{G}''^\top + \mathbf{G}''\mathbf{G}^\top - \frac{1}{2}\mathbf{G}(\boldsymbol{\varrho}^\top \circ \boldsymbol{\sigma}^\top) - \frac{1}{2}(\boldsymbol{\varrho} \circ \boldsymbol{\sigma})\mathbf{G}^\top],\\
\mathbf{H}_{n1} &= \frac{1}{n}\mathbf{G}[\boldsymbol{\epsilon}^\top - \frac{1}{2}(\boldsymbol{\varrho}^\top \circ \boldsymbol{\xi}^\top)],\\
\mathbf{H}_{n2} &= \frac{1}{n}\mathbf{G}\mathbf{G}^\top - E\{\mathrm{G}(U)\mathrm{G}(U)^\top\}.
\end{aligned}$$

Because $\boldsymbol{\Lambda}$ is a real symmetric matrix, and V_j is the normalized eigenvector associated with a simple eigenvalue λ_j of $\mathbf{\Lambda}$ for $j = 1, ..., p$. Then by the results in Magnus (1985), a real-valued function $\mathfrak{u}_j$ and a vector function $\mathcal{V}_j$ $(j = 1, ..., p)$ are defined for all $\mathbf{\Lambda}^*$ in some neighbourhood $N(\mathbf{\Lambda})$ of $\mathbf{\Lambda}$ such that

$$\mathfrak{u}_j(\mathbf{\Lambda}) = \lambda_j, \quad \mathcal{V}_j(\mathbf{\Lambda}) = \mathrm{V}_j, \quad \mathfrak{u}_j(\mathbf{\Lambda}_{\hat{\mathbf{G}}}) = \hat{\lambda}_j, \quad \mathcal{V}_j(\mathbf{\Lambda}_{\hat{\mathbf{G}}}) = \hat{\mathrm{V}}_j,$$

$$\mathbf{\Lambda}^*\mathcal{V}_j = \mathfrak{u}_j\mathcal{V}_j, \quad \mathcal{V}_j^\top\mathcal{V}_j = 1, \quad \mathbf{\Lambda}^* \in N(\mathbf{\Lambda}).$$

Moreover, the functions $\mathfrak{u}_j$ and $\mathcal{V}_j$ are ∞ times differentiable, and the differentials at $\mathbf{\Lambda}$ are

$$\begin{aligned} d\mathfrak{u}_j &= \mathrm{V}_j^\top d\mathbf{\Lambda}^*\mathrm{V}_j, \\ d\mathcal{V}_j &= (\lambda_j\mathbf{I} - \mathbf{\Lambda})^+ d\mathbf{\Lambda}^*\mathrm{V}_j, \end{aligned} \tag{2.37}$$

where $\mathbf{I}$ is the identity matrix of size M, and $(\lambda_j\mathbf{I} - \mathbf{\Lambda})^+$ is the Moore-Penrose inverse of $\lambda_j\mathbf{I} - \mathbf{\Lambda}$.

Recall the definition of λ_j, V_j and $\hat{\lambda}_j$, $\hat{V}_j$, by applying (2.37) and Taylor's expansion,

$$\begin{aligned} \hat{\lambda}_j - \lambda_j &= V_j^\top(\mathbf{\Lambda}_{\hat{\mathbf{G}}} - \mathbf{\Lambda})V_j + o_p(\frac{1}{\sqrt{n}}) \\ &= V_j^\top\mathbf{E}_n V_j + o_p(\frac{1}{\sqrt{n}}), \end{aligned} \tag{2.38}$$

$$\begin{aligned} \hat{V}_j - V_j &= (\lambda_j\mathbf{I} - \mathbf{\Lambda})^+(\mathbf{\Lambda}_{\hat{\mathbf{G}}} - \mathbf{\Lambda})V_j + o_p(\frac{1}{\sqrt{n}}) \\ &= (\lambda_j\mathbf{I} - \mathbf{\Lambda})^+\mathbf{E}_n V_j + o_p(\frac{1}{\sqrt{n}}). \end{aligned} \tag{2.39}$$

(i) Since we have assumed that $\mathrm{F}(U) = \mathbf{V}_1^{*\top}\mathrm{G}(U)$, i.e. $F_j(U) = \mathrm{V}_j^\top\mathrm{G}(U)$,

$e_j^\top \mathbf{F} = \mathrm{V}_j^\top \mathbf{G}$, $e_j^\top \mathbf{F}'' = \mathrm{V}_j^\top \mathbf{G}''$, and $b_j = \mathrm{V}_j$ by (2.15),

$$\begin{aligned}\hat{\lambda}_j - \lambda_j &= V_j^\top \Big[(\frac{1}{2} w_2^K h^2)\mathbf{W}_n + \mathbf{H}_{n1} + \mathbf{H}_{n1}^\top + \mathbf{H}_{n2} \Big] V_j + o_p(\frac{1}{\sqrt{n}}) \\ &= (\frac{1}{2} w_2^K h^2) \mathrm{V}_j^\top \mathbf{W}_n \mathrm{V}_j + \mathrm{V}_j^\top (\mathbf{H}_{n1} + \mathbf{H}_{n1}^\top) \mathrm{V}_j + \mathrm{V}_j^\top \mathbf{H}_{n2} \mathrm{V}_j + o_p(\frac{1}{\sqrt{n}}),\end{aligned}$$

with

$$\begin{aligned}(\frac{1}{2} w_2^K h^2) \mathrm{V}_j^\top \mathbf{W}_n \mathrm{V}_j &= (\frac{1}{2} w_2^K h^2) [\frac{2}{n} \mathrm{V}_j^\top \mathbf{G}\mathbf{G}''^\top \mathrm{V}_j - \frac{1}{n} \mathrm{V}_j^\top \mathbf{G} (\boldsymbol{\varrho}^\top \circ \boldsymbol{\sigma}^\top) \mathrm{V}_j] \\ &= (\frac{1}{2} w_2^K h^2) [\frac{2}{n} \sum_{t=1}^n F_j(U_t) F_j''(U_t) - \frac{1}{n} \sum_{t=1}^n \mathrm{V}_j^\top (\varrho(U_t) \circ \sigma(U_t)) F_j(U_t)] \\ &= (\frac{1}{2} w_2^K h^2) [\frac{2}{n} \sum_{t=1}^n F_j(U_t) F_j''(U_t) - \frac{1}{n} \sum_{t=1}^n F_j(U_t) b_j^\top (\varrho(U_t) \circ \sigma(U_t))] \\ \mathrm{V}_j^\top (\mathbf{H}_{n1} + \mathbf{H}_{n1}^\top) \mathrm{V}_j &= \frac{2}{n} \mathrm{V}_j^\top \mathbf{G} \boldsymbol{\epsilon}^\top \mathrm{V}_j - \frac{1}{n} \mathrm{V}_j^\top \mathbf{G} (\boldsymbol{\varrho}^\top \circ \boldsymbol{\xi}^\top) \mathrm{V}_j \\ &= \frac{2}{n} \sum_{t=1}^n F_j(U_t) \mathrm{V}_j^\top \tilde{\epsilon}_t - \frac{1}{n} \sum_{t=1}^n F_j(U_t) \mathrm{V}_j^\top (\varrho(U_t) \circ \tilde{\xi}_t) \\ &= \frac{1}{n} \sum_{t=1}^n F_j(U_t) b_j^\top [2\tilde{\epsilon}_t - (\varrho(U_t) \circ \tilde{\xi}_t)] \\ \mathrm{V}_j^\top \mathbf{H}_{n2} \mathrm{V}_j &= \mathrm{V}_j^\top [\frac{1}{n} \mathbf{G}\mathbf{G}^\top - E\{G(U) G(U)^\top\}] \mathrm{V}_j \\ &= \frac{1}{n} \sum_{t=1}^n F_j^2(U_t) - E F_j^2(U),\end{aligned}$$

where $\boldsymbol{\epsilon} = (\tilde{\epsilon}_1, ..., \tilde{\epsilon}_n)$, and $\boldsymbol{\xi} = (\tilde{\xi}_1, ..., \tilde{\xi}_n)$.

Then, because $F_j(\cdot)$, $F_j''(\cdot)$ are uniformly continuous by (C4), together with (C1) and (C2), we could show that $E|F_j(U_t)|^{4+\delta} < \infty$, $E|F_j(U_t) F_j''(U_t)|^{2+\delta} < \infty$, $E|b_j^\top (\varrho(U_t) \circ \sigma(U_t)) F_j(U_t)|^{2+\delta} < \infty$, $E|b_j^\top \tilde{\epsilon}_t F_j(U_t)|^{2+\delta} < \infty$, $E|b_j^\top (\varrho(U_t) \circ \tilde{\xi}_t) F_j(U_t)|^{2+\delta} < \infty$, and by Hölder's inequality, $E|2 b_j^\top \tilde{\epsilon}_t F_j(U_t) - b_j^\top (\varrho(U_t) \circ \tilde{\xi}_t) F_j(U_t) + F_j^2(U_t)|^{2+\delta} < \infty$ could be obtained accordingly.

Under the α-mixing condition (C3), $\sum_{N=1}^\infty \alpha(N)^{\frac{\delta}{2+\delta}} \le \sum_{N=1}^\infty N^{-(2+\frac{2}{\delta})(\frac{\delta}{2+\delta})}$

$= \sum_{N=1}^{\infty} N^{-2+\frac{2}{2+\delta}} < \infty$, then

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} F_j(U_t)F_j''(U_t) &= E\{F_j(U)F''(U)\} + O(\frac{1}{\sqrt{n}}), \\ \frac{1}{n}\sum_{t=1}^{n} F_j(U_t)b_j^\top(\varrho(U_t)\circ\sigma(U_t)) &= E\{F_j(U)b_j^\top(\varrho(U)\circ\sigma(U))\} + O(\frac{1}{\sqrt{n}}), \end{aligned}$$

thus

$$\begin{aligned} & \sqrt{n}\Big(\hat{\lambda}_j - \lambda_j - (\frac{1}{2}w_2^K h^2)E\{2F_j(U)F''(U) - b_j^\top F_j(U)(\varrho(U)\circ\sigma(U))\}\Big) \\ = & \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big[2b_j^\top\tilde{\epsilon}_t F_j(U_t) - b_j^\top(\varrho(U_t)\circ\tilde{\xi}_t)F_j(U_t) + F_j^2(U_t) - EF_j^2(U)\big] + o(1). \end{aligned}$$

Let $I_{j,t} = 2b_j^\top\tilde{\epsilon}_t F_j(U_t) - b_j^\top(\varrho(U_t)\circ\tilde{\xi}_t)F_j(U_t) + F_j^2(U_t) - EF_j^2(U)$, since $E\{\tilde{\epsilon}_t|U_t\} = 0$, $E\{\tilde{\xi}_t|U_t\} = 0$, for a particular t,

$$\begin{aligned} \mathrm{Var}(I_{j,t}) &= \mathrm{Var}\{2b_j^\top\tilde{\epsilon}_t F_j(U_t) - b_j^\top(\varrho(U_t)\circ\tilde{\xi}_t)F_j(U_t)\} + \mathrm{Var}\{F_j^2(U_t) - EF_j^2(U)\} \\ &= \mathrm{Var}\{b_j^\top[2\tilde{\epsilon}_t - (\varrho(U_t)\circ\tilde{\xi}_t)]F_j(U_t)\} + E\{F_j^4(U_t)\} - \lambda_j^2, \end{aligned}$$

for time t and $t+s$,

$$\begin{aligned} & \mathrm{Cov}(I_{j,t}, I_{j,t+s}) \\ = & \mathrm{Cov}\Big(b_j^\top[2\tilde{\epsilon}_t - (\varrho(U_t)\circ\tilde{\xi}_t)]F_j(U_t) + F_j^2(U_t) - EF_j^2(U), b_j^\top[2\tilde{\epsilon}_{t+s} - \\ & (\varrho(U_{t+s})\circ\tilde{\xi}_{t+s})]F_j(U_{t+s}) + F_j^2(U_{t+s}) - EF_j^2(U)\Big) \\ = & E\Big\{b_j^\top[2\tilde{\epsilon}_t - (\varrho(U_t)\circ\tilde{\xi}_t)][2\tilde{\epsilon}_{t+s}^\top - (\varrho^\top(U_{t+s})\circ\tilde{\xi}_{t+s}^\top)]b_j F_j(U_t)F_j(U_{t+s})\Big\} \\ & + E\{F_j^2(U_t)F_j^2(U_{t+s})\}. \end{aligned}$$

hence, by CLT result for α-mixing series,

$$\sqrt{n}\left(\hat{\lambda}_j - \lambda_j - (\frac{1}{2}w_2^K h^2)E\Big\{2F_j(U)F''(U) - b_j^\top F_j(U)(\varrho(U)\circ\sigma(U))\Big\}\right) \xrightarrow{d} N(0, \sigma^2_{\lambda_j}),$$

where

$$\begin{aligned}\sigma^2_{\lambda_j} &= E\{I^2_{j,1}\} + 2\sum_{s=1}^{\infty} E\{I_{j,1}, I_{j,s+1}\} \\ &= E\Big\{F_j^2(U_1)b_j^\top \mathrm{Cov}\big(2\tilde{\epsilon}_1 - \varrho(U_1)\circ\tilde{\xi}_1\big|U_1\big)b_j\Big\} + E\{F_j^4(U_1)\} - \lambda_j^2 \\ &+2\sum_{s=1}^{\infty} E\Big\{F_j(U_1)F_j(U_{s+1})b_j^\top \mathrm{Cov}\big(2\tilde{\epsilon}_1 - \varrho(U_1)\circ\tilde{\xi}_1, 2\tilde{\epsilon}_{s+1} - \varrho(U_{s+1})\circ\tilde{\xi}_{s+1}\big|U_1, U_{s+1}\big)b_j\Big\} \\ &+2\sum_{s=1}^{\infty} E\{(F_j^2(U_1) - \lambda_j)(F_j^2(U_{s+1}) - \lambda_j)\}\end{aligned}$$

(ii) Similarly, consider the asymptotic properties of the estimated eigenvector $\hat{\mathrm{V}}_j$. Let $\mathbf{I}_p$ be the identity matrix of size p, then substitute (2.36) into (2.39),

$$\begin{aligned}\hat{\mathrm{V}}_j - \mathrm{V}_j &= (\lambda_j\mathbf{I} - \mathbf{\Lambda})^+\mathbf{E}_n\mathrm{V}_j + o_p(\frac{1}{\sqrt{n}}) \\ &= (\lambda_j\mathbf{I} - \mathbf{\Lambda})^+\Big[(\frac{1}{2}w_2^K h^2)\mathbf{W}_n + \mathbf{H}_{n1} + \mathbf{H}_{n1}^\top + \mathbf{H}_{n2}\Big]\mathrm{V}_j + o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

Specifically, $\mathbf{G} = \mathbf{BF} = \mathbf{V}_1^*\mathbf{F}$ by (2.15), and $\sum_{i=1}^p \mathrm{V}_i\mathrm{V}_i^\top = \mathbf{V}_1^*\mathbf{V}_1^{*\top}$, $\mathbf{V}_1^{*\top}\mathbf{V}_1^* = \mathbf{I}_p$. Moreover, $(\lambda_j\mathbf{I} - \mathbf{\Lambda})^\top\mathbf{\Lambda}\mathrm{V}_j = (\lambda_j\mathbf{I} - \mathbf{\Lambda})\lambda_j\mathrm{V}_j = 0$, which means that $(\lambda_j\mathbf{I} - \mathbf{\Lambda})^+\mathbf{\Lambda}\mathrm{V}_j = 0$. Thus,

$$\begin{aligned}&(\frac{1}{2}w_2^K h^2)(\lambda_j\mathbf{I} - \mathbf{\Lambda})^+\mathbf{W}_n\mathrm{V}_j \\ &= (\frac{1}{2}w_2^K h^2)(\lambda_j\mathbf{I} - \mathbf{\Lambda})^+\frac{1}{n}\sum_{t=1}^n\Big[\sum_{i=1}^p \mathrm{V}_iF_i(U_t)F_j''(U_t) + \sum_{i=1}^p \mathrm{V}_iF_j(U_t)F_i''(U_t)\end{aligned}$$

$$-\frac{1}{2}\sum_{i=1}^{p} V_i F_i(U_t) V_j^\top (\varrho(U_t)\circ\sigma(U_t)) - \frac{1}{2}F_j(U_t)(\varrho(U_t)\circ\sigma(U_t))\Big],$$

$$\begin{aligned}
&(\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \mathbf{H}_{n1} V_j \\
&\quad = \quad (\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{p} V_i V_j^\top [\tilde\epsilon_t - \frac{1}{2}(\varrho(U_t)\circ\tilde\xi_t)] F_i(U_t), \\
&(\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \mathbf{H}_{n1}^\top V_j \\
&\quad = \quad (\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \frac{1}{n}\sum_{t=1}^{n}[\tilde\epsilon_t - \frac{1}{2}(\varrho(U_t)\circ\tilde\xi_t)] F_j(U_t), \\
&(\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \mathbf{H}_{n2} V_j \\
&\quad = \quad (\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ [\frac{1}{n}\mathbf{G}\mathbf{G}^\top - \mathbf{\Lambda}] V_j = (\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{p} V_i F_i(U_t) F_j(U_t).
\end{aligned}$$

To investigate the asymptotic normality of the eigenvector $\hat{V}_j$, we consider the asymptotic result of $y^\top \hat{V}_j$ for $y \in \mathbb{R}^M$. Under the α-mixing condition (C3), $\sum_{N=1}^{\infty}\alpha(N)^{\frac{\delta}{2+\delta}} \le \sum_{N=1}^{\infty} N^{-(2+\frac{2}{\delta})(\frac{\delta}{2+\delta})} = \sum_{N=1}^{\infty} N^{-2+\frac{2}{2+\delta}} < \infty$. Let

$$\begin{aligned}
W_{j,t} &= (\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \Big[\sum_{i=1}^{p} V_i F_i(U_t) F_j''(U_t) + \sum_{i=1}^{p} V_i F_j(U_t) F_i''(U_t) \\
&\quad -\frac{1}{2}\sum_{i=1}^{p} V_i F_i(U_t) V_j^\top (\varrho(U_t)\circ\sigma(U_t)) - \frac{1}{2}F_j(U_t)(\varrho(U_t)\circ\sigma(U_t))\Big], \\
H_{j,t} &= (\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \Big[\sum_{i=1}^{p} V_i V_j^\top \big(\tilde\epsilon_t - \frac{1}{2}(\varrho(U_t)\circ\tilde\xi_t)\big) F_i(U_t) + \big(\tilde\epsilon_t - \frac{1}{2}(\varrho(U_t)\circ\tilde\xi_t)\big) F_j(U_t) \\
&\quad + \sum_{i=1}^{p} V_i F_i(U_t) F_j(U_t)\Big] \\
&= (\lambda_j \mathbf{I}-\mathbf{\Lambda})^+ \Big[\Big(\sum_{i=1}^{p} V_i V_j^\top F_i(U_t) + F_j(U_t)\mathbf{I}\Big)\big(\tilde\epsilon_t - \frac{1}{2}\varrho(U_t)\circ\tilde\xi_t\big) \\
&\quad + \sum_{i=1}^{p} V_i F_i(U_t) F_j(U_t)\Big],
\end{aligned}$$

by (C1), (C2) and (C4), for the same δ in the assumptions, $E|y^\top H_{j,t}|^{2+\delta} < \infty$, and $E|y^\top W_{j,t}|^{2+\delta} < \infty$, then for an arbitrary linear combination $y^\top \hat{V}_j$

for $y \in \mathbb{R}^M$,

$$\sqrt{n}\Big(\mathrm{y}^\top \hat{\mathrm{V}}_j - \mathrm{y}^\top \mathrm{V}_j - (\frac{1}{2}w_2^K h^2) E\mathrm{y}^\top \mathrm{W}_{j,1}\Big) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{y}^\top \mathrm{H}_{j,t} + o_p(1),$$

hence, by CLT result for α-mixing series, which means that

$$\sqrt{n}\Big(\mathrm{y}^\top \hat{\mathrm{V}}_j - \mathrm{y}^\top \mathrm{V}_j - (\frac{1}{2}w_2^K h^2) E\mathrm{y}^\top \mathrm{W}_{j,1}\Big) \xrightarrow{d} N(0, \mathrm{y}^\top \boldsymbol{\Sigma}_{\mathrm{V}_j}\mathrm{y})$$

where

$$\begin{aligned}
E\mathrm{W}_{j,1} = E\Big\{&(\lambda_j \mathbf{I} - \boldsymbol{\Lambda})^+ \Big[\sum_{i=1}^{p} \mathrm{V}_i F_i(U_t) F_j''(U_t) + \sum_{i=1}^{p} \mathrm{V}_i F_j(U_t) F_i''(U_t) \\
&-\frac{1}{2}\sum_{i=1}^{p} \mathrm{V}_i F_i(U_t) \mathrm{V}_j^\top (\varrho(U_t) \circ \sigma(U_t)) - \frac{1}{2} F_j(U_t)(\varrho(U_t) \circ \sigma(U_t))\Big]\Big\},
\end{aligned}$$

$$\begin{aligned}
&\boldsymbol{\Sigma}_{\mathrm{V}_j} = \mathrm{Cov}(\mathrm{H}_{j,1}) + 2\sum_{s=1}^{\infty} \mathrm{Cov}(\mathrm{H}_{j,1}, \mathrm{H}_{j,s+1}) \\
&= (\lambda_j \mathbf{I} - \boldsymbol{\Lambda})^+ \Bigg[\Big(\sum_{i=1}^{p} \mathrm{V}_i \mathrm{V}_j^\top F_i(U_1) + F_j(U_1)\mathbf{I}\Big) \mathrm{Cov}\big(\tilde{\epsilon}_1 - \frac{1}{2}\varrho(U_1) \circ \tilde{\xi}_1 | U_1\big) \\
&\Big(\sum_{i=1}^{p} \mathrm{V}_j \mathrm{V}_i^\top F_i(U_1) + F_j(U_1)\mathbf{I}\Big) + 2\sum_{s=1}^{\infty} \mathrm{Cov}\Big(\big(\sum_{i=1}^{p} \mathrm{V}_i \mathrm{V}_j^\top F_i(U_1) + F_j(U_1)\mathbf{I}\big)\big(\tilde{\epsilon}_1 - \frac{1}{2}\varrho(U_1) \circ \tilde{\xi}_1\big), \\
&\big(\sum_{i=1}^{p} \mathrm{V}_i \mathrm{V}_j^\top F_i(U_{s+1}) + F_j(U_{s+1})\mathbf{I}\big)\big(\tilde{\epsilon}_{s+1} - \frac{1}{2}\varrho(U_{s+1}) \circ \tilde{\xi}_{s+1}\big)\Big) \\
&+\Big(\sum_{i=1}^{p} \mathrm{V}_i F_i(U_1) F_j(U_1)\Big)\Big(\sum_{i=1}^{p} \mathrm{V}_i^\top F_i(U_1) F_j(U_1)\Big) \\
&+2\sum_{s=1}^{\infty} \mathrm{Cov}\Big(\sum_{i=1}^{p} \mathrm{V}_i F_i(U_1) F_j(U_1), \sum_{i=1}^{p} \mathrm{V}_i F_i(U_{s+1}) F_j(U_{s+1})\Big)\Bigg](\lambda_j \mathbf{I} - \boldsymbol{\Lambda})^+,
\end{aligned}$$

therefore, by cramér-wold theorem,

$$\sqrt{n}\Big(\hat{\mathrm{V}}_j - \mathrm{V}_j - (\frac{1}{2}w_2^K h^2) E\mathrm{W}_{j,1}\Big) \xrightarrow{d} \mathbf{N}_M(0, \boldsymbol{\Sigma}_{\mathrm{V}_j}).$$

□

Proof of Theorem 2.3. From (2.35), we could directly have the following equation,

$$\hat{\mathrm{G}}(u) = \mathrm{G}(u) + \frac{1}{2}w_2^K h^2[\varrho''(u) - \frac{1}{2}\varrho(u)\circ\sigma(u)] + \frac{1}{n}\boldsymbol{\epsilon}K_f(u)$$
$$-\frac{1}{2n}\varrho(u)\circ(\boldsymbol{\sigma}_\xi(u)K * K_f(u)) - \frac{1}{2}w_2^K h^2\Big[\frac{1}{n}\sum_{t=1}^{n}\varrho''(U_t) - \frac{1}{2n}\sum_{t=1}^{n}\varrho(U_t)\circ\sigma(U_t)\Big]$$
$$-\frac{1}{n}\sum_{t=1}^{n}\mathrm{G}(U_t) - \frac{1}{n}\sum_{t=1}^{n}\Big[\frac{1}{n}\boldsymbol{\epsilon}K_f(U_t) - \frac{1}{2n}\varrho(U_t)\circ(\boldsymbol{\sigma}_\xi(U_t)K * K_f(U_t))\Big] + \delta_n,$$

and recall (2.15), (2.36), $\varrho''(u) = \mathbf{B}\mathrm{F}''(u) = \mathbf{V}_1^*\mathrm{F}''(u)$, $\boldsymbol{\varrho}'' = \mathbf{B}\mathbf{F}'' = \mathbf{V}_1^*\mathbf{F}$, hence,

$$\begin{aligned}
&\hat{F}_j(u) - F_j(u)\\
&= \hat{\mathrm{V}}_j^\top\hat{\mathrm{G}}(u) - \mathrm{V}_j^\top\mathrm{G}(u) = \big(\mathrm{V}_j + (\lambda_j\mathbf{I} - \boldsymbol{\Lambda})^+\mathbf{E}_n\mathrm{V}_j\big)^\top\big(\mathrm{G}(u) + \hat{\mathrm{G}}(u) - \mathrm{G}(u)\big) - \mathrm{V}_j^\top\mathrm{G}(u)\\
&= \mathrm{V}_j^\top(\frac{1}{2}w_2^K h^2)\mathbf{W}_n(\lambda_j\mathbf{I} - \boldsymbol{\Lambda})^+\mathrm{G}(u) + \mathrm{V}_j^\top(\hat{\mathrm{G}}(u) - \mathrm{G}(u)) + \delta_n\\
&= \mathrm{V}_j^\top(\frac{1}{2}w_2^K h^2)\mathbf{W}_n(\lambda_j\mathbf{I} - \boldsymbol{\Lambda})^+\mathbf{V}_1^*\mathrm{F}(u) + (\frac{1}{2}w_2^K h^2)\Big[F_j''(u) - \frac{1}{2}\mathrm{V}_j^\top\Big(\varrho(u)\circ\sigma(u)\Big)\\
&\quad -\Big(\frac{1}{n}e_j^\top\mathbf{F}''1_n - \frac{1}{2n}\mathrm{V}_j^\top(\boldsymbol{\varrho}\circ\boldsymbol{\sigma})1_n\Big)\Big] + \frac{1}{n}\mathrm{V}_j^\top\Big[\boldsymbol{\epsilon}K_f(u) - \frac{1}{2}\varrho(u)\circ\Big(\boldsymbol{\sigma}_\xi(u)K * K_f(u)\Big)\Big]\\
&\quad +\delta_n\\
&= (\frac{1}{2}w_2^K h^2)A_1(u) + A_2(u) + \delta_n,
\end{aligned}$$

where

$$\begin{aligned}
A_1(u) &= V_j^\top\mathbf{W}_n(\lambda_j\mathbf{I} - \boldsymbol{\Lambda})^+\mathbf{V}_1^*\mathrm{F}(u) + F_j''(u) - \frac{1}{2}\mathrm{V}_j^\top\Big(\varrho(u)\circ\sigma(u)\Big)\\
&\qquad -\Big(\frac{1}{n}e_j^\top\mathbf{F}''1_n - \frac{1}{2n}\mathrm{V}_j^\top(\boldsymbol{\varrho}\circ\boldsymbol{\sigma})1_n\Big)\\
&= \frac{1}{n}\sum_{t=1}^{n}\Big[\sum_{i=1}^{p}F_i(U_t)F_j''(U_t)\mathrm{V}_i^\top + \sum_{i=1}^{p}F_j(U_t)F_i''(U_t)\mathrm{V}_i^\top\\
&-\frac{1}{2}\sum_{i=1}^{p}\mathrm{V}_j^\top(\varrho(U_t)\circ\sigma(U_t))F_i(U_t)\mathrm{V}_i^\top - \frac{1}{2}(\varrho^\top(U_t)\circ\sigma^\top(U_t))F_j(U_t)\Big](\lambda_j\mathbf{I} - \boldsymbol{\Lambda})^+\mathbf{V}_1^*\mathrm{F}(u)\\
&+\Big[F_j''(u) - \frac{1}{2}\mathrm{V}_j^\top\Big(\varrho(u)\circ\sigma(u)\Big) - \Big(\frac{1}{n}\sum_{t=1}^{n}F_j''(U_t) - \frac{1}{2n}\sum_{t=1}^{n}\mathrm{V}_j^\top(\varrho(U_t)\circ\sigma(U_t))\Big)\Big],\\
A_2(u) &= \frac{1}{n}\mathrm{V}_j^\top\Big[\boldsymbol{\epsilon}K_f(u) - \frac{1}{2}\varrho(u)\circ\Big(\boldsymbol{\sigma}_\xi(u)K * K_f(u)\Big)\Big]
\end{aligned}$$

$$= \frac{1}{nf_U(u)} \sum_{t=1}^{n} \Big[V_j^\top \tilde{\epsilon}_t K_h(U_t - u) - \frac{1}{2} \mathrm{V}_j^\top \Big(\varrho(u) \circ \tilde{\sigma}_{\xi,t}(u) \Big) K * K_h(U_t - u) \Big].$$

Then

$$\begin{aligned} EA_1(u) = {} & [F_j''(u) - EF_j''(U)] - \frac{1}{2} \mathrm{V}_j^\top \Big[\Big(\varrho(u) \circ \sigma(u) \Big) - E\Big(\varrho(U) \circ \sigma(U) \Big) \Big] \\ & + E\Big[F_i(U) F_j''(U) \mathrm{V}_i^\top + F_j(U) F_i''(U) \mathrm{V}_i^\top - \frac{1}{2} \mathrm{V}_j^\top (\varrho(U) \circ \sigma(U)) F_i(U) \mathrm{V}_i^\top \\ & - \frac{1}{2} (\varrho^\top(U) \circ \sigma^\top(U)) F_j(U) \Big] (\lambda_j \mathbf{I} - \mathbf{\Lambda})^+ \mathbf{V}_1^* \mathrm{F}(u), \\ EA_2(u) = {} & \frac{1}{nf_U(u)} \Big[V_j^\top \tilde{\epsilon}_1 K_h(U_1 - u) - \frac{1}{2} \mathrm{V}_j^\top \Big(\varrho(u) \circ \tilde{\sigma}_{\xi,1}(u) \Big) K * K_h(U_1 - u) \Big] = 0, \end{aligned}$$

and let

$$A_2(u) = \frac{1}{n} \sum_{t=1}^{n} R_t(u),$$

with

$$R_t(u) = \frac{1}{f_U(u)} \Big[V_j^\top \tilde{\epsilon}_t K_h(U_t - u) - \frac{1}{2} \mathrm{V}_j^\top \Big(\varrho(u) \circ \tilde{\sigma}_{\xi,t}(u) \Big) K * K_h(U_t - u) \Big].$$

Note that

$$\begin{aligned} \mathrm{Var}(R_1(u)) = {} & \frac{1}{hf_U(u)} V_j^\top \Big[\nu_K^2 \mathrm{Var}(\tilde{\epsilon}_1) + \frac{1}{4} \nu_{K*K}^2 \Big((\varrho(u) \varrho^\top(u)) \circ \mathrm{Var}(\tilde{\xi}_1) \Big) \\ & - \nu_{K,K*K} E\{ \tilde{\epsilon}_1 (\varrho^\top(u) \circ \tilde{\xi}_1^\top) \} \Big] V_j + o(\frac{1}{h}), \end{aligned}$$

by stationarity in (C3), we have

$$\mathrm{Var}(A_2(u)) = \frac{1}{n} \mathrm{Var}(R_1(u)) + \frac{2}{n} \sum_{s=1}^{n-1} (1 - \frac{s}{n}) \mathrm{Cov}(R_1(u), R_{s+1}(u)).$$

Let $d_n \to \infty$ be a sequence of integers such that $hd_n \to 0$. Define

$$Q_1 = \sum_{s=1}^{d_n-1} |\text{Cov}(R_1(u), R_{s+1}(u))|, \quad Q_2 = \sum_{s=d_n}^{n-1} |\text{Cov}(R_1(u), R_{s+1}(u))|.$$

Conditioning on (U_1, U_{s+1}), and by (C2), (C4) and (C5),

$$\begin{aligned}
&|\text{Cov}(R_1(u), R_{s+1}(u))| \\
&= |E\{E(R_1(u), R_{s+1}(u)|U_1, U_{s+1})\}| \\
&= \Big|E\Big\{E\Big(\frac{1}{f^2(u)}\Big[V_j^\top \tilde{\epsilon}_1 K_h(U_1-u) - \frac{1}{2}\mathrm{V}_j^\top\Big(\varrho(u)\circ\tilde{\sigma}_{\xi,1}(u)\Big)K*K_h(U_1-u)\Big] \\
&\quad \Big[V_j^\top \tilde{\epsilon}_t K_h(U_{s+1}-u) - \frac{1}{2}\mathrm{V}_j^\top\Big(\varrho(u)\circ\tilde{\sigma}_{\xi,s+1}(u)\Big)K*K_h(U_{s+1}-u)\Big]\Big)\Big|U_1, U_{s+1}\Big\}\Big| \\
&\le CL_2 \le M_0
\end{aligned}$$

for $M_0 > 0$ and some generic constant $C > 0$. Then it follows that $Q_1 \le d_n M_0$. We now consider the contribution of Q_2. For this α-mixing process, by Davydov's lemma,

$$|\text{Cov}(R_1(u), R_{s+1}(u))| = E|R_1(u)R_{s+1}(u)| \le 8[\alpha(s)]^{\frac{\delta}{1+\delta}}\{E|R_1|^{2(1+\delta)}\}^{\frac{1}{1+\delta}}.$$

By conditioning on U_1, and using (C2) and (C3),

$$\begin{aligned}
E|R_1|^{2(1+\delta)} &= E\left|\frac{V_j^\top \tilde{\epsilon}_1 K_h(U_1-u) - \frac{1}{2}\mathrm{V}_j^\top\Big(\varrho(u)\circ\tilde{\sigma}_{\xi,1}(u)\Big)K*K_h(U_1-u)}{f_U(u)}\right|^{2(1+\delta)} \\
&\le CL_2 h^{-2(1+\delta)+1}.
\end{aligned}$$

Hence, for $\frac{\delta}{1+\delta} < \gamma < 1$,

$$Q_2 \le \sum_{s=d_n}^{n-1} 8[\alpha(s)]^{\frac{\delta}{1+\delta}}\{E|R_1|^{2(1+\delta)}\}^{\frac{1}{1+\delta}} \le (CL_2)^{\frac{1}{1+\delta}} 8(h^{-2(1+\delta)+1})^{\frac{1}{1+\delta}} \sum_{s=d_n}^{\infty} [s^{-\beta}]^{\frac{\delta}{1+\delta}}$$

$$\leq M_1 h^{-2+\frac{1}{1+\delta}} \sum_{s=d_n}^{\infty} s^{-2} = M_1 h^{-2+\frac{1}{1+\delta}} d_n^{-\gamma} \sum_{s=d_n}^{\infty} s^{-2+\gamma} = o(1/h)$$

by taking $h^{-1+\frac{1}{1+\delta}} d_n^{-\gamma} = 1$. Together with the above results,

$$\sum_{s=1}^{n-1} \mathrm{Cov}(R_1(u), R_{s+1}(u)) = o(1/h),$$

and

$$\begin{aligned}
&\mathrm{Var}(A_2(u)) \\
= &\frac{V_j^\top \Big[\nu_K^2 \mathrm{Var}(\tilde{\epsilon}_1) + \frac{1}{4}\nu_{K*K}^2 \Big((\varrho(u)\varrho^\top(u)) \circ \mathrm{Var}(\tilde{\xi}_1)\Big) - \nu_{K,K*K} E\{\tilde{\epsilon}_1(\varrho^\top(u) \circ \tilde{\xi}_1^\top)\}\Big] V_j}{nhf_U(u)} \\
&+o(\frac{1}{nh}).
\end{aligned}$$

Therefore, as $n \to \infty$, $h \to 0$, similar to other nonparametric estimators for strong mixing time series, the following asymptotic normality could be established,

$$\sqrt{nh}\left(\hat{F}_j(u) - F_j(u) - (\frac{1}{2} w_2^K h^2) E A_1(u)\right) \xrightarrow{d} N(0, \sigma_{F_j}^2),$$

where

$$\begin{aligned}
EA_1(u) = &[F_j''(u) - EF_j''(U)] - \frac{1}{2}\mathrm{V}_j^\top \Big[\Big(\varrho(u) \circ \sigma(u)\Big) - E\Big(\varrho(U) \circ \sigma(U)\Big)\Big] \\
&+E\Big[F_i(U)F_j''(U)\mathrm{V}_i^\top + F_j(U)F_i''(U)\mathrm{V}_i^\top - \frac{1}{2}\mathrm{V}_j^\top(\varrho(U) \circ \sigma(U))F_i(U)\mathrm{V}_i^\top \\
&-\frac{1}{2}(\varrho^\top(U) \circ \sigma^\top(U))F_j(U)\Big](\lambda_j - \mathbf{\Lambda})^+ \mathbf{V}_1^* \mathrm{F}(u), \\
\sigma_{F_j}^2 = &\frac{V_j^\top \Big[\nu_K^2 \mathrm{Var}(\tilde{\epsilon}_1) + \frac{1}{4}\nu_{K*K}^2 \Big((\varrho(u)\varrho^\top(u)) \circ \mathrm{Var}(\tilde{\xi}_1)\Big) - \nu_{K,K*K} E\{\tilde{\epsilon}_1(\varrho^\top(u) \circ \tilde{\xi}_1^\top)\}\Big] V_j}{f_U(u)}.
\end{aligned}$$

□

Proof of Theorem 2.4. Let

$$\hat{\tilde{\epsilon}}_t = \begin{pmatrix} \hat{\epsilon}_{1,2,t} \\ \vdots \\ \hat{\epsilon}_{1,m,t} \\ \hat{\epsilon}_{2,3,t} \\ \vdots \\ \hat{\epsilon}_{2,m,t} \\ \vdots \\ \hat{\epsilon}_{m-1,m,t} \end{pmatrix}, \ \hat{\tilde{\varepsilon}}_t = \begin{pmatrix} \hat{\varepsilon}_{1,t}\hat{\varepsilon}_{2,t} \\ \vdots \\ \hat{\varepsilon}_{1,t}\hat{\varepsilon}_{m,t} \\ \hat{\varepsilon}_{2,t}\hat{\varepsilon}_{3,t} \\ \vdots \\ \hat{\varepsilon}_{2,t}\hat{\varepsilon}_{m,t} \\ \vdots \\ \hat{\varepsilon}_{m-1,t}\hat{\varepsilon}_{m,t} \end{pmatrix}, \ \hat{\mathbf{B}} = (\hat{\mathrm{V}}_1, ..., \hat{\mathrm{V}}_p), \ \hat{\mathrm{F}}_{[1:p]}(U_t) = \begin{pmatrix} \hat{F}_1(U_t) \\ \vdots \\ \hat{F}_p(U_t) \end{pmatrix}.$$

Lemma 2.3. Let $\hat{p}$ be the minimizer of the information criteria defined in (2.20) among $0 \le p \le p_{max}$ with $p_{max} > p_0$ being a fixed search limit. Consider the cases that $p \le p_0$, under the regularity conditions given before, $\hat{\sigma}^2_{[p]} - \frac{1}{M}E\|\tilde{\epsilon}_1\|^2 \to \frac{1}{M}tr(\mathbf{D}^*_{[p+1:p_0]})$ in probability and $tr(\mathbf{D}^*_{[p+1:p_0]}) = 0$ for $p = p_0$.

Proof of Lemma 2.3. For $p \le p_0$,

$$\begin{aligned} M\hat{\sigma}^2_{[p]} &= \frac{1}{n}\sum_{t=1}^{n}\|\hat{\tilde{\varepsilon}}_t - \hat{a} - \hat{\mathbf{B}}\hat{\mathrm{F}}(U_t)\|^2 = \frac{1}{n}\sum_{t=1}^{n}\|\hat{\tilde{\varepsilon}}_t - \hat{a} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathrm{F}}_{[1:p]}(U_t)\|^2 \\ &= \frac{1}{n}\sum_{t=1}^{n}\|\hat{\tilde{\varepsilon}}_t - \hat{a} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathbf{V}}^{*\top}_{1,[p]}\hat{\mathrm{G}}(U_t)\|^2 \\ &= \frac{1}{n}\sum_{t=1}^{n}\|\hat{\tilde{\varepsilon}}_t - \hat{\varrho}(U_t) + (\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathbf{V}}^{*\top}_{1,[p]})(\hat{\varrho}(U_t) - \hat{a})\|^2. \end{aligned}$$

Define $M\sigma^{*2}_{[p]} = \frac{1}{n}\sum_{t=1}^{n}\|\tilde{\varepsilon}_t - a - \mathbf{V}^*_{1,[p]}\mathbf{V}^{*\top}_{1,[p]}\mathrm{G}(U_t)\|^2$, recall that $\varrho(U_t) = a + \mathrm{G}(U_t)$, $\mathrm{F}(U_t) = \mathrm{F}_{[1:p]}(U_t) = \mathrm{V}^{*\top}_{1,[p]}\mathrm{G}(U_t)$, $\mathbf{B} = \mathbf{V}^*_{1,[p]}$, and $\mathbf{V}^{*\top}_{1,[p]}\mathbf{V}^*_{1,[p]} = \mathbf{I}_p$, $\mathbf{V}^{*\top}_{1,[p]}\mathbf{V}^*_{1,[p+1:p_0]} = 0$, thus

$$M\sigma^{*2}_{[p]} = \frac{1}{n}\sum_{t=1}^{n}\|\tilde{\varepsilon}_t - a - \mathbf{V}^*_{1,[p]}\mathbf{V}^{*\top}_{1,[p]}\mathrm{G}(U_t)\|^2$$

$$= \frac{1}{n}\sum_{t=1}^{n}\|\tilde{\varepsilon}_t - \varrho(U_t) + \varrho(U_t) - a - \mathbf{V}^*_{1,[p]}\mathbf{V}^{*\top}_{1,[p]}\mathrm{G}(U_t)\|^2$$

$$= \frac{1}{n}\sum_{t=1}^{n}\|\tilde{\varepsilon}_t - \varrho(U_t) + \big(\mathbf{I} - \mathbf{V}^*_{1,[p]}\mathbf{V}^{*\top}_{1,[p]}\big)\big(\mathbf{V}^*_{1,[p]}\mathrm{F}_{[1:p]}(U_t) + \mathbf{V}^*_{1,[p+1:p_0]}\mathrm{F}_{[p+1:p_0]}(U_t)\big)\|^2$$

$$= \frac{1}{n}\sum_{t=1}^{n} \|\tilde{\epsilon}_t + \mathbf{V}^*_{1,[p+1:p_0]}\mathrm{F}_{[p+1:p_0]}(U_t))\|^2$$
$$= \frac{1}{n}\sum_{t=1}^{n} \Big[\tilde{\epsilon}_t^\top \tilde{\epsilon}_t + 2\mathrm{F}^\top_{[p+1:p_0]}(U_t)\mathbf{V}^{*\top}_{1,[p+1:p_0]}\tilde{\epsilon}_t + \mathrm{F}^\top_{[p+1:p_0]}(U_t)\mathrm{F}_{[p+1:p_0]}(U_t)\Big].$$

Therefore, by law of large numbers,

$$M\sigma^{*2}_{[p]} \to E\tilde{\epsilon}_1^\top \tilde{\epsilon}_1 + E(\mathrm{F}^\top_{[p+1:p_0]}(U_1)\mathrm{F}_{[p+1:p_0]}(U_1)) = E\tilde{\epsilon}_1^\top \tilde{\epsilon}_1 + tr(\mathbf{D}^*_{[p+1:p_0]}) \quad a.s.$$

Furthermore,

$$\begin{aligned}
&M(\hat{\sigma}^2_{[p]} - \sigma^{*2}_{[p]})\\
&= \frac{1}{n}\sum_{t=1}^{n} \Big[(\hat{\tilde{\varepsilon}}_t - \hat{\varrho}(U_t))^\top(\hat{\tilde{\varepsilon}}_t - \hat{\varrho}(U_t)) + 2\hat{\mathrm{G}}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathbf{V}}^{*\top}_{1,[p]})(\hat{\tilde{\epsilon}}_t - \hat{\varrho}(U_t))\\
&\quad + \hat{\mathrm{G}}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathbf{V}}^{*\top}_{1,[p]})\hat{\mathrm{G}}(U_t)\Big] - \frac{1}{n}\sum_{t=1}^{n}\Big[\tilde{\epsilon}_t^\top \tilde{\epsilon}_t + 2\mathrm{F}^\top_{[p+1:p_0]}(U_t)\mathbf{V}^{*\top}_{1,[p+1:p_0]}\tilde{\epsilon}_t\\
&\quad + \mathrm{F}^\top_{[p+1:p_0]}(U_t)\mathrm{F}_{[p+1:p_0]}(U_t)\Big]\\
&= \frac{1}{n}\sum_{t=1}^{n} \Big[(\hat{\tilde{\varepsilon}}_t - \hat{\varrho}(U_t))^\top(\hat{\tilde{\varepsilon}}_t - \hat{\varrho}(U_t)) - \tilde{\epsilon}_t^\top \tilde{\epsilon}_t + 2\hat{G}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathbf{V}}^{*\top}_{1,[p]})(\hat{\tilde{\epsilon}}_t - \hat{\varrho}(U_t))\\
&\quad - 2\mathrm{F}^\top_{[p+1:p_0]}(U_t)\mathbf{V}^{*\top}_{1,[p+1:p_0]}\tilde{\epsilon}_t + \hat{\mathrm{G}}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathbf{V}}^{*\top}_{1,[p]})\hat{\mathrm{G}}(U_t)\\
&\quad - \mathrm{F}^\top_{[p+1:p_0]}(U_t)\mathrm{F}_{[p+1:p_0]}(U_t)\Big]\\
&= O_p(h^2 + \left(\frac{\log n}{nh}\right)^{\frac{1}{2}}),
\end{aligned}$$

by the convergence results of $\hat{\tilde{\varepsilon}}_t$, $\hat{\varrho}(U_t)$, $\hat{\mathrm{G}}(U_t)$ and $\hat{\mathbf{V}}^*_{1,[p]}$, it means that

$$M(\hat{\sigma}^2_{[p]} - \sigma^{*2}_{[p]}) \to 0 \text{ in probability for } p \le p_0.$$

Hence, we could deduce that $\hat{\sigma}^2_{[p]} - \frac{1}{M}E\|\tilde{\epsilon}_1\|^2 \to \frac{1}{M}tr(\mathbf{D}^*_{[p+1:p_0]})$ in probability.

□

Lemma 2.4. For $p > p_0$, under the same regularity conditions, $\hat{\sigma}^2_{[p]} - \hat{\sigma}^2_{[p_0]} =$

$O_p(h^2 + \big(\frac{\log n}{nh}\big)^{\frac{1}{2}})$.

Proof of Lemma 2.4. For $p > p_0$,

$$M\hat{\sigma}^2_{[p]} = \frac{1}{n}\sum_{t=1}^{n} \|\hat{\hat{\varepsilon}}_t - \hat{a} - \hat{\mathbf{B}}\hat{\mathrm{F}}(U_t)\|^2 = \frac{1}{n}\sum_{t=1}^{n} \|\hat{\hat{\varepsilon}}_t - \hat{a} - \hat{\mathbf{V}}^*_{1,[p]}\hat{\mathrm{F}}_{[1:p]}(U_t)\|^2$$
$$= \frac{1}{n}\sum_{t=1}^{n} \|\hat{\hat{\varepsilon}}_t - \hat{\varrho}(U_t) + (\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p_0]}\hat{\mathbf{V}}^{*\top}_{1,[p_0]} - \hat{\mathbf{V}}^*_{1,[p_0+1:p]}\hat{\mathbf{V}}^{*\top}_{1,[p_0:p]})(\hat{\varrho}(U_t) - \hat{a})\|^2$$
$$= \frac{1}{n}\sum_{t=1}^{n} \Big[(\hat{\hat{\epsilon}}_t - \hat{\varrho}(U_t))^\top(\hat{\hat{\epsilon}}_t - \hat{\varrho}(U_t)) + 2\hat{\mathrm{G}}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p_0]}\hat{\mathbf{V}}^{*\top}_{1,[p_0]} - \hat{\mathbf{V}}^*_{1,[p_0+1:p]}$$
$$\hat{\mathbf{V}}^{*\top}_{1,[p_0+1:p]})(\hat{\hat{\epsilon}}_t - \hat{\varrho}(U_t)) + \hat{\mathrm{G}}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p_0]}\hat{\mathbf{V}}^{*\top}_{1,[p_0]} - \hat{\mathbf{V}}^*_{1,[p_0+1:p]}\hat{\mathbf{V}}^{*\top}_{1,[p_0+1:p]})\hat{\mathrm{G}}(U_t)\Big],$$

$$M\hat{\sigma}^2_{[p_0]} = \frac{1}{n}\sum_{t=1}^{n} \Big[(\hat{\hat{\epsilon}}_t - \hat{\varrho}(U_t))^\top(\hat{\hat{\epsilon}}_t - \hat{\varrho}(U_t)) + 2\hat{\mathrm{G}}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p_0]}\hat{\mathbf{V}}^{*\top}_{1,[p_0]})(\hat{\hat{\epsilon}}_t - \hat{\varrho}(U_t))$$
$$+\hat{\mathrm{G}}^\top(U_t)(\mathbf{I} - \hat{\mathbf{V}}^*_{1,[p_0]}\hat{\mathbf{V}}^{*\top}_{1,[p_0]})\hat{\mathrm{G}}(U_t)\Big],$$

together with $\mathrm{G}(U_t) = \mathbf{V}^*_{1,[p_0]}\mathrm{F}(U_t)$, $\mathbf{V}^{*\top}_{1,[p_0]}\mathbf{V}^*_{1,[p_0+1:p]} = \mathbf{0}$, then

$$\begin{aligned} M|\hat{\sigma}^2_{[p]} - \hat{\sigma}^2_{[p_0]}| &= \Big|\frac{1}{n}\sum_{t=1}^{n} 2\hat{\mathrm{G}}^\top(U_t)\hat{\mathbf{V}}^*_{1,[p_0+1:p]}\hat{\mathbf{V}}^{*\top}_{1,[p_0+1:p]})(\hat{\hat{\epsilon}}_t - \hat{\varrho}(U_t)) \\ &\quad +\frac{1}{n}\sum_{t=1}^{n} \hat{\mathrm{G}}^\top(U_t)\hat{\mathbf{V}}^*_{1,[p_0+1:p]}\hat{\mathbf{V}}^{*\top}_{1,[p_0+1:p]}\hat{\mathrm{G}}(U_t)\Big| \\ &= \Big|\frac{1}{n}\sum_{t=1}^{n} 2\mathrm{G}^\top(U_t)\mathbf{V}^*_{1,[p_0+1:p]}\mathbf{V}^{*\top}_{1,[p_0+1:p]})(\tilde{\epsilon}_t - \varrho(U_t)) \\ &\quad +\frac{1}{n}\sum_{t=1}^{n} \mathrm{G}^\top(U_t)\mathbf{V}^*_{1,[p_0+1:p]}\mathbf{V}^{*\top}_{1,[p_0+1:p]}\mathrm{G}(U_t) + O_p(h^2 + \Big(\frac{\log n}{nh}\Big)^{\frac{1}{2}})\Big| \\ &= O_p(h^2 + \Big(\frac{\log n}{nh}\Big)^{\frac{1}{2}}). \end{aligned}$$

□

Now we only discuss the consistency of $IC(p)$.

For $p < p_0$, by Lemma 2.3 and $g_n \xrightarrow{P} 0$,

$$\begin{aligned} IC(p) - IC(p_0) &= \log(\hat{\sigma}^2_{[p]}) - \log(\hat{\sigma}^2_{[p_0]}) + (p - p_0)g_n \\ &= \log\left(1 + \frac{\hat{\sigma}^2_{[p]} - \hat{\sigma}^2_{[p_0]}}{\hat{\sigma}^2_{[p_0]}}\right) + (p - p_0)g_n \\ &= \left(\frac{\hat{\sigma}^2_{[p]} - \hat{\sigma}^2_{[p_0]}}{\hat{\sigma}^2_{[p_0]}}\right)(1 + o(1)) + (p - p_0)g_n \\ &\xrightarrow{P} \frac{tr(\mathbf{D}^*_{[p+1:p_0]})}{E\|\tilde{\epsilon}_1\|^2} > 0. \end{aligned}$$

Then $IC(p) > IC(p_0)$ with probability tending to 1.

For $p > p_0$, by Lemma 2.4 and $g_n / \left(h^2 + \left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right) \xrightarrow{P} \infty$,

$$\begin{aligned} IC(p) - IC(p_0) &= \log(\hat{\sigma}^2_{[p]}) - \log(\hat{\sigma}^2_{[p_0]}) + (p - p_0)g_n \\ &= \left(\frac{\hat{\sigma}^2_{[p]} - \hat{\sigma}^2_{[p_0]}}{\hat{\sigma}^2_{[p_0]}}\right)(1 + o(1)) + (p - p_0)g_n \\ &= O_p(h^2 + \left(\frac{\log n}{nh}\right)^{\frac{1}{2}}) + (p - p_0)g_n > 0. \end{aligned}$$

Therefore, $\hat{p}$ which minimizes $IC(p)$ converge to p_0 with probability going to 1. □

CHAPTER 3

A Hybrid Estimation of Conditional Variance or Covariance Function

3.1 Introduction

The estimation of variance functions without estimation of the mean functions have been extensively discussed in literature. The mechanism behind is the cross-difference can itself remove the mean automatically. In particular, difference-based estimators for estimating constant variances have a long history. Here we present the definition of the difference scheme.

Definition 3.1. A difference scheme of order $m \in \mathbb{N}$ is a vector $d = (d_r)_{r=0,\cdots,m} \in \mathbb{R}^{m+1}$ satisfying

$$\sum_{r=0}^{m} d_r = 0, \qquad \sum_{r=0}^{m} d_r^2 = 1.$$

The original idea could date back to Von Neumann (1941). Later, Rice (1984) introduced a popular first-order difference-based estimator

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2, \tag{3.1}$$

with $d_0 = \frac{1}{2}$ and $d_1 = -\frac{1}{2}$. Hall, Kay and Titterinton (1990) provided the m-th order variance estimator

$$\hat{\sigma}^2 = \frac{1}{(n-m)} \sum_{i=1}^{n-m} \left(\sum_{r=0}^{m} d_r Y_{i+r} \right)^2, \tag{3.2}$$

based on optimal difference sequences $\{d_r : r = 0, \cdots, m\}$ minimizing the asymptotic MSE of this variance estimator. Recently, Müller, Schick and Wefelmeyer (2003) proposed a covariate-matched U-statistic for noise variances in nonparametric regression. Except for estimation under univariate setting, difference-based constant variance estimators also have been examined for multivariate case. A generalized difference scheme is defined as follows.

Definition 3.2. A generalized difference scheme is a vector $d = (d_r)_{r \in J}$, satisfying

$$\sum_{r \in J} d_r = 0, \qquad \sum_{r \in J} d_r^2 = 1,$$

where $J \subset \mathbb{Z}^p$ denotes a particular index set.

Munk, Bissantz, Wagner and Freitag (2005) investigated the influences of dimensionality and smoothness of $m(\mathbf{x})$ on the optimal convergence rate of the difference-based approach, and suggested a class of estimators with a generalized polynomial weighted difference scheme.

The difference sequence scheme have also been utilized for nonconstant variance function estimation. Brown and Levine (2007) constructed a sort of difference-based variance estimator of order m by applying local polynomial smoothing based on squared pseudo residual of order m. Moreover, based upon the square of first-order differences, Wang, Brown, Cai and Levine (2008) presented a kernel estimator and established the result that the performance of their estimator is much better compared with the residual-based estimator, while the conditional mean function is not smooth enough. For multivariate case, Cai, Levine and Wang (2009) extended the difference sequence approach described in Munk, Bissantz, Wagner and Freitag (2005) to multidimensional nonparametric regression models, and derived the minimax convergence rate of this estimator. However, for multidimensional $\mathbf{X}$, the construction of the cross-difference is not easy due to the index sets selection; see for example Munk, Bissantz, Wagner and Freitag (2005), Cai, Levine and Wang (2009). In this chapter, we propose a novel approach to the estimation of conditional variance function. We do not define the complex difference sequence scheme, but construct a local variance estimate based on kernel weighted squared differences at the first stage, then estimate variance functions by kernel smoothing. Therefore, our method is a combination of the techniques in kernel smoothing and difference-based approach.

The remainder of this chapter is organized as follows. Section 3.2 presents our new strategy for estimating the unknown conditional variance function (or conditional covariance function). In Section 3.3, we establish the asymptotic normality of our proposed estimator. We then perform a simulation study to make a comparison with two existing methods in Section 3.4. All the technical proofs are given in Section 3.6.

3.2 Estimation

In this chapter, we consider the nonparametric multivariate regression model

$$\begin{aligned} Y &= m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon, \\ Z &= g(\mathbf{X}) + \phi(\mathbf{X})\varepsilon, \end{aligned} \tag{3.3}$$

where $\mathbf{X}$ is p-dimensional covariate, Y and Z are scalar responses, and ε, is the error term with $E(\varepsilon|\mathbf{X}) = 0$, and $E(\varepsilon^2|\mathbf{X}) = 1$. In this section, we will present our estimation method for both the conditional variance function and conditional covariance function. Therefore, the estimation of

$$\sigma^2(\mathbf{x}) = \mathrm{Var}(Y|\mathbf{X} = \mathbf{x}), \quad \phi^2(\mathbf{x}) = \mathrm{Var}(Z|\mathbf{X} = \mathbf{x})$$

and

$$\sigma_{Y,Z}(\mathbf{x}) = \mathrm{Cov}(Y, Z|\mathbf{X} = \mathbf{x}) = E\Big((Y - m(\mathbf{X}))(Z - g(\mathbf{X}))|\mathbf{X} = \mathbf{x}\Big)$$

is the primary focus of this chapter.

3.2.1 The hybrid estimation of conditional variance function

In order to illustrate the estimation procedure of conditional variance function, we take the nonparametric regression model

$$Y_i = m(\mathbf{X}_i) + \sigma(\mathbf{X}_i)\varepsilon_i, \quad i = 1, ..., n, \tag{3.4}$$

as an example, where $\{(\mathbf{X}_i, Y_i) : i = 1, \cdots, n\}$ have the same distribution as $(\mathbf{X}, Y)$ and $E(\varepsilon_i|\mathbf{X}_i) = 0$, $\mathrm{Var}(\varepsilon_i^2|\mathbf{X}_i) = 1$. It is noteworthy that both the mean function $m(\cdot)$ and the variance function $\sigma^2(\cdot)$ are unknown, and the estimation of conditional variance function $\sigma^2(\cdot)$ is of our interest. Consider the unconditional variance $\sigma_0^2 = E(Y - E(Y))^2$, we could observe that its estimate is usually done by its sample version, i.e.,

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \bar{Y}\}^2.$$

where $\bar{Y}$ is the sample mean. However, it is not easy to see that $\hat{\sigma}_0^2$ can also be written as

$$\hat{\sigma}_0^2 = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{Y_i - Y_j\}^2. \tag{3.5}$$

As a result, this enlightens us to propose a new estimator based on a local version of (3.5).

Without estimating the conditional mean $m(\mathbf{x})$ at first, our conditional variance estimator $\hat{\sigma}^2(\mathbf{x})$ is described as follows,

$$\hat{\sigma}^2(\mathbf{x}) = \frac{1}{2}\frac{\sum_{i=1}^{n} K_h(\mathbf{X}_i - \mathbf{x}) s_{i,b}^2}{\sum_{i=1}^{n} K_h(\mathbf{X}_i - \mathbf{x})},$$

with (3.6)

$$s_{i,b}^2 = \frac{\sum_{j=1}^{n} \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)(Y_i - Y_j)^2}{\sum_{j=1}^{n} \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)},$$

where $K(\cdot)$ and $\mathcal{K}(\cdot)$ are kernel functions, h and b are two bandwidths, $K_h(\cdot) = h^{-p}K(\cdot/h)$, and $\mathcal{K}_b(\cdot) = b^{-p}\mathcal{K}(\cdot/b)$. Alternatively, we could define

$\hat{\sigma}^2(\mathbf{x})$ as

$$\hat{\sigma}^2(\mathbf{x}) = \frac{1}{2}\frac{\sum_{i=1}^n W_{n,h}(\mathbf{X}_i - \mathbf{x}) s_{i,b}^2}{\sum_{i=1}^n W_{n,h}(\mathbf{X}_i - \mathbf{x})},$$

with (3.7)

$$s_{i,b}^2 = \frac{\sum_{j=1}^n \mathcal{W}_{n,b}(\mathbf{X}_i - \mathbf{X}_j)(Y_i - Y_j)^2}{\sum_{j=1}^n \mathcal{W}_{n,b}(\mathbf{X}_i - \mathbf{X}_j)},$$

where h and b are also bandwidths,

$$W_{n,h}(\mathbf{X}_i - \mathbf{x}) = w_{n,h,2} K_h(\mathbf{X}_i - \mathbf{x}) - w_{n,h,1} K_h(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x}),$$

$$\mathcal{W}_{n,b}(\mathbf{X}_i - \mathbf{x}) = \omega_{n,b,2} \mathcal{K}_b(\mathbf{X}_i - \mathbf{x}) - \omega_{n,b,1} \mathcal{K}_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x}),$$

and $w_{n,h,r} = \sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^r$, $\omega_{n,b,r} = \sum_{i=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^r$ for $r = 0, 1, 2$. The basic idea behind is that we first use $s_{i,b}^2$ to measure the local variation around Y_i, then apply kernel smoothing on those $s_{i,b}^2$. Note that the reason we define $s_{i,b}^2$ in this way is that we could quantify local variation relying on nearest neighbours without constructing a complicated difference sequence scheme.

3.2.2 The hybrid estimation of conditional covariance function

Due to the fact that

$$\sigma_{Y,Z}(\mathbf{x}) = \mathrm{Cov}(Y, Z|\mathbf{X} = \mathbf{x}) = \frac{1}{4}\left[\mathrm{Var}(Y + Z|\mathbf{X} = \mathbf{x}) - \mathrm{Var}(Y - Z|\mathbf{X} = \mathbf{x})\right],$$

we could estimate the conditional covariance function $\sigma_{Y,Z}(\mathbf{x})$ by separately estimating conditional variances of $Y + Z$ and $Y - Z$. Let $\sigma_+^2(\mathbf{x}) = \mathrm{Var}(Y + Z|\mathbf{X} = \mathbf{x})$, $\sigma_-^2(\mathbf{x}) = \mathrm{Var}(Y - Z|\mathbf{X} = \mathbf{x})$, and denote the nonparametric estimator for $\sigma_+^2(\mathbf{x})$ and $\sigma_-^2(\mathbf{x})$ by $\hat{\sigma}_+^2(\mathbf{x})$ and $\hat{\sigma}_-^2(\mathbf{x})$ respectively. Therefore, by applying the same estimation strategy for conditional variance function, our nonparametric estimator $\hat{\sigma}_{Y,Z}(\mathbf{x})$ for $\sigma_{Y,Z}(\mathbf{x})$ could be defined based on $\hat{\sigma}_+^2(\mathbf{x})$ and $\hat{\sigma}_-^2(\mathbf{x})$ accordingly. Similar to (3.6) and (3.7), we define $\hat{\sigma}_+^2(\mathbf{x})$ and $\hat{\sigma}_-^2(\mathbf{x})$ as follows,

$$\hat{\sigma}_+^2(\mathbf{x}) = \frac{1}{2}\frac{\sum_{i=1}^n K_{h_+}(\mathbf{X}_i - \mathbf{x})v_{i,b_+}^2}{\sum_{i=1}^n K_{h_+}(\mathbf{X}_i - \mathbf{x})}, \quad \hat{\sigma}_-^2(\mathbf{x}) = \frac{1}{2}\frac{\sum_{i=1}^n K_{h_-}(\mathbf{X}_i - \mathbf{x})v_{i,b_-}^2}{\sum_{i=1}^n K_{h_-}(\mathbf{X}_i - \mathbf{x})}, \tag{3.8}$$

where

$$\begin{aligned} v_{i,b_+}^2 &= \frac{\sum_{j=1}^n \mathcal{K}_{b_+}(\mathbf{X}_i - \mathbf{X}_j)[(Y_i + Z_i) - (Y_j + Z_j)]^2}{\sum_{j=1}^n \mathcal{K}_{b_+}(\mathbf{X}_i - \mathbf{X}_j)}, \\ v_{i,b_-}^2 &= \frac{\sum_{j=1}^n \mathcal{K}_{b_-}(\mathbf{X}_i - \mathbf{X}_j)[(Y_i - Z_i) - (Y_j - Z_j)]^2}{\sum_{j=1}^n \mathcal{K}_{b_-}(\mathbf{X}_i - \mathbf{X}_j)}; \end{aligned}$$

or,

$$\hat{\sigma}_+^2(\mathbf{x}) = \frac{1}{2}\frac{\sum_{i=1}^n W_{n,h_+}(\mathbf{X}_i - \mathbf{x})v_{i,b_+}^2}{\sum_{i=1}^n W_{n,h_+}(\mathbf{X}_i - \mathbf{x})}, \quad \hat{\sigma}_-^2(\mathbf{x}) = \frac{1}{2}\frac{\sum_{i=1}^n W_{n,h_-}(\mathbf{X}_i - \mathbf{x})v_{i,b_-}^2}{\sum_{i=1}^n W_{n,h_-}(\mathbf{X}_i - \mathbf{x})}, \tag{3.9}$$

where

$$\begin{aligned} v_{i,b_+}^2 &= \frac{\sum_{j=1}^n \mathcal{W}_{n,b_+}(\mathbf{X}_i - \mathbf{X}_j)[(Y_i + Z_i) - (Y_j + Z_j)]^2}{\sum_{j=1}^n \mathcal{W}_{n,b_+}(\mathbf{X}_i - \mathbf{X}_j)}, \\ v_{i,b_-}^2 &= \frac{\sum_{j=1}^n \mathcal{W}_{n,b_-}(\mathbf{X}_i - \mathbf{X}_j)[(Y_i - Z_i) - (Y_j - Z_j)]^2}{\sum_{j=1}^n \mathcal{W}_{n,b_-}(\mathbf{X}_i - \mathbf{X}_j)}, \end{aligned}$$

with functions $K_{b_+}(\cdot)$, $K_{b_-}(\cdot)$, $\mathcal{K}_{b_+}(\cdot)$, $\mathcal{K}_{b_-}(\cdot)$, $W_{n,b_+}(\cdot)$, $W_{n,b_-}(\cdot)$, $\mathcal{W}_{n,b_+}(\cdot)$, $\mathcal{W}_{n,b_-}(\cdot)$, and bandwidths b_+, b_-, h_+, h_- defined similar to those in (3.6) and (3.7). Subsequently, our estimator for the conditional covariance function

$\sigma_{Y,Z}(\mathbf{x})$ is

$$\hat{\sigma}_{Y,Z}(\mathbf{x}) = \frac{1}{4}[\hat{\sigma}_{+}^{2}(\mathbf{x}) - \hat{\sigma}_{-}^{2}(\mathbf{x})].$$

3.3 Asymptotics

In this section, we will discuss the asymptotic properties of our proposed estimators for conditional variance and conditional covariance function. We will only present the results for estimators in the form of (3.6) and (3.9), and results for estimators of other forms could be derived similarly.

For the nonparametric conditional variance estimator $\hat{\sigma}^2(\mathbf{x})$, we could establish the following asymptotic result. We only present here the asymptotic normality for $\hat{\sigma}^2(\mathbf{x})$, and results for $\hat{\phi}^2(\mathbf{x})$ is similar.

Theorem 3.1. Suppose that the regularity conditions (C1)-(C6) in Section 3.6 hold, then for a particular $\mathbf{x}$, as $n \to \infty$, we have

$$\sqrt{nh^p}\{\hat{\sigma}^2(\mathbf{x}) - \sigma^2(\mathbf{x}) - B(\mathbf{x})\} \xrightarrow{d} N\left(0, \frac{\sigma^4(\mathbf{x})\mathrm{Var}(\varepsilon_1^2)}{4f(\mathbf{x})}\nu_{K,K}^2\right), \qquad (3.10)$$

where

$$\begin{aligned} B(\mathbf{x}) = &\frac{1}{2}h^2 \mathrm{tr}\left\{\frac{2\mathbf{M}_2^K \nabla\sigma^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})} + \mathbf{M}_2^K \mathbf{S}(\mathbf{x})\right\} + \frac{1}{2}b^2 \mathrm{tr}\Big\{\mathbf{M}_2^K \nabla m(\mathbf{x})(\nabla m(\mathbf{x}))^\top \\ &+\mathbf{M}_2^K \nabla\sigma^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top + \mathbf{M}_2^K \mathbf{S}(\mathbf{x})\Big\}, \\ \nu_{K,K}^2 = &\int \left[\int \Big(K(\mathbf{v}) + K(\mathbf{v} + \frac{b\mathbf{u}}{h})\Big) K(\mathbf{u}) d\mathbf{u}\right]^2 d\mathbf{v}, \end{aligned}$$

with $\mathbf{M}_2^K = \int K(\mathbf{v})\mathbf{v}\mathbf{v}^\top d\mathbf{v}$, $\nabla m(\mathbf{x}) = \partial m(\mathbf{x})/\partial \mathbf{x}$, $\nabla\sigma^2(\mathbf{x}) = \partial\sigma^2(\mathbf{x})/\partial\mathbf{x}$, $\mathbf{H}(\mathbf{x}) = \partial^2 m(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}$, and $\mathbf{S}(\mathbf{x}) = \partial^2\sigma^2(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}$.

Remark 3.1. It can be observed from (3.10) that if $b = o(h)$,

$$\begin{aligned} B(\mathbf{x}) &= \frac{1}{2}h^2 \text{tr}\left\{ \frac{2\mathbf{M}_2^K \nabla\sigma^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})} + \mathbf{M}_2^K \mathbf{S}(\mathbf{x}) \right\} + o(h^2), \\ \nu_{K,K}^2 &= \int \left[\int \left(K(\mathbf{v}) + K(\mathbf{v} + \frac{b\mathbf{u}}{h}) \right) K(\mathbf{u}) d\mathbf{u} \right]^2 d\mathbf{v} = 4 \int K^2(v)\ d\mathbf{v} + o(1). \end{aligned}$$

This indicates that our estimator $\hat{\sigma}^2(\mathbf{x})$ has the same order of asymptotic bias, and the same asymptotic variance as the residual-based estimator proposed in Fan and Yao (1998), when the bandwidth b is of smaller order than bandwidth h. Therefore, our newly developed estimator $\hat{\sigma}^2(\mathbf{x})$ could achieve the same asymptotic efficiency as the estimator applied in Fan and Yao (1998), if we choose the bandwidth b very small compared with h.

Next, we could also study the asymptotic property of $\hat{\sigma}_{Y,Z}(\mathbf{x})$ in a similar way. Recall that

$$\begin{aligned} \sigma_+^2(\mathbf{x}) &= \text{Var}(Y + Z|\mathbf{X} = \mathbf{x}) = \sigma^2(\mathbf{x}) + \phi^2(\mathbf{x}) + 2\sigma_{Y,Z}(\mathbf{x}), \\ \sigma_-^2(\mathbf{x}) &= \text{Var}(Y - Z|\mathbf{X} = \mathbf{x}) = \sigma^2(\mathbf{x}) + \phi^2(\mathbf{x}) - 2\sigma_{Y,Z}(\mathbf{x}), \end{aligned}$$

and let

$$\begin{aligned} \mu_+(\mathbf{x}) = m(\mathbf{x}) + g(\mathbf{x}), &\quad \mu_-(\mathbf{x}) = m(\mathbf{x}) - g(\mathbf{x}), \\ \nabla\mu_+(\mathbf{x}) = \partial\mu_+(\mathbf{x})/\partial\mathbf{x}, &\quad \nabla\mu_-(\mathbf{x}) = \partial\mu_-(\mathbf{x})/\partial\mathbf{x}, \\ \nabla\sigma_+^2(\mathbf{x}) = \partial\sigma_+^2(\mathbf{x})/\partial\mathbf{x}, &\quad \nabla\sigma_-^2(\mathbf{x}) = \partial\sigma_-^2(\mathbf{x})/\partial\mathbf{x}, \\ \mathbf{H}_+(\mathbf{x}) = \partial^2\mu_+(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}, &\quad \mathbf{H}_-(\mathbf{x}) = \partial^2\mu_-(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}, \\ \mathbf{S}_+(\mathbf{x}) = \partial^2\sigma_+^2(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}, &\quad \mathbf{S}_-(\mathbf{x}) = \partial^2\sigma_-^2(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}. \end{aligned}$$

based on the definitions of estimators $\hat{\sigma}_+^2(\mathbf{x})$, $\hat{\sigma}_-^2(\mathbf{x})$ and $\hat{\sigma}_{Y,Z}(\mathbf{x})$, we could obtain the following asymptotic results.

Theorem 3.2. Let $b_+ = o(h_+)$, $b_- = o(h_-)$, and $h_- = O(h_+)$. Under the regularity conditions (C1)-(C6) in section 3.6 with h and b replaced by h_+, h_- and b_+, b_-, then for a particular $\mathbf{x}$, as $n \to \infty$, we have

$$\sqrt{nh_-^p}\{\hat{\sigma}_{Y,Z}(\mathbf{x}) - \sigma_{Y,Z}(\mathbf{x}) - B_{Y,Z}(\mathbf{x})\} \xrightarrow{d} N\left(0, \frac{\gamma_{cov}^2(\mathbf{x})\nu_{K,K,+}^2 + \frac{\gamma^2(\mathbf{x})\nu_{K,K,-}^2}{4} - \frac{\theta(\mathbf{x})\nu_{K,K,2}}{2}}{4f(\mathbf{x})}\right), \quad (3.11)$$

where

$$\begin{aligned}
B_{Y,Z}(\mathbf{x}) &= \frac{1}{8}h_+^2 \mathrm{tr}\left\{\frac{2\mathbf{M}_2^K \nabla\sigma_+^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})} + \mathbf{M}_2^K \mathbf{S}_+(\mathbf{x})\right\} \\
&\quad - \frac{1}{8}h_-^2 \mathrm{tr}\left\{\frac{2\mathbf{M}_2^K \nabla\sigma_-^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})} + \mathbf{M}_2^K \mathbf{S}_-(\mathbf{x})\right\}, \\
\gamma^2(\mathbf{x}) &= E\Big\{\left[\sigma^2(\mathbf{X}_1)(\varepsilon_1^2 - 1) + \phi^2(\mathbf{X}_1)(\varepsilon_1^2 - 1)\right]^2 | \mathbf{X}_1 = \mathbf{x}\Big\}, \\
\gamma_{cov}^2(\mathbf{x}) &= E\Big\{\sigma^2(\mathbf{X}_1)\phi^2(\mathbf{X}_1)(\varepsilon_1^2 - 1)^2 | \mathbf{X}_1 = \mathbf{x}\Big\}, \\
\theta(\mathbf{x}) &= E\Big\{\left[\sigma^2(\mathbf{X}_1)(\varepsilon_1^2 - 1) + \phi^2(\mathbf{X}_1)(\varepsilon_1^2 - 1)\right]\left[\sigma(\mathbf{X}_1)\phi(\mathbf{X}_1)\right. \\
&\qquad \left.(\varepsilon_1^2 - 1)\right] | \mathbf{X}_1 = \mathbf{x}\Big\}, \\
\nu_{K,K,+}^2 &= \int \left[\frac{h_-^p}{h_+^p} K\left(\frac{h_-\mathbf{v}}{h_+}\right) + K(\mathbf{v})\right]^2 d\mathbf{v}, \\
\nu_{K,K,-}^2 &= \int \left[\frac{h_-^p}{h_+^p} K\left(\frac{h_-\mathbf{v}}{h_+}\right) - K(\mathbf{v})\right]^2 d\mathbf{v}, \\
\nu_{K,K,2} &= \int \left[\frac{h_-^{2p}}{h_+^{2p}} K^2(\frac{h_-\mathbf{v}}{h_+}) - K^2(\mathbf{v})\right] d\mathbf{v}.
\end{aligned}$$

Remark 3.2. Note that if the selected bandwidths $h_+ = h_-$, then as $n \to \infty$,

$$\sqrt{nh_-^p}\{\hat{\sigma}_{Y,Z}(\mathbf{x}) - \sigma_{Y,Z}(\mathbf{x}) - B_{Y,Z}(\mathbf{x})\} \xrightarrow{d} N\left(0, \frac{\gamma_{cov}^2(\mathbf{x})\nu_{K,K,+}^2}{4f(\mathbf{x})}\right),$$

where

$$B_{Y,Z}(\mathbf{x}) = \frac{1}{2}h_-^2 \operatorname{tr}\left\{\frac{2\mathbf{M}_2^K \nabla\sigma_{Y,Z}^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})} + \mathbf{M}_2^K \mathbf{S}_{Y,Z}(\mathbf{x})\right\},$$

$$\nabla\sigma_{Y,Z}(\mathbf{x}) = \partial\sigma_{Y,Z}(\mathbf{x})/\partial\mathbf{x}, \quad \mathbf{S}_{Y,Z}(\mathbf{x}) = \partial^2\sigma_{Y,Z}(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}.$$

3.4 Numerical Results

In this section, simulation studies for the finite sample performance of our method for estimating conditional variance and covariance functions will be presented. Particularly, the residual-based method of Fan and Yao (1998) and the difference-based method of Cai, Levine and Wang (2009) are also investigated for comparison.

3.4.1 A simulation study for conditional variance estimation

For the estimation problem of conditional variance function, we will mainly compare the performance of our variance estimator $\hat{\sigma}^2(\mathbf{x})$ with the residual-based local linear estimator $\hat{\sigma}_{FY}^2(\mathbf{x})$ of Fan and Yao (1998) and the difference-based estimator $\hat{\sigma}_{CLW}^2(\mathbf{x})$ of Cai, Levine and Wang (2009).

We simulate 500 random realizations of size n from the following model

$$Y_i = m(\mathbf{X}_i) + \sigma(\mathbf{X}_i)\varepsilon_i,$$

and we assume that $\varepsilon_i \sim N(0,1)$, $\mathbf{X}_i = (X_{1i}, \cdots, X_{pi})$ for $i = 1, \cdots, n$. In order to examine the performance of our estimator, we will conduct a simulation study for the cases $p = 1, 2, 3, 4, 5$ with $n = 200, 500, 1000$, and consider the following set of conditional mean functions and variance functions,

Model 1: $m_1(\mathbf{x}) = \frac{3}{4}\sin(\frac{5}{2}\pi x_1)$, $\sigma_1^2(\mathbf{x}) = 0.5$.

Model 2: $m_2(\mathbf{x}) = \frac{3}{4}\sin(5\pi x_1)$, $\sigma_1^2(\mathbf{x}) = 0.5$.

Model 3: $m_3(\mathbf{x}) = \frac{3}{4}\sin(10\pi x_1)$, $\sigma_1^2(\mathbf{x}) = 0.5$.

Model 4: $m_4(\mathbf{x}) = \frac{3}{2}\sin(\frac{5}{2}\pi x_1)$, $\sigma_1^2(\mathbf{x}) = 0.5$.

Model 5: $m_5(\mathbf{x}) = \frac{3}{2}\sin(5\pi x_1)$, $\sigma_1^2(\mathbf{x}) = 0.5$.

Model 6: $m_6(\mathbf{x}) = \frac{3}{2}\sin(10\pi x_1)$, $\sigma_1^2(\mathbf{x}) = 0.5$.

Model 7: $m_1(\mathbf{x}) = \frac{3}{4}\sin(\frac{5}{2}\pi x_1)$, $\sigma_2^2(\mathbf{x}) = x_p^2 + 0.5$.

Model 8: $m_2(\mathbf{x}) = \frac{3}{4}\sin(5\pi x_1)$, $\sigma_2^2(\mathbf{x}) = x_p^2 + 0.5$.

Model 9: $m_3(\mathbf{x}) = \frac{3}{4}\sin(10\pi x_1)$, $\sigma_2^2(\mathbf{x}) = x_p^2 + 0.5$.

Model 10: $m_4(\mathbf{x}) = \frac{3}{2}\sin(\frac{5}{2}\pi x_1)$, $\sigma_2^2(\mathbf{x}) = x_p^2 + 0.5$.

Model 11: $m_5(\mathbf{x}) = \frac{3}{2}\sin(5\pi x_1)$, $\sigma_2^2(\mathbf{x}) = x_p^2 + 0.5$.

Model 12: $m_6(\mathbf{x}) = \frac{3}{2}\sin(10\pi x_1)$, $\sigma_2^2(\mathbf{x}) = x_p^2 + 0.5$.

Model 13: $m_7(\mathbf{x}) = 0.2\sin(\pi x_p) + 0.4e^{-16x_p^2}$, $\sigma_3^2(\mathbf{x}) = (0.4e^{-2x_1^2} + 0.2)^2$.

Model 14: $m_8(\mathbf{x}) = \sin(\pi x_p) + 2e^{-16x_p^2}$, $\sigma_3^2(\mathbf{x}) = (0.4e^{-2x_1^2} + 0.2)^2$.

Model 15: $m_9(\mathbf{x}) = 2\sin(\pi x_p) + 4e^{-16x_p^2}$, $\sigma_3^2(\mathbf{x}) = (0.4e^{-2x_1^2} + 0.2)^2$.

Note that only three variance functions are being studied, while several different mean functions are investigated for each variance function. The variance functions $\sigma_1^2(\mathbf{x})$ in Model 1-6 are the same constant, $\sigma_2^2(\mathbf{x})$ in

Model 7-12 are the same quadratic functions in x_p-direction, and $\sigma_3^2(\mathbf{x})$ in Model 13-15 are functions in x_1-direction, which have been considered in Fan and Yao (1998). Meanwhile, to investigate the influence of oscillation of mean functions on the estimated variances, we assumed sinusoid functions for conditional means. Most importantly, the mean functions examined for each variance function are arranged based on the amplitude of oscillation in particular directions.

From the theoretical results in Theorem 3.1, if the selected bandwidth b is of smaller order than bandwidth h, then our variance estimator $\hat{\sigma}^2(\mathbf{x})$ has the same asymptotic efficiency as $\hat{\sigma}^2_{FY}(\mathbf{x})$. As a result, to choose appropriate bandwidth b is of great significance in the estimation process. In this case, instead of selecting a common bandwidth when computing $s^2_{i,b}$ $(i = 1, \cdots, n)$, we adopt the following method to obtain $s^2_{i,b}$. At the beginning, we could compute the distances between all $\mathbf{X}_i$ $(i = 1, \cdots, n)$ and their respective neighbours. For a particular $\mathbf{X}_i$, we define the distance between $\mathbf{X}_i$ and $\mathbf{X}_j$ $(j = 1, \cdots, n)$ as the L_2-norm of $\mathbf{X}_i - \mathbf{X}_j$ $(\|\mathbf{X}_i - \mathbf{X}_j\|)$, denoted by $d(i, j)$. Let $\mathbf{X}_{i_l}$ be the l-th nearest neighbour of $\mathbf{X}_i$, which means that the distance $d(i, i_l)$ is the l-th smallest among all $d(i, j)$ for $j = 1, \cdots, n$. Therefore, we employ a set of $d(i, i_l)$ with different values of l as the bandwidths to quantify local variation, and then compare the obtained conditional variance estimators for different l. For the sake of convenience, we will denote $d(i, i_l)$ by $d_i(l)$ hereafter, and let $\bar{d}^2(l) = \frac{1}{n}\sum_{i=1}^{n} d_i^2(l)$.

Our estimation procedure is described as follows. At the first stage, we estimate the local variance around a particular Y_i $(i = 1, \cdots, n)$ by using bandwidth $d_i(l)$ for $l = 2, \cdots, L$ (L large enough). In addition, we use

$$\begin{aligned} &\mathcal{K}\left(\frac{\mathbf{X}_i - \mathbf{X}_j}{d_i(l)}\right) \\ \propto\ &\left(1 - \frac{d^2(i,j)}{d_i^2(l)}\right) I(d(i,j) < d_i(l)) + 10\left(1 - \frac{d^2(i,j)}{\bar{d}^2(l)}\right) I(d(i,j) < \bar{d}(l)) \end{aligned}$$

as the kernel weight at $\mathbf{X}_j$ when estimating local variance around $(\mathbf{X}_i, Y_i)$, here $I(d(i,j) < d_i(l)) = 1$ if $d(i,j) < d_i(l)$, otherwise $I(d(i,j) < d_i(l)) = 0$; and the function $I(d(i,j) < \bar{d}(l))$ is defined similarly. Note that more weight is given for $\mathbf{X}_j$ when $\mathbf{X}_j$ is very close to $\mathbf{X}_i$. In this step, we compute the local variance estimator $s^2_{i,d_i(l)}$ for different values of l, which may contribute a lot to the following step. Next, to obtain the conditional variance estimator $\hat{\sigma}^2(\mathbf{x})$, we choose product kernels created from the normal kernel and employ bandwidth h selected by cross-validation. Because of using different bandwidths $d_i(l)$ in the first step, we could obtain a set of conditional variance estimators, denoted by $\hat{\sigma}^2(\mathbf{x}, l)$ $(l = 2, \cdots, L)$.

In order to obtain the ultimate conditional variance estimator, we need to determine the value l_0 such that $\hat{\sigma}^2(\mathbf{x}, l_0)$ could approximate the true conditional variance well in a sense. To this end, we first compute the standard deviation of $(\hat{\sigma}^2(\mathbf{X}_i, l-1), \hat{\sigma}^2(\mathbf{X}_i, l), \hat{\sigma}^2(\mathbf{X}_i, l+1))$ for this particular l and each $\mathbf{X}_i$, thereby we could observe the change of conditional variance estimator at all $(\mathbf{X}_i, Y_i)$, $i = 1, \cdots, n$, when the bandwidth at the first stage is slightly adjusted. Let $sd(i,l)$ be the standard deviation of $(\hat{\sigma}^2(\mathbf{X}_i, l-1), \hat{\sigma}^2(\mathbf{X}_i, l), \hat{\sigma}^2(\mathbf{X}_i, l+1))$, and denote the average of these standard deviations by $sd(l) = \frac{1}{n}\sum_{i=1}^n sd(i,l)$. Therefore, for different values of k, starting from $l = p+1$ for $l \in [p+1, \frac{L}{2}]$, if we could find an integer l^* such that $sd(l^*) < \min(sd(l^*+1), sd(l^*+2), \cdots, sd(l^*+k))$, then we select $l_0(k) = l^*$, or $l^* + 1$, or $l^* + 2$ by observing the values of $sd(l^* + 1), sd(l^* + 2), \cdots, sd(l^* + k + 1)$, and $sd(l^* + k + 2)$; otherwise, we may assign a suitable value for l_0, for example $\lceil \frac{L}{2} \rceil$, where $\lceil \frac{L}{2} \rceil$ is the smallest integer larger than $\frac{L}{2}$. Consequently, to observe the performance of our conditional variance estimator with different k, denoted by $\hat{\sigma}^2(\mathbf{x}, l_0(k))$, we compute the results for the cases $k = 2, 3, \cdots, 6$. Finally, our conditional variance estimator is constructed as $\hat{\sigma}^2(\mathbf{x}) = \min_{2 \le k \le 6} \hat{\sigma}^2(\mathbf{x}, l_0(k))$.

For comparison purposes, the same Model 1-16 are investigated by using residual-based method in Fan and Yao (1998) and difference-based method in Cai, Levine and Wang (2009) with difference scheme of order p. We consider sample sizes $n = 200, 500, 1000$, and a random design of the sample $\mathbf{X}_i$. In order to compare the performance of the three estimators, we further compute the discrete mean squared error (DMSE) of $\hat{\sigma}^2(\mathbf{x})$, $\hat{\sigma}^2_{FY}(\mathbf{x})$ and $\hat{\sigma}^2_{CLW}(\mathbf{x})$, denoted by $DMSE(\hat{\sigma}^2)$, $DMSE(\hat{\sigma}^2_{FY})$ and $DMSE(\hat{\sigma}^2_{CLW})$ respectively. Accordingly, the discrete mean squared error (DMSE) of these three estimators are defined as

$$\begin{aligned} DMSE(\hat{\sigma}^2) &= \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\sigma}^2(\mathbf{X}_i) - \sigma^2(\mathbf{X}_i)\right]^2, \\ DMSE(\hat{\sigma}^2_{FY}) &= \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\sigma}^2_{FY}(\mathbf{X}_i) - \sigma^2(\mathbf{X}_i)\right]^2, \\ DMSE(\hat{\sigma}^2_{CLW}) &= \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\sigma}^2_{CLW}(\mathbf{X}_i) - \sigma^2(\mathbf{X}_i)\right]^2. \end{aligned}$$

Therefore, we report the mean of $DMSE(\hat{\sigma}^2)$, $DMSE(\hat{\sigma}^2_{FY})$, $DMSE(\hat{\sigma}^2_{CLW})$, denoted by $MDMSE(\hat{\sigma}^2)$, $MDMSE(\hat{\sigma}^2_{FY})$, $MDMSE(\hat{\sigma}^2_{CLW})$, based on 500 simulations for comparison.

Table 3.1: Performance of conditional variance estimators for $p = 1$

n	Model	$m(\mathbf{x})$	$\sigma^2(\mathbf{x})$	$MDMSE(\hat{\sigma}^2)$	$MDMSE(\hat{\sigma}^2_{CLW})$	$MDMSE(\hat{\sigma}^2_{FY})$
200	1	m_1	σ_1^2	**0.0114**	0.0175	0.0135
	2	m_2		**0.0148**	0.0214	0.0177
	3	m_3		**0.0311**	0.0462	0.0346
	4	m_4		**0.0130**	0.0209	0.0162
	5	m_5		**0.0170**	0.0356	0.0182
	6	m_6		**0.0252**	0.2000	0.0399
	7	m_1	σ_2^2	**0.4851**	0.9323	0.6010
	8	m_2		**0.5462**	0.9615	0.7014
	9	m_3		**0.5444**	0.9204	0.7679
	10	m_4		**0.5212**	1.0041	0.6189
	11	m_5		**0.6430**	1.1640	0.7111
	12	m_6		**0.9109**	1.5801	1.0575
	13	m_7	σ_3^2	**0.0026**	0.0044	0.0033
	14	m_8		**0.0028**	0.0048	0.0036
	15	m_9		**0.0030**	0.0060	0.0040
500	1	m_1	σ_1^2	**0.0042**	0.0052	0.0044
	2	m_2		**0.0047**	0.0057	0.0061
	3	m_3		**0.0075**	0.0083	0.0135
	4	m_4		**0.0044**	0.0055	0.0050
	5	m_5		**0.0054**	0.0059	0.0059
	6	m_6		**0.0092**	0.0126	0.0108
	7	m_1	σ_2^2	**0.2323**	0.4077	0.2618
	8	m_2		**0.2413**	0.4152	0.2735
	9	m_3		**0.2709**	0.4259	0.3100
	10	m_4		**0.2316**	0.4130	0.2650
	11	m_5		**0.2578**	0.4332	0.2963
	12	m_6		**0.3382**	0.5274	0.4513
	13	m_7	σ_3^2	**0.0013**	0.0019	0.0014
	14	m_8		**0.0013**	0.0020	0.0016
	15	m_9		**0.0013**	0.0021	0.0016
1000	1	m_1	σ_1^2	0.0029	0.0027	**0.0021**
	2	m_2		**0.0025**	0.0028	0.0027
	3	m_3		**0.0032**	0.0033	0.0059
	4	m_4		0.0027	0.0027	**0.0022**
	5	m_5		**0.0024**	0.0027	0.0026
	6	m_6		0.0039	**0.0033**	0.0043
	7	m_1	σ_2^2	0.1369	0.1903	**0.1278**
	8	m_2		0.1389	0.1979	**0.1349**
	9	m_3		**0.1364**	0.1994	0.1410
	10	m_4		0.1381	0.1943	**0.1314**
	11	m_5		**0.1377**	0.2071	0.1466
	12	m_6		**0.1610**	0.2299	0.2056
	13	m_7	σ_3^2	**0.0008**	0.0010	0.0008
	14	m_8		**0.0008**	0.0010	0.0008
	15	m_9		**0.0008**	0.0010	0.0008

Table 3.2: Performance of conditional variance estimators for $p = 2$

n	Model	$m(\mathbf{x})$	$\sigma^2(\mathbf{x})$	$MDMSE(\hat{\sigma}^2)$	$MDMSE(\hat{\sigma}^2_{CLW})$	$MDMSE(\hat{\sigma}^2_{FY})$
200	1	m_1	σ_1^2	**0.0428**	0.0492	0.0632
	2	m_2		**0.0329**	0.1138	0.0976
	3	m_3		**0.0361**	0.1286	0.0972
	4	m_4		0.3297	0.3574	**0.0527**
	5	m_5		**0.5984**	1.2252	1.2510
	6	m_6		**0.6588**	1.4591	1.2665
	7	m_1	σ_2^2	**0.9989**	1.7368	1.3541
	8	m_2		**1.0011**	2.1181	1.3825
	9	m_3		**0.9864**	2.1002	1.3532
	10	m_4		**1.7410**	2.5037	2.2720
	11	m_5		**1.5757**	4.0130	3.0497
	12	m_6		**1.6313**	4.3611	3.0924
	13	m_7	σ_3^2	**0.0050**	0.0087	0.0061
	14	m_8		0.0129	0.0302	**0.0103**
	15	m_9		0.0409	0.2536	**0.0165**
500	1	m_1	σ_1^2	0.0164	**0.0157**	0.0202
	2	m_2		0.0399	**0.0362**	0.0863
	3	m_3		**0.0396**	0.0999	0.0859
	4	m_4		0.0602	0.0572	**0.0306**
	5	m_5		0.7777	0.5265	**0.1022**
	6	m_6		**0.7941**	1.3589	1.2519
	7	m_1	σ_2^2	**0.5966**	0.9692	0.7802
	8	m_2		**0.6121**	1.0326	0.8262
	9	m_3		**0.6277**	1.2634	0.8726
	10	m_4		0.9927	1.4304	**0.9411**
	11	m_5		**1.3888**	1.6348	2.3735
	12	m_6		**1.3803**	3.0961	2.4325
	13	m_7	σ_3^2	**0.0031**	0.0044	0.0034
	14	m_8		**0.0056**	0.0077	0.0058
	15	m_9		0.0105	0.0347	**0.0081**
1000	1	m_1	σ_1^2	**0.0037**	0.0057	0.0110
	2	m_2		0.0457	**0.0181**	0.0232
	3	m_3		**0.0413**	0.0641	0.0844
	4	m_4		**0.0097**	0.0134	0.0150
	5	m_5		0.5760	0.2124	**0.0591**
	6	m_6		**0.8023**	0.9086	1.2765
	7	m_1	σ_2^2	**0.4006**	0.6481	0.4010
	8	m_2		**0.4188**	0.6075	0.5491
	9	m_3		**0.4131**	0.7483	0.5445
	10	m_4		**0.4647**	0.7542	0.5389
	11	m_5		1.3853	1.1947	**0.9391**
	12	m_6		**1.1931**	1.8995	2.0324
	13	m_7	σ_3^2	**0.0020**	0.0026	0.0020
	14	m_8		**0.0027**	0.0034	0.0033
	15	m_9		0.0051	0.0092	**0.0047**

Table 3.3: Performance of conditional variance estimators for $p = 3$

n	Model	$m(\mathbf{x})$	$\sigma^2(\mathbf{x})$	$MDMSE(\hat{\sigma}^2)$	$MDMSE(\hat{\sigma}^2_{CLW})$	$MDMSE(\hat{\sigma}^2_{FY})$
200	1	m_1	σ_1^2	**0.0228**	0.1206	0.1014
	2	m_2		**0.0258**	0.1615	0.1071
	3	m_3		**0.0246**	0.1553	0.1041
	4	m_4		**0.2839**	1.0942	1.2129
	5	m_5		**0.3347**	1.5679	1.2821
	6	m_6		**0.3412**	1.5807	1.2863
	7	m_1	σ_2^2	**1.5367**	3.2758	1.9891
	8	m_2		**1.5690**	3.5507	2.0891
	9	m_3		**1.5458**	3.5723	2.0407
	10	m_4		**1.8297**	4.9991	3.6569
	11	m_5		**1.8898**	5.9557	3.8234
	12	m_6		**1.8956**	6.2148	3.8821
	13	m_7	σ_3^2	**0.0080**	0.0148	0.0095
	14	m_8		0.0513	0.1711	**0.0165**
	15	m_9		0.5824	2.4709	**0.0773**
500	1	m_1	σ_1^2	**0.0174**	0.0514	0.0858
	2	m_2		**0.0193**	0.1107	0.0909
	3	m_3		**0.0197**	0.1124	0.0897
	4	m_4		**0.5327**	0.5920	0.7617
	5	m_5		**0.5565**	1.3843	1.2689
	6	m_6		**0.5693**	1.3947	1.2637
	7	m_1	σ_2^2	**1.0863**	1.8944	1.2750
	8	m_2		**1.0912**	2.1377	1.3336
	9	m_3		**1.1016**	2.1341	1.3187
	10	m_4		**1.4148**	2.7029	2.8344
	11	m_5		**1.4788**	4.2447	3.0893
	12	m_6		**1.4970**	4.2369	3.0303
	13	m_7	σ_3^2	**0.0058**	0.0091	0.0060
	14	m_8		0.0283	0.0674	**0.0112**
	15	m_9		0.1640	0.8965	**0.0286**
1000	1	m_1	σ_1^2	**0.0192**	0.0333	0.0431
	2	m_2		**0.0181**	0.0805	0.0843
	3	m_3		**0.0186**	0.0962	0.0842
	4	m_4		0.5335	0.3980	**0.0471**
	5	m_5		**0.5986**	1.1097	1.2680
	6	m_6		**0.6160**	1.3300	1.2668
	7	m_1	σ_2^2	**0.8327**	1.2420	0.8816
	8	m_2		**0.8289**	1.4408	0.8961
	9	m_3		**0.8314**	1.4853	0.8985
	10	m_4		**1.2610**	1.9433	1.6739
	11	m_5		**1.2250**	2.9768	2.5112
	12	m_6		**1.2493**	3.3439	2.5205
	13	m_7	σ_3^2	0.0044	0.0062	**0.0042**
	14	m_8		0.0229	0.0312	**0.0088**
	15	m_9		0.0795	0.3701	**0.0163**

Table 3.4: Performance of conditional variance estimators for $p = 4$

n	Model	$m(\mathbf{x})$	$\sigma^2(\mathbf{x})$	$MDMSE(\hat{\sigma}^2)$	$MDMSE(\hat{\sigma}^2_{CLW})$	$MDMSE(\hat{\sigma}^2_{FY})$
200	1	m_1	σ_1^2	**0.0254**	0.1809	0.1157
	2	m_2		**0.0251**	0.1722	0.1232
	3	m_3		**0.0241**	0.1693	0.1145
	4	m_4		**0.1143**	1.6459	1.2592
	5	m_5		**0.1233**	1.5848	1.3293
	6	m_6		**0.1193**	1.5781	1.2715
	7	m_1	σ_2^2	**2.2042**	3.6416	2.5415
	8	m_2		**2.2128**	3.5563	2.5768
	9	m_3		**2.2074**	3.4584	2.5208
	10	m_4		**2.0660**	6.1529	4.3181
	11	m_5		**2.0913**	6.1030	4.5845
	12	m_6		**2.0794**	5.8240	4.2987
	13	m_7	σ_3^2	**0.0113**	0.0171	0.0120
	14	m_8		0.0867	0.4408	**0.0279**
	15	m_9		1.5112	6.8118	**0.2281**
500	1	m_1	σ_1^2	**0.0111**	0.1105	0.0924
	2	m_2		**0.0117**	0.1227	0.0963
	3	m_3		**0.0123**	0.1238	0.0962
	4	m_4		**0.2196**	1.2878	1.2383
	5	m_5		**0.2496**	1.4056	1.2798
	6	m_6		**0.2466**	1.4233	1.2729
	8	m_1	σ_2^2	**1.7345**	2.8007	1.9064
	9	m_2		**1.7246**	2.8585	1.9579
	10	m_3		**1.7266**	2.8745	1.9564
	11	m_4		**1.6882**	4.7443	3.5924
	12	m_5		**1.6949**	5.0090	3.6887
	13	m_6		**1.7164**	5.0945	3.7712
	14	m_7	σ_3^2	0.0091	0.0137	**0.0089**
	15	m_8		0.0674	0.2460	**0.0168**
	16	m_9		0.9129	3.7876	**0.0847**
1000	1	m_1	σ_1^2	**0.0073**	0.0747	0.0837
	2	m_3		**0.0067**	0.0975	0.0850
	3	m_4		**0.0072**	0.1013	0.0860
	4	m_5		**0.2883**	0.9650	1.2222
	5	m_6		**0.3197**	1.3256	1.2584
	6	m_7		**0.3315**	1.3511	1.2680
	7	m_1	σ_2^2	1.4646	1.8917	**1.3736**
	8	m_2		1.4493	1.9445	**1.3546**
	9	m_3		1.4550	2.0138	**1.3913**
	10	m_4		**1.4587**	3.3513	3.0218
	11	m_5		**1.4519**	3.8572	3.0308
	12	m_6		**1.4748**	4.0691	3.1315
	13	m_8	σ_3^2	0.0076	0.0104	**0.0068**
	14	m_9		0.0614	0.1452	**0.0135**
	15	m_{10}		0.6737	2.1888	**0.0428**

Table 3.5: Performance of conditional variance estimators for $p = 5$

n	Model	$m(\mathbf{x})$	$\sigma^2(\mathbf{x})$	$MDMSE(\hat{\sigma}^2)$	$MDMSE(\hat{\sigma}^2_{CLW})$	$MDMSE(\hat{\sigma}^2_{FY})$
200	1	m_1	σ_1^2	**0.0551**	0.2107	0.1343
	2	m_2		**0.0538**	0.2036	0.1318
	3	m_3		**0.0518**	0.1974	0.1297
	4	m_4		**0.0683**	1.7009	1.3031
	5	m_5		**0.0669**	1.7528	1.3467
	6	m_6		**0.0674**	1.6897	1.3107
	7	m_1	σ_2^2	**3.0061**	4.4049	3.2052
	8	m_2		**2.9846**	4.4441	3.1885
	9	m_3		**3.0004**	4.3137	3.0978
	10	m_4		**2.7058**	7.0634	5.1131
	11	m_5		**2.6253**	7.1578	5.1083
	12	m_6		**2.6287**	6.7993	4.9117
	13	m_7	σ_3^2	0.0150	0.0220	**0.0147**
	14	m_8		**0.0659**	0.6959	0.0810
	15	m_9		1.1080	10.6302	**0.7867**
500	1	m_1	σ_1^2	**0.0272**	0.1289	0.0964
	2	m_2		**0.0265**	0.1308	0.0994
	3	m_3		**0.0275**	0.1363	0.1021
	4	m_4		**0.0768**	1.4567	1.2520
	5	m_5		**0.0783**	1.4477	1.2809
	6	m_6		**0.0837**	1.4674	1.2945
	7	m_1	σ_2^2	2.4954	3.3306	**2.3558**
	8	m_2		2.4945	3.3315	**2.3797**
	9	m_3		2.4761	3.4566	**2.4436**
	10	m_4		**2.2329**	5.4682	4.0433
	11	m_5		**2.2176**	5.4968	4.1140
	12	m_6		**2.2079**	5.8292	4.2921
	13	m_7	σ_3^2	0.0806	0.5102	**0.0224**
	14	m_8		0.0128	0.0177	**0.0115**
	15	m_9		1.3894	7.8304	**0.1575**
1000	1	m_1	σ_1^2	**0.0183**	0.1049	0.0885
	2	m_3		**0.0167**	0.1070	0.0891
	3	m_4		**0.0174**	0.1090	0.0896
	4	m_5		**0.1133**	1.3064	1.2469
	5	m_6		**0.1271**	1.3627	1.2702
	6	m_7		**0.1309**	1.3769	1.2782
	7	m_2	σ_2^2	2.1991	2.7969	**2.0402**
	8	m_3		2.1864	2.7528	**2.0001**
	9	m_4		2.1870	2.7703	**2.0050**
	10	m_5		**1.9742**	4.9037	3.8248
	11	m_6		**1.9671**	4.8539	3.7687
	12	m_7		**1.9646**	4.9279	3.8135
	13	m_7	σ_3^2	0.0111	0.0153	**0.0091**
	14	m_8		0.0840	0.3736	**0.0176**
	15	m_9		1.3978	5.7884	**0.0839**

Tables 3.1-3.5 summarize the simulation results using three methods for $p = 1, 2, 3, 4, 5$ individually. For different sample sizes $n = 200, 500, 1000$, we present the $MDMSE(\hat{\sigma}^2)$, $MDMSE(\hat{\sigma}^2_{FY})$, $MDMSE(\hat{\sigma}^2_{CLW})$ for each model, and the minimum value of each row is marked in bold font. In Table 3.1, results for three conditional variance estimators are displayed for Model 1-16 when $p = 1$. In this case, the mean functions and variance functions are functions of x_1. From this table, it can be observed that our estimator $\hat{\sigma}^2(\mathbf{x})$ outperforms $\hat{\sigma}^2_{CLW}(\mathbf{x})$ and $\hat{\sigma}^2_{FY}(\mathbf{x})$ in general. The $MDMSE$s for Model 13-15 with $\sigma^2_3(\mathbf{x}) = (0.4e^{-2x_1^2} + 0.2)^2$ are the smallest, while the $MDMSE$s for Model 7-12 with $\sigma^2_2(\mathbf{x}) = x_1^2 + 0.5$ are the largest. Particularly, for the case that the real conditional variance function is constant $\sigma^2_1(\mathbf{x}) = 0.5$, our conditional variance estimator definitely performs better than other two estimators for the sample sizes $n = 200, 500$. However, the performance of these three estimators does not differ too much, when we increase the sample size to $n = 1000$. Additionally, by observing the results for larger sample sizes $n = 500, 1000$, the accuracy of $\hat{\sigma}^2_{CLW}(\mathbf{x})$ seems higher than that of $\hat{\sigma}^2_{FY}(\mathbf{x})$ for Model 3 and Model 6 with heavily oscillating mean functions m_3 and m_6, which is in accordance with the finding in Wang, Brown, Cai and Levine (2008). Here our conditional variance estimator $\hat{\sigma}^2(\mathbf{x})$ could further improve accuracy in general, except for several special cases. Next, for the case that the real conditional variance function is $\sigma^2_2(\mathbf{x}) = x_1^2 + 0.5$, our variance estimator also performs better than other two estimators in general, but surprisingly its performance is slightly worse than $\hat{\sigma}^2_{FY}(\mathbf{x})$ when $n = 1000$ for not strongly oscillating mean functions, such as m_1, m_2 and m_4. Finally, when the variance function considered is $\sigma^2_3(\mathbf{x}) = (0.4e^{-2x_1^2} + 0.2)^2$, the estimation errors are smallest among all considered models. In particular for sample sizes $n = 200, 500$, it could be easily seen from Table 3.1 that $DMSE(\hat{\sigma}^2)$ for Model 13-15 are the smallest, $DMSE(\hat{\sigma}^2_{CLW})$ for those three models are the largest, and the estimation errors increase as the amplitude of fluctuation becomes larger. When $n = 1000$, this phenomenon is not obviously observed but $\hat{\sigma}^2(\mathbf{x})$ and $\hat{\sigma}^2_{FY}(\mathbf{x})$ still possess higher accuracy. Therefore, we could conclude that not only the oscillation and smoothness may affect the performance of variance estimators, but also the variation pattern of the variance functions may have an influence on it.

A detailed description of simulation results for $p = 1$ has been provided above, whereas the results for models with multivariate mean and variance functions are also of great importance. For higher dimensions $p > 1$, the similar studies are also conducted based on three estimation approaches. Table 3.2 shows the comparison results for $p = 2$. It is apparent that the $MDMSE(\hat{\sigma}^2)$s are still smaller than $MDMSE(\hat{\sigma}^2_{CLW})$s and $MDMSE(\hat{\sigma}^2_{FY})$s generally, which indicates that our variance estimator are superior to other two estimators in most cases. Specifically, when the constant variance function $\sigma_1^2(\mathbf{x})$ is considered, our estimator $\hat{\sigma}^2(\mathbf{x})$ has significantly the smallest DMSEs especially for Model 3 and Model 6, considering all sample sizes ($n = 200, 500, 1000$). But for other models with variance function $\sigma_1^2(\mathbf{x})$, either $\hat{\sigma}^2_{CLW}(\mathbf{x})$ or $\hat{\sigma}^2_{FY}(\mathbf{x})$ may be superior to $\hat{\sigma}^2(\mathbf{x})$ by taking different sample sizes into account. Next, for models with variance function $\sigma_2^2(\mathbf{x})$, our estimator seems competitive except for two special cases ($n = 500$, Model 10; $n = 1000$, Model 11). However, when the variance function $\sigma_3^2(\mathbf{x})$ is examined, the MDMSEs of two difference-based estimators increase apparently as the the amplitude of fluctuation of mean functions grows. Particularly, it is worth noting that even though the $MDMSE(\hat{\sigma}^2)$s are significantly smaller than $MDMSE(\hat{\sigma}^2_{CLW})$s, they are larger than $MDMSE(\hat{\sigma}^2_{FY})$ when estimating Model 15 with mean function m_9.

For the case that $p = 3$, $DMSE(\hat{\sigma}^2)$s are the smallest among models considered with variance functions σ_1^2 and σ_2^2, therefore it may be reasonable to apply our proposed estimator to estimate conditional variances for similar cases. Unfortunately, the performance of our variance estimator is not very satisfactory for models with variance function σ_3^2, since $\hat{\sigma}^2_{FY}$ could obtain the smallest DMSEs especially for Model 14 and Model 15 with mean function m_8 and m_9. Moreover, the results for higher dimensions $p = 4, 5$ agree with previous findings for $p = 3$, except that $\hat{\sigma}^2_{FY}(\mathbf{x})$ may perform better than our $\hat{\sigma}^2(\mathbf{x})$ for Model 7-9 when considering larger sample sizes $n = 500, 1000$. This finding indicates that $\hat{\sigma}^2_{FY}(\mathbf{x})$ may be much more appropriate for estimating variance functions like σ_2^2 with not strongly oscillating mean functions.

3.4.2 A simulation study for conditional covariance estimation

For the estimation problem of conditional covariance function, we will also focus on comparing the performance of our conditional covariance estimator with those estimators by means of residual-based local linear smoothing and conventional difference-based method.

We simulate 500 random realizations of size n from the following model

$$\begin{aligned} Y_i &= m(\mathbf{X}_i) + 2(X_{1i} - 0.3)e_{0i} + e_{1i}, \\ Z_i &= g(\mathbf{X}_i) + 2(X_{1i} - 0.5)e_{0i} + e_{2i}, \end{aligned} \tag{3.12}$$

and we assume that e_{0i}, e_{1i}, e_{2i} are independent and generated from $N(0, 1)$,

$\mathbf{X}_i = (X_{1i}, \cdots, X_{pi})$ for $i = 1, \cdots, n$. Here we consider conditional mean functions $m(\mathbf{x}) = \Psi\cos(\tau\pi x_1)$ and $g(\mathbf{x}) = \Psi\sin(\tau\pi x_1)$, and then the true conditional covariance of Y and Z given $\mathbf{X} = \mathbf{x}$ is actually $\sigma_{Y,Z}(\mathbf{x}) = 4(x_1 - 0.3)(x_1 - 0.5)$.

Following the estimation procedure described in Section 3.2.2, we will also compare the performance of three types of conditional covariance estimators in a similar way, by investigating the situations with different values for Ψ and τ. Since our estimation approach is based on the estimated conditional variances of $Y + Z$ and $Y - Z$, the same method presented in Section 3.4.1 could be directly employed for estimating conditional covariance functions. In the following, we will study the influence of oscillation and fluctuation of mean functions on the accuracy of estimating conditional covariance functions. Therefore, analogous to Section 3.4.1, the three types of conditional covariance estimators are denoted by $\hat{\sigma}_{Y,Z}(\mathbf{x})$, $\hat{\sigma}_{Y,Z_{FY}}(\mathbf{x})$ and $\hat{\sigma}_{Y,Z_{CLW}}(\mathbf{x})$, and the accuracy of these estimators are measured by $MDMSE(\hat{\sigma}_{Y,Z})$, $MDMSE(\hat{\sigma}_{Y,Z_{FY}})$ and $MDMSE(\hat{\sigma}_{Y,Z_{CLW}})$ respectively.

Table 3.6 displays the simulation results for the conditional covariance estimators based on three estimation strategies. For the case that the true conditional covariance function $\sigma_{Y,Z}(\mathbf{x}) = 4(x_1 - 0.3)(x_1 - 0.5)$, our conditional covariance estimator outperforms other two types of covariance estimators. Compared with the estimator $\hat{\sigma}_{Y,Z_{FY}}(\mathbf{x})$ utilizing residual-based local linear smoothing of Fan and Yao (1998), our estimator $\hat{\sigma}_{Y,Z}(\mathbf{x})$ approximates much more accurately especially when the conditional mean functions $m(\mathbf{x})$ and $g(\mathbf{x})$ are highly oscillating. Meanwhile, our estimation approach seems slightly better than the conventional difference-based method in all

Table 3.6: Performance of conditional covariance estimators for $p = 1$

n	$\sigma_{Y,Z}(\mathbf{x})$	Ψ	τ	$MDMSE(\hat{\sigma}^2_{Y,Z})$	$MDMSE(\hat{\sigma}^2_{Y,Z_{CLW}})$	$MDMSE(\hat{\sigma}^2_{Y,Z_{FY}})$
100	$4(x_1-0.3)(x_1-0.5)$	0	-	**0.0965**	0.2039	0.1157
		1	1	**0.0966**	0.2045	0.1169
		1	5	**0.0979**	0.2086	0.1411
		1	10	**0.1074**	0.2077	0.1710
		1	20	**0.1522**	0.2274	0.1740
		4	1	**0.0969**	0.2071	0.1530
		4	5	**0.1007**	0.2486	0.8468
		4	10	**0.1185**	0.3049	1.9635
		4	20	**0.1368**	0.6413	2.0301
		10	1	**0.0981**	0.2167	0.9586
		10	5	**0.1094**	0.5455	16.9590
		10	10	**0.1315**	1.5478	46.3867
		10	20	**0.6343**	8.6348	47.5867
500	$4(x_1-0.3)(x_1-0.5)$	0	-	**0.0264**	0.0449	0.0268
		1	1	**0.0264**	0.0449	0.0269
		1	5	**0.0264**	0.0449	0.0314
		1	10	**0.0263**	0.0449	0.0365
		1	20	**0.0271**	0.0452	0.0418
		4	1	**0.0264**	0.0449	0.0299
		4	5	**0.0265**	0.0450	0.3670
		4	10	**0.0271**	0.0453	0.4643
		4	20	**0.0297**	0.0474	0.6028
		10	1	**0.0264**	0.0450	0.1266
		10	5	**0.0273**	0.0459	11.5837
		10	10	**0.0294**	0.0481	12.1740
		10	20	**0.0335**	0.0630	15.6445
1000	$4(x_1-0.3)(x_1-0.5)$	0	-	**0.0152**	0.0226	0.0153
		1	1	**0.0152**	0.0226	0.0153
		1	5	**0.0152**	0.0226	0.0175
		1	10	**0.0152**	0.0226	0.0209
		1	20	**0.0152**	0.0226	0.0247
		4	1	**0.0152**	0.0226	0.0164
		4	5	**0.0152**	0.0226	0.3445
		4	10	**0.0154**	0.0225	0.2259
		4	20	**0.0160**	0.0227	0.3925
		10	1	**0.0152**	0.0226	0.0553
		10	5	**0.0151**	0.0227	12.2780
		10	10	**0.0157**	0.0227	5.5099
		10	20	**0.0175**	0.0240	10.5522

situations. It is worth mentioning that the estimation errors of our estimator does not significantly increase when the amplitude of oscillation and fluctuation is enlarged, especially for larger sample sizes.

3.5 Discussion

This chapter presents a novel approach to modelling the conditional variance (or covariance) function by applying the techniques both in kernel smoothing and difference-based methods. Without estimating the conditional mean or constructing a complicated difference scheme, our proposed variance estimator $\hat{\sigma}^2(\mathbf{x})$ defined in (3.6) or (3.7) could possess desirable asymptotic properties and exhibit good performance in most cases. Specifically, it could be directly observed from Theorem 3.1 that, our conditional variance estimator $\hat{\sigma}^2(\mathbf{x})$ has the same asymptotic variance as the residual-based estimator $\hat{\sigma}^2_{FY}(\mathbf{x})$ when the bandwidth $b = o(h)$, which demonstrates the validity of our newly developed variance estimator from the theoretical point of view. More importantly, to understand the influence of oscillation and fluctuation of true functions on estimation accuracy, we have examined the estimators' performance for models with three conditional variance functions and different sinusoid conditional mean functions. Note that not only the effect of oscillation and fluctuation of the mean functions have been considered, variance functions of different forms are also studied. As illustrated in our simulation study, our proposed estimator outperforms other two estimators in most cases. Focusing on the results for larger sample sizes, our variance estimator performs slightly better than other two estimators for the true conditional variance function being constant especially in case of strongly oscillating mean functions when $p = 1, 2$, whereas our estimation strategy is absolutely superior to other two methods for higher dimensions

$p = 3, 4, 5$. For the true conditional variance function being a quadratic function in x_p direction, the performance of two existing estimators is hardly comparable to our estimator when the mean function is heavily oscillating and fluctuating, while the residual-based estimator may gain higher accuracy in case of not heavily oscillating and fluctuating mean functions. However, our estimator seems no longer competitive when modelling the variance function considered in Fan and Yao (1998), especially for mean functions with large amplitude of fluctuation.

3.6 Proofs

Before presenting the asymptotic results, we introduce the assumptions needed for our analysis. The following regularity conditions are assumed.

(C1) Let $f(\cdot)$ denote the marginal density of $\mathbf{X}_i$, and $f_\ell(\cdot,\cdot)$ denote the joint density of $(\mathbf{X}_i, \mathbf{X}_{i+\ell})$. Suppose that $f(\cdot)$ has a closed and bounded support, such as $D \in \mathbb{R}^p$, $f(\mathbf{x}) > 0$, and $|f(\mathbf{x}) - f(\mathbf{x}')| \le \Delta_1 \|\mathbf{x} - \mathbf{x}'\|$ for all given $\mathbf{x}, \mathbf{x}' \in D$ and some $\Delta_1 > 0$, also $f_\ell(\mathbf{x}_0, \mathbf{x}_\ell) > 0$ for $\mathbf{x}_0, \mathbf{x}_\ell \in D$. Meanwhile, $\sup_{\mathbf{x}\in D} f(\mathbf{x}) \le L_0 < \infty$, $\sup_{\mathbf{x}_0,\mathbf{x}_\ell \in D} f_\ell(\mathbf{x}_0, \mathbf{x}_\ell) \le L_0 < \infty$. Further, denote the gradient and Hessian matrix of $f(\mathbf{x})$ by $\nabla f(\mathbf{x})$ and $\mathbf{H}_f(\mathbf{x})$.

(C2) $E|Y_i|^{4(1+\delta)} \le L_2 < \infty$, $E|Z_i|^{4(1+\delta)} \le L_2 < \infty$, and $E(|Y_i|^{4(1+\delta)}|\mathbf{X}_i) \le L_2 < \infty$, $E(|Z_i|^{4(1+\delta)}|\mathbf{X}_i) \le L_2 < \infty$, $i = 1, ..., n$, for L_2 large enough and constant $\delta > 0$.

(C3) The process $\{(\mathbf{X}_i, Y_i, Z_i) : t = 1, ..., n\}$ is strictly stationary and strong mixing with mixing coefficient $\alpha(N) \leq CN^{-\beta}$ for some $C > 0$ and $\beta > 2 + \frac{2}{\delta}$ for the same δ as in (C2).

(C4) $m(\cdot)$, $g(\cdot)$, and $\sigma^2(\cdot)$, $\phi^2(\cdot)$ are differentiable, and the Hessian matrix of $m(\cdot)$, $g(\cdot)$ and $\sigma^2(\cdot)$, $\phi^2(\cdot)$ are uniformly continuous.

(C5) The continuous symmetric kernel function $K(\cdot)$ has the following properties, and the assumptions are also required for kernel $\mathcal{K}(\cdot)$.

(i) $\int |K(\mathbf{v})|d\mathbf{v} < \infty$, $\int K^2(\mathbf{v})d\mathbf{v} < \infty$, and $\int K(\mathbf{v})d\mathbf{v} = 1$, $\int \mathbf{v}K(\mathbf{v})d\mathbf{v} = 0$, $\int K(\mathbf{v})\mathbf{v}\mathbf{v}^\top d\mathbf{v} = \mathbf{M}_2^K$.

(ii) For a vector c_0 with all positive elements and a scalar Δ_2 with $0 < \Delta_2 < \infty$, either $K(\cdot)$ is a bounded function with a bounded support on $\mathbb{R}^p$, satisfying the Lipschitz condition, i.e. $|K(\mathbf{v}_1) - K(\mathbf{v}_2)| \leq \Delta_2\|\mathbf{v}_1 - \mathbf{v}_2\|$, or $K(\cdot)$ is differentiable, when $\|\mathbf{v}\| \to \infty$, $K(\mathbf{v})e^{c_0^\top \mathbf{v}} \to 0$.

(C6) As $n \to \infty$, $h \to 0$, $b \to 0$, such that $nh^{p+2} \to \infty$, $nb^{p+2} \to \infty$ and $b = o(h)$.

Lemma 3.1. Under the regularity conditions (C1)-(C6) in Section 3.6, for model (3.3) where $(\mathbf{X}_i, Y_i)$ is a strictly stationary time series, for a particular $\mathbf{x}$, we have uniformly for $\mathbf{x}^* \in D$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_b(\mathbf{X}_i - \mathbf{x}^*)Y_i = E\left\{\mathcal{K}_b(\mathbf{X}_1 - \mathbf{x}^*)Y_1\right\}\left(1 + O_p(\{\log n/(nb^p)\}^{1/2})\right), \quad (3.13)$$

$$\frac{1}{n}\sum_{i=1}^{n} K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{x}^*)Y_i = E\left\{K_h(\mathbf{X}_1 - \mathbf{x})\mathcal{K}_b(\mathbf{X}_1 - \mathbf{x}^*)Y_1\right\}\Big[1 + O_p(\{\log n/(nb^p)\}^{1/2})\Big], \quad (3.14)$$

The results in Lemma 3.1 could be derived similarly as those presented in Mack and Silverman (1982) and Fan and Yao (2003).

Proof of Theorem 3.1. From model (3.3), it is easily seen that we can write

$$(Y_i - Y_j)^2 = (m(\mathbf{X}_i) - m(\mathbf{X}_j))^2 + \sigma^2(\mathbf{X}_i) + \sigma^2(\mathbf{X}_j) + \eta_{ij}, \tag{3.15}$$

where

$$\begin{aligned}\eta_{ij} &= 2(m(\mathbf{X}_i) - m(\mathbf{X}_j))[\sigma(\mathbf{X}_i)\varepsilon_i - \sigma(\mathbf{X}_j)\varepsilon_j] - 2\sigma(\mathbf{X}_i)\sigma(\mathbf{X}_j)\varepsilon_i\varepsilon_j \\ &\quad + \sigma^2(\mathbf{X}_i)(\varepsilon_i^2 - 1) + \sigma^2(\mathbf{X}_j)(\varepsilon_j^2 - 1),\end{aligned}$$

with

$$E(\eta_{ij}|\mathbf{X}_s, s = 1, ..., n) = 0.$$

By Condition (C4) and Taylor's expansion to the second order, for $\mathbf{X}_i$ and $\mathbf{X}_j$ in the local neighbourhood of $\mathbf{x}$,

$$\begin{aligned}m(\mathbf{X}_i) - m(\mathbf{X}_j) &= (\mathbf{X}_i - \mathbf{x})^\top \nabla m(\mathbf{x}) - (\mathbf{X}_j - \mathbf{x})^\top \nabla m(\mathbf{x}) \\ &\quad + \frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) - \frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_j - \mathbf{x}) \\ &\quad + O(||\mathbf{X}_i - \mathbf{x}||^3 + ||\mathbf{X}_j - \mathbf{x}||^3),\end{aligned} \tag{3.16}$$

where $\nabla m(\mathbf{x}) = \partial m(\mathbf{x})/\partial \mathbf{x}, \mathbf{H}(\mathbf{x}) = \partial^2 m(\mathbf{x})/\partial \mathbf{x} \partial \mathbf{x}^\top$. Analogously, we have

$$\begin{aligned}\sigma^2(\mathbf{X}_i) + \sigma^2(\mathbf{X}_j) &= 2\sigma^2(\mathbf{x}) + (\mathbf{X}_i - \mathbf{x})^\top \nabla \sigma^2(\mathbf{x}) + (\mathbf{X}_j - \mathbf{x})^\top \nabla \sigma^2(\mathbf{x}) \\ &\quad + \frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) + \frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_j - \mathbf{x}) \\ &\quad + O(||\mathbf{X}_i - \mathbf{x}||^3 + ||\mathbf{X}_j - \mathbf{x}||^3),\end{aligned} \tag{3.17}$$

where $\nabla\sigma^2(\mathbf{x}) = \partial\sigma^2(\mathbf{x})/\partial\mathbf{x}, \mathbf{S}(\mathbf{x}) = \partial^2\sigma^2(\mathbf{x})/\partial\mathbf{x}\partial^\top\mathbf{x}$. By Condition (C1) and Taylor's expansion, for $\mathbf{X}_i$ in the neighbourhood of $\mathbf{x}$,

$$f(\mathbf{X}_i) = f(\mathbf{x}) + (\mathbf{X}_i - \mathbf{x})^\top \nabla f(\mathbf{x}) + \frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{H}_f(\mathbf{x})(\mathbf{X}_i - \mathbf{x}). \quad (3.18)$$

From (3.6), we could obtain that

$$\begin{aligned}
\hat{\sigma}^2(\mathbf{x}) - \sigma^2(\mathbf{x}) &= \frac{1}{2}\frac{\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})\left[\frac{\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)(Y_i - Y_j)^2}{\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)} - 2\sigma^2(\mathbf{x})\right]}{\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})}, \\
&= \frac{1}{2}\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})\left\{\frac{\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)\left[(Y_i - Y_j)^2 - 2\sigma^2(\mathbf{x})\right]}{\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})}\right\} \\
&= \frac{1}{2}\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})\left\{\frac{\frac{1}{n^2}\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)\left[(Y_i - Y_j)^2 - 2\sigma^2(\mathbf{x})\right]}{\frac{1}{n}\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)\frac{1}{n}\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})}\right\},
\end{aligned}$$

where $K_h(.) = h^{-p}K(./h)$, $\mathcal{K}_b(.) = b^{-p}K(./b)$, then by Condition (C5)(i) and Lemma 3.1,

$$\frac{1}{n}\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x}) = f(\mathbf{x})(1 + O(h^2) + O_p\left(\sqrt{\frac{logn}{nh^p}}\right)),$$

$$\frac{1}{n}\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j) = f(\mathbf{X}_i)(1 + O(b^2) + O_p\left(\sqrt{\frac{logn}{nb^p}}\right)).$$

Therefore,

$$\begin{aligned}
\hat{\sigma}^2(\mathbf{x}) - \sigma^2(\mathbf{x}) &= \frac{1}{2}\sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})\left\{\frac{\frac{1}{n^2}\sum_{j=1}^n \mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)\left[(Y_i - Y_j)^2 - 2\sigma^2(\mathbf{x})\right]}{f(\mathbf{X}_i)f(\mathbf{x})}\right\} \\
&\quad \left(1 + O(h^2) + O(b^2) + O_p\left(\sqrt{\frac{logn}{nh^p}}\right) + O_p\left(\sqrt{\frac{logn}{nb^p}}\right)\right), \\
&= \frac{1}{2n^2 f(\mathbf{x})}\sum_{i=1}^n\sum_{j=1}^n \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)}\left[\{Y_i - Y_j\}^2 - 2\sigma^2(x)\right]
\end{aligned}$$

$$\left(1 + O(h^2) + O(b^2) + O_p\left(\sqrt{\frac{logn}{nh^p}}\right) + O_p\left(\sqrt{\frac{logn}{nb^p}}\right)\right).$$

Moreover, by Taylor's expansion (3.16) and (3.17),

$$\begin{aligned}
&\frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \left[\{Y_i - Y_j\}^2 - 2\sigma^2(x)\right] \\
&= A_1(\mathbf{x}) + A_2(\mathbf{x}) + A_3(\mathbf{x}) + A_4(\mathbf{x}) + A_5(\mathbf{x}) + A_6(\mathbf{x}) \\
&\quad + \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} O\left(\|\mathbf{X}_i - \mathbf{x}\|^3 + \|\mathbf{X}_j - \mathbf{x}\|^3\right),
\end{aligned}$$

where

$$\begin{aligned}
A_1(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[(\mathbf{X}_i - \mathbf{x})^\top \nabla m(\mathbf{x}) \\
&\qquad - (\mathbf{X}_j - \mathbf{x})^\top \nabla m(\mathbf{x})\Big]^2, \\
A_2(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) \\
&\qquad - \frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_j - \mathbf{x})\Big]^2, \\
A_3(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[(\mathbf{X}_i - \mathbf{x})^\top \nabla m(\mathbf{x}) \\
&\quad - (\mathbf{X}_j - \mathbf{x})^\top \nabla m(\mathbf{x})\Big] \left[\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) - \frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_j - \mathbf{x})\right], \\
A_4(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[(\mathbf{X}_i - \mathbf{x})^\top \nabla \sigma^2(\mathbf{x}) \\
&\qquad + (\mathbf{X}_j - \mathbf{x})^\top \nabla \sigma^2(\mathbf{x})\Big], \\
A_5(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) \\
&\qquad - \frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_j - \mathbf{x})\Big], \\
A_6(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \eta_{ij},
\end{aligned}$$

with

$$\eta_{ij} = 2(m(\mathbf{X}_i) - m(\mathbf{X}_j))[\sigma(\mathbf{X}_i)\varepsilon_i - \sigma(\mathbf{X}_j)\varepsilon_j] - 2\sigma(\mathbf{X}_i)\sigma(\mathbf{X}_j)\varepsilon_i\varepsilon_j + \sigma^2(\mathbf{X}_i)(\varepsilon_i^2 - 1) + \sigma^2(\mathbf{X}_j)(\varepsilon_j^2 - 1).$$

By results in Lemma 3.1, for a particular $\mathbf{x}$, we could approximate the above terms as follows,

$$\begin{aligned}
A_1(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[(\mathbf{X}_i - \mathbf{x})^\top \nabla m(\mathbf{x}) \\
&\qquad - (\mathbf{X}_j - \mathbf{x})^\top \nabla m(\mathbf{x})\Big]^2 \\
&= \frac{1}{2nh^p f(\mathbf{x})} \sum_{j=1}^{n} \int K(\frac{\mathbf{X}_j - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})\Big[(\mathbf{X}_j - \mathbf{x} + b\mathbf{u})^\top \nabla m(\mathbf{x}) \\
&\qquad - (\mathbf{X}_j - \mathbf{x})^\top \nabla m(\mathbf{x})\Big]^2 d\mathbf{u} \left[1 + O_p\left(\sqrt{\frac{logn}{nb^p}}\right)\right] \\
&= \frac{1}{2}\iint K(\mathbf{v} + \frac{b\mathbf{u}}{h})\mathcal{K}(u)[(b\mathbf{u})^\top \nabla m(\mathbf{x})]^2 \frac{f(\mathbf{x} + h\mathbf{v})}{f(\mathbf{x})} d\mathbf{u}d\mathbf{v} + O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right) \\
&= \frac{1}{2}\iint K(\mathbf{t})\mathcal{K}(u)[(b\mathbf{u})^\top \nabla m(\mathbf{x})]^2 \frac{f(\mathbf{x} + h\mathbf{t} - b\mathbf{u})}{f(\mathbf{x})} d\mathbf{u}d\mathbf{t} + O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right) \\
&= \frac{1}{2} b^2 \mathrm{tr}\{\mathbf{M}_2^K \nabla m(\mathbf{x})(\nabla m(\mathbf{x}))^\top\} + O(h^2 b^2) + O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right),
\end{aligned}$$

where the last equality holds because of Taylor's expansion.

Based on Conditions (C1)-(C6), Lemma 3.1 and the similar arguments, we have

$$\begin{aligned}
A_2(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) \\
&\qquad - \frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_j - \mathbf{x})\Big]^2 \\
&= \frac{1}{2nh^p f(\mathbf{x})} \sum_{j=1}^{n} \int K(\frac{\mathbf{X}_j - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})\Big[\frac{1}{2}(\mathbf{X}_j - \mathbf{x} + b\mathbf{u})^\top \mathbf{H}(\mathbf{x})(\mathbf{X}_j - \mathbf{x} + b\mathbf{u})
\end{aligned}$$

$$
\begin{aligned}
&\quad -\frac{1}{2}(\mathbf{X}_j-\mathbf{x})^\top\mathbf{H}(\mathbf{x})(\mathbf{X}_j-\mathbf{x})\Big]^2 d\mathbf{u}\left[1+O_p\left(\sqrt{\frac{logn}{nb^p}}\right)\right]\\
&=\frac{1}{2}\iint K(\mathbf{v}+\frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})\big[(b\mathbf{u})^\top\mathbf{H}(\mathbf{x})(h\mathbf{v})+\frac{1}{2}(b\mathbf{u})^\top\mathbf{H}(\mathbf{x})(b\mathbf{u})\big]^2\frac{f(\mathbf{x}+h\mathbf{v})}{f(\mathbf{x})}d\mathbf{u}d\mathbf{v}\\
&\qquad +O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right)\\
&=\frac{1}{2}\iint K(\mathbf{t})\mathcal{K}(\mathbf{u})\big[(b\mathbf{u})^\top\mathbf{H}(\mathbf{x})(h\mathbf{t}-b\mathbf{u})+\frac{1}{2}(b\mathbf{u})^\top\mathbf{H}(\mathbf{x})(b\mathbf{u})\big]^2\frac{f(\mathbf{x}+h\mathbf{t}-b\mathbf{u})}{f(\mathbf{x})}d\mathbf{u}d\mathbf{t}\\
&\qquad +O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right)\\
&=O(h^2b^2+b^4)+O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right),
\end{aligned}
$$

where the last equality is obtained due to Condition (C5) and Taylor's expansion. Besides,

$$
\begin{aligned}
A_3(\mathbf{x})&=\frac{1}{2n^2f(\mathbf{x})}\sum_{i=1}^n\sum_{j=1}^n\frac{K_h(\mathbf{X}_i-\mathbf{x})\mathcal{K}_b(\mathbf{X}_i-\mathbf{X}_j)}{f(\mathbf{X}_i)}\Big[(\mathbf{X}_i-\mathbf{x})^\top\nabla m(\mathbf{x})\\
&\quad -(\mathbf{X}_j-\mathbf{x})^\top\nabla m(\mathbf{x})\Big]\left[\frac{1}{2}(\mathbf{X}_i-\mathbf{x})^\top\mathbf{H}(\mathbf{x})(\mathbf{X}_i-\mathbf{x})-\frac{1}{2}(\mathbf{X}_j-\mathbf{x})^\top\mathbf{H}(\mathbf{x})(\mathbf{X}_j-\mathbf{x})\right]\\
&=\frac{1}{2nh^pf(\mathbf{x})}\sum_{j=1}^n\int K(\frac{\mathbf{X}_j-\mathbf{x}}{h}+\frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})\Big[(\mathbf{X}_j-\mathbf{x}+b\mathbf{u})^\top\nabla m(\mathbf{x})\\
&\quad -(\mathbf{X}_j-\mathbf{x})^\top\nabla m(\mathbf{x})\Big]\Big[(\mathbf{X}_j-\mathbf{x}+b\mathbf{u})^\top\mathbf{H}(\mathbf{x})(\mathbf{X}_j-\mathbf{x}+b\mathbf{u})\\
&\quad -(\mathbf{X}_j-\mathbf{x})^\top\mathbf{H}(\mathbf{x})(\mathbf{X}_j-\mathbf{x})\Big]d\mathbf{u}\left[1+O_p\left(\sqrt{\frac{logn}{nb^p}}\right)\right]\\
&=\frac{1}{2}\iint K(\mathbf{v}+\frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})\big[(b\mathbf{u})^\top\nabla m(\mathbf{x})\big]\big[(b\mathbf{u})^\top\mathbf{H}(\mathbf{x})(b\mathbf{u})+2(b\mathbf{u})^\top\mathbf{H}(\mathbf{x})(h\mathbf{v})\big]\\
&\quad \frac{f(\mathbf{x}+h\mathbf{v})}{f(\mathbf{x})}d\mathbf{u}d\mathbf{v}+O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right)\\
&=o(b^2)+O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right),
\end{aligned}
$$

$$
A_4(\mathbf{x})=\frac{1}{2n^2f(\mathbf{x})}\sum_{i=1}^n\sum_{j=1}^n\frac{K_h(\mathbf{X}_i-\mathbf{x})\mathcal{K}_b(\mathbf{X}_i-\mathbf{X}_j)}{f(\mathbf{X}_i)}\Big[(\mathbf{X}_i-\mathbf{x})^\top\nabla\sigma^2(\mathbf{x})
$$

$$+(\mathbf{X}_j - \mathbf{x})^\top \nabla\sigma^2(\mathbf{x})\Big],$$

$$= \frac{1}{2}\iint K(\mathbf{v} + \frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})[(b\mathbf{u} + h\mathbf{v})^\top \nabla\sigma^2(\mathbf{x}) + (h\mathbf{v})^\top \nabla\sigma^2(\mathbf{x})]\frac{f(\mathbf{x} + h\mathbf{v})}{f(\mathbf{x})}d\mathbf{u}d\mathbf{v}$$

$$+O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right)$$

$$= \frac{1}{2}\iint K(\mathbf{t})\mathcal{K}(\mathbf{u})[(h\mathbf{t})^\top \nabla\sigma^2(\mathbf{x}) + (h\mathbf{t} - b\mathbf{u})^\top \nabla\sigma^2(\mathbf{x})]\frac{f(\mathbf{x} + h\mathbf{t} - b\mathbf{u})}{f(\mathbf{x})}d\mathbf{u}d\mathbf{t}$$

$$+O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right)$$

$$= (h^2 + \frac{1}{2}b^2)[\nabla f(\mathbf{x})]^\top \mathbf{M}_2^K \nabla\sigma^2(\mathbf{x})/f(\mathbf{x}) + o(h^2 + b^2) + O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right),$$

and

$$A_5(\mathbf{x}) = \frac{1}{2n^2 f(\mathbf{x})}\sum_{i=1}^n\sum_{j=1}^n \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)}\Big[\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_i - \mathbf{x})$$

$$-\frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_j - \mathbf{x})\Big]$$

$$= \frac{1}{2nh^p f(\mathbf{x})}\sum_{j=1}^n \int K(\frac{\mathbf{X}_j - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})\Big[\frac{1}{2}(\mathbf{X}_j - \mathbf{x} + b\mathbf{u})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_j - \mathbf{x} + b\mathbf{u})$$

$$-\frac{1}{2}(\mathbf{X}_j - \mathbf{x})^\top \mathbf{S}(\mathbf{x})(\mathbf{X}_j - \mathbf{x})\Big]d\mathbf{u}\left[1 + O_p\left(\sqrt{\frac{logn}{nb^p}}\right)\right]$$

$$= \frac{1}{2}\iint K(\mathbf{v} + \frac{b\mathbf{u}}{h})\mathcal{K}(\mathbf{u})\Big[\frac{1}{2}(h\mathbf{v} + b\mathbf{u})^\top \mathbf{S}(\mathbf{x})(h\mathbf{v} + b\mathbf{u}) - \frac{1}{2}(h\mathbf{v})^\top \mathbf{S}(\mathbf{x})(h\mathbf{v})\Big]$$

$$\frac{f(\mathbf{x} + h\mathbf{v})}{f(\mathbf{x})}d\mathbf{u}d\mathbf{v} + O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right)$$

$$= (\frac{1}{2}h^2 + \frac{1}{4}b^2)\mathrm{tr}\{\mathbf{M}_2^K \mathbf{S}(\mathbf{x})\} + O(h^3) + O_p\left(b^2\sqrt{\frac{logn}{nb^p}}\right).$$

More importantly,

$$A_6(\mathbf{x}) = \frac{1}{2n^2 f(\mathbf{x})}\sum_{i=1}^n\sum_{j=1}^n \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)}\eta_{ij}$$

$$= A_{61}(\mathbf{x}) + A_{62}(\mathbf{x}) + A_{63}(\mathbf{x}),$$

where

$$A_{61}(\mathbf{x}) = \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[\sigma^2(\mathbf{X}_i)(\varepsilon_i^2 - 1) + \sigma^2(\mathbf{X}_j)(\varepsilon_j^2 - 1)\Big],$$

$$A_{62}(\mathbf{x}) = \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[2(m(\mathbf{X}_i) - m(\mathbf{X}_j)) (\sigma(\mathbf{X}_i)\varepsilon_i - \sigma(\mathbf{X}_j)\varepsilon_j)\Big],$$

$$A_{63}(\mathbf{x}) = \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[- 2\sigma(\mathbf{X}_i)\sigma(\mathbf{X}_j)\varepsilon_i\varepsilon_j\Big],$$

because of the definition of η_{ij}.

Note that by Lemma 3.1 and the similar technique utilized before, we could obtain that

$$A_{62}(\mathbf{x}) = o(h^2 + b^2), \quad A_{63}(\mathbf{x}) = o(h^2 + b^2),$$

consequently we will focus on A_{61} in the following. By Lemma 3.1, we approximate

$$\begin{aligned} A_{61}(\mathbf{x}) &= \frac{1}{2n^2 f(\mathbf{x})} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{K_h(\mathbf{X}_i - \mathbf{x})\mathcal{K}_b(\mathbf{X}_i - \mathbf{X}_j)}{f(\mathbf{X}_i)} \Big[\sigma^2(\mathbf{X}_i)(\varepsilon_i^2 - 1) + \sigma^2(\mathbf{X}_j)(\varepsilon_j^2 - 1)\Big] \\ &= \frac{1}{2nh^p f(\mathbf{x})} \sum_{j=1}^{n} \left\{ \int \left[K(\frac{\mathbf{X}_j - \mathbf{x}}{h}) + K(\frac{\mathbf{X}_j - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h}) \right] K(\mathbf{u}) d\mathbf{u} \right\} \sigma^2(\mathbf{X}_j)(\varepsilon_j^2 - 1) \\ &\quad \left[1 + O_p\left(\sqrt{\frac{logn}{nb^p}} \right) \right], \\ &= \frac{1}{n} \sum_{j=1}^{n} I_j(\mathbf{x}) \left[1 + O_p\left(\sqrt{\frac{logn}{nb^p}} \right) \right], \end{aligned}$$

by letting

$$I_j(\mathbf{x}) = \frac{1}{2h^p f(\mathbf{x})} \left\{ \int \left[K(\frac{\mathbf{X}_j - \mathbf{x}}{h}) + K(\frac{\mathbf{X}_j - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h}) \right] K(\mathbf{u}) d\mathbf{u} \right\} \sigma^2(\mathbf{X}_j)(\varepsilon_j^2 - 1).$$

Note that by Condition (C2),

$$\mathrm{Var}(I_1(\mathbf{x})) = \frac{\sigma^4(\mathbf{x})\mathrm{Var}(\varepsilon_1^2)}{4h^p f(\mathbf{x})} \int \left[\int \Big(K(\mathbf{v}) + K(\mathbf{v} + \frac{b\mathbf{u}}{h}) \Big) K(\mathbf{u}) d\mathbf{u} \right]^2 d\mathbf{v} + o(\frac{1}{h^p}),$$

by stationarity in (C3), we have

$$\mathrm{Var}(A_{61}(\mathbf{x})) = \frac{1}{n}\mathrm{Var}(I_1(\mathbf{x})) + \frac{2}{n}\sum_{\ell=1}^{n-1}(1 - \frac{\ell}{n})\mathrm{Cov}(I_1(\mathbf{x}), I_{\ell+1}(\mathbf{x})).$$

Let $d_n \to \infty$ be a sequence of integers such that $h^p d_n \to 0$. Define

$$R_1(\mathbf{x}) = \sum_{\ell=1}^{d_n - 1} |\mathrm{Cov}(I_1(\mathbf{x}), I_{\ell+1}(\mathbf{x}))|, \quad R_2(\mathbf{x}) = \sum_{\ell=d_n}^{n-1} |\mathrm{Cov}(I_1(\mathbf{x}), I_{\ell+1}(\mathbf{x}))|.$$

by Condition (C2), (C4) and (C5) and conditioning on $(\mathbf{X}_1, \mathbf{X}_{\ell+1})$, we have

$$\begin{aligned}
&|\mathrm{Cov}(I_1(\mathbf{x}), I_{\ell+1}(\mathbf{x}))| \\
=& |E\{E(I_1(\mathbf{x})I_{\ell+1}(\mathbf{x})|\mathbf{X}_1, \mathbf{X}_{\ell+1})\}| \\
=& \Big| E\Big\{ E\Big(\frac{1}{4h^{2p}f^2(\mathbf{x})} \Big[\Big\{ \int \Big[K(\frac{\mathbf{X}_1 - \mathbf{x}}{h}) + K(\frac{\mathbf{X}_1 - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h}) \Big] K(\mathbf{u}) d\mathbf{u} \Big\} \\
& \sigma^2(\mathbf{X}_1)(\varepsilon_1^2 - 1) \Big] \Big[\Big\{ \int \Big[K(\frac{\mathbf{X}_{\ell+1} - \mathbf{x}}{h}) + K(\frac{\mathbf{X}_{\ell+1} - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h}) \Big] K(\mathbf{u}) d\mathbf{u} \Big\} \\
& \sigma^2(\mathbf{X}_{\ell+1})(\varepsilon_{\ell+1}^2 - 1) \Big] \Big) \Big| \mathbf{X}_1, \mathbf{X}_{\ell+1} \Big\} \Big| \\
\le& CM_0,
\end{aligned}$$

for $M_0 > 0$ and some generic constant $C > 0$. Then it follows that $R_1(\mathbf{x}) \le Cd_n M_0$. We now consider the contribution of $R_2(\mathbf{x})$. Because of the property of α-mixing process, then by Davydov's lemma,

$$|\mathrm{Cov}(I_1(\mathbf{x}), I_{\ell+1}(\mathbf{x}))| = E|I_1(\mathbf{x})I_{\ell+1}(\mathbf{x})| \leq 8[\alpha(\ell)]^{\frac{\delta}{1+\delta}}\{E|I_1(\mathbf{x})|^{2(1+\delta)}\}^{\frac{1}{1+\delta}}.$$

By conditioning on $\mathbf{X}_1$, and using (C2) and (C3),

$$\begin{aligned} & E|I_1(\mathbf{x})|^{2(1+\delta)} \\ &= E\Big|\frac{1}{2h^p f(\mathbf{x})}\Big\{\int\Big[K(\frac{\mathbf{X}_j - \mathbf{x}}{h}) + K(\frac{\mathbf{X}_j - \mathbf{x}}{h} + \frac{b\mathbf{u}}{h})\Big]K(\mathbf{u})d\mathbf{u}\Big\} \\ &\qquad \sigma^2(\mathbf{X}_j)(\varepsilon_j^2 - 1)\Big|^{2(1+\delta)} \\ &\leq CM_1 h^{-2p(1+\delta)+p}, \end{aligned}$$

for $M_1 > 0$. Hence, for $\frac{\delta}{1+\delta} < \gamma < 1$,

$$\begin{aligned} R_2(\mathbf{x}) &\leq \sum_{\ell=d_n}^{n-1} 8[\alpha(\ell)]^{\frac{\delta}{1+\delta}}\{E|I_1(\mathbf{x})|^{2(1+\delta)}\}^{\frac{1}{1+\delta}} \\ &\leq (CM_1)^{\frac{1}{1+\delta}} 8(h^{-2p(1+\delta)+p})^{\frac{1}{1+\delta}} \sum_{\ell=d_n}^{\infty}[\ell^{-\beta}]^{\frac{\delta}{1+\delta}} \\ &\leq M_2 h^{-2p+\frac{p}{1+\delta}} \sum_{\ell=d_n}^{\infty} \ell^{-2} = M_2 h^{-2p+\frac{p}{1+\delta}} d_n^{-\gamma} \sum_{\ell=d_n}^{\infty} \ell^{-2+\gamma} = o(1/h^p), \end{aligned}$$

by taking $h^{-p+\frac{p}{1+\delta}} d_n^{-\gamma} = 1$. Together with above results,

$$\sum_{s=1}^{n-1} \mathrm{Cov}(I_1(u), I_{s+1}(u)) = o(1/h^p),$$

and

$$\mathrm{Var}(\frac{1}{n}\sum_{j=1}^{n} I_j(\mathbf{x})) = \frac{\sigma^4(\mathbf{x})\mathrm{Var}(\varepsilon_1^2)}{4h^p f(\mathbf{x})} \int\Big[\int\Big(K(\mathbf{v}) + K(\mathbf{v} + \frac{b\mathbf{u}}{h})\Big)K(\mathbf{u})d\mathbf{u}\Big]^2 d\mathbf{v} + o(\frac{1}{h^p}).$$

Subsequently, as $n \to \infty$, $h \to 0$, similar to other nonparametric estimators for strong mixing time series, the asymptotic normality of the estimator $\hat{\sigma}^2(\mathbf{x})$ could be established by employing the so-called small-block and large-block arguments, thus

$$\sqrt{nh^p}\{\hat{\sigma}^2(\mathbf{x}) - \sigma^2(\mathbf{x}) - B(\mathbf{x})\} \xrightarrow{d} N\left(0, \frac{\sigma^4(\mathbf{x})\mathrm{Var}(\varepsilon_1^2)}{4f(\mathbf{x})}\nu_{K,K}^2\right),$$

where

$$\begin{aligned} B(\mathbf{x}) = & \frac{1}{2}h^2\mathrm{tr}\left\{\frac{2\mathbf{M}_2^K\nabla\sigma^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})} + \mathbf{M}_2^K\mathbf{S}(\mathbf{x})\right\} \\ & + \frac{1}{2}b^2\mathrm{tr}\Big\{\mathbf{M}_2^K\nabla m(\mathbf{x})(\nabla m(\mathbf{x}))^\top + \mathbf{M}_2^K\nabla\sigma^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top + \mathbf{M}_2^K\mathbf{S}(\mathbf{x})\Big\}, \\ \nu_{K,K}^2 = & \int\left[\int\Big(K(\mathbf{v}) + K(\mathbf{v} + \frac{b\mathbf{u}}{h})\Big)K(\mathbf{u})d\mathbf{u}\right]^2 d\mathbf{v}. \end{aligned}$$

□

Proof of Theorem 3.2. Note that

$$\begin{aligned} \sigma_+^2(\mathbf{x}) &= \mathrm{Var}(Y + Z|\mathbf{X} = \mathbf{x}) = \sigma^2(\mathbf{x}) + \phi^2(\mathbf{x}) + 2\sigma_{Y,Z}(\mathbf{x}), \\ \sigma_-^2(\mathbf{x}) &= \mathrm{Var}(Y - Z|\mathbf{X} = \mathbf{x}) = \sigma^2(\mathbf{x}) + \phi^2(\mathbf{x}) - 2\sigma_{Y,Z}(\mathbf{x}), \end{aligned}$$

and $\sigma_{Y,Z}(\mathbf{x}) = \frac{1}{4}[\sigma_+^2(\mathbf{x}) - \sigma_-^2(\mathbf{x})]$, based on the definitions of estimators $\hat{\sigma}_+^2(\mathbf{x})$, $\hat{\sigma}_-^2(\mathbf{x})$ and $\hat{\sigma}_{Y,Z}(\mathbf{x})$, we could obtain the following expressions immediately from the conclusions in Theorem 3.1. Therefore,

$$\begin{aligned} & \hat{\sigma}_+^2(\mathbf{x}) - \sigma_+^2(\mathbf{x}) \\ = & \frac{1}{2}h_+^2\mathrm{tr}\left\{\frac{2\mathbf{M}_2^K\nabla\sigma_+^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})} + \mathbf{M}_2^K\mathbf{S}_+(\mathbf{x})\right\} + \frac{1}{2}b_+^2\mathrm{tr}\Big\{\mathbf{M}_2^K\nabla\mu_+(\mathbf{x})(\nabla\mu_+(\mathbf{x}))^\top \\ & + \mathbf{M}_2^K\nabla\sigma_+^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top + \mathbf{M}_2^K\mathbf{S}_+(\mathbf{x})\Big\} + \frac{1}{2nh_+^p f(\mathbf{x})}\sum_{j=1}^n\Big\{\int\Big[K(\frac{\mathbf{X}_j - \mathbf{x}}{h_+}) \end{aligned}$$

$$+K(\frac{\mathbf{X}_j-\mathbf{x}}{h_+}+\frac{b_+\mathbf{u}}{h_+})\Big]K(\mathbf{u})d\mathbf{u}\Big\}\ \Big\{\sigma^2(\mathbf{X}_j)(\varepsilon_j^2-1)+\phi^2(\mathbf{X}_j)(\varepsilon_j^2-1)$$
$$+2\sigma(\mathbf{X}_j)\phi(\mathbf{X}_j)[\varepsilon_j^2-1]\Big\}+o_p(h_+^2+\sqrt{\frac{logn}{nh_+^p}}),$$

and

$$\hat{\sigma}_-^2(\mathbf{x})-\sigma_-^2(\mathbf{x})$$
$$=\frac{1}{2}h_-^2\,\mathrm{tr}\left\{\frac{2\mathbf{M}_2^K\nabla\sigma_-^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})}+\mathbf{M}_2^K\mathbf{S}_-(\mathbf{x})\right\}+\frac{1}{2}b_-^2\,\mathrm{tr}\Big\{\mathbf{M}_2^K\nabla\mu_-(\mathbf{x})(\nabla\mu_-(\mathbf{x}))^\top$$
$$+\mathbf{M}_2^K\nabla\sigma_-^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top+\mathbf{M}_2^K\mathbf{S}_-(\mathbf{x})\Big\}+\frac{1}{2nh_-^pf(\mathbf{x})}\sum_{j=1}^n\Big\{\int\Big[K(\frac{\mathbf{X}_j-\mathbf{x}}{h_-})$$
$$+K(\frac{\mathbf{X}_j-\mathbf{x}}{h_-}+\frac{b_-\mathbf{u}}{h_-})\Big]K(\mathbf{u})d\mathbf{u}\Big\}\ \Big\{\sigma^2(\mathbf{X}_j)(\varepsilon_j^2-1)+\phi^2(\mathbf{X}_j)(\varepsilon_j^2-1)$$
$$-2\sigma(\mathbf{X}_j)\phi(\mathbf{X}_j)[\varepsilon_j^2-1]\Big\}+o_p(h_-^2+\sqrt{\frac{logn}{nh_-^p}}),$$

consequently, if $b_+=o(h_+)$, $b_-=o(h_-)$ and $h_-=O(h_+)$, then

$$\hat{\sigma}_{Y,Z}(\mathbf{x})-\sigma_{Y,Z}(\mathbf{x})$$
$$=\frac{1}{4}\big[\hat{\sigma}_+^2(\mathbf{x})-\hat{\sigma}_-^2(\mathbf{x})-(\sigma_+^2(\mathbf{x})-\sigma_-^2(\mathbf{x}))\big]$$
$$=\frac{1}{8}h_+^2\,\mathrm{tr}\left\{\frac{2\mathbf{M}_2^K\nabla\sigma_+^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})}+\mathbf{M}_2^K\mathbf{S}_+(\mathbf{x})\right\}$$
$$-\frac{1}{8}h_-^2\,\mathrm{tr}\left\{\frac{2\mathbf{M}_2^K\nabla\sigma_-^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})}+\mathbf{M}_2^K\mathbf{S}_-(\mathbf{x})\right\}$$
$$+\frac{1}{8nf(\mathbf{x})}\sum_{j=1}^n K*K_{h_+,h_-}^-(\mathbf{X}_j-\mathbf{x})\big[\sigma^2(\mathbf{X}_j)(\varepsilon_j^2-1)+\phi^2(\mathbf{X}_j)(\varepsilon_j^2-1)\big]$$
$$+\frac{1}{8nf(\mathbf{x})}\sum_{j=1}^n K*K_{h_+,h_-}^+(\mathbf{X}_j-\mathbf{x})\big[2\sigma(\mathbf{X}_j)\phi(\mathbf{X}_j)(\varepsilon_j^2-1)\big]$$
$$+o_p\left(h_-^2+\sqrt{\frac{logn}{nh_-^p}}\right),$$

where

$$\begin{aligned}
K*K^{-}_{h_+,h_-}(\mathbf{X}_j-\mathbf{x}) &= \frac{1}{h_+^p}\int\Big[K(\frac{\mathbf{X}_j-\mathbf{x}}{h_+})+K(\frac{\mathbf{X}_j-\mathbf{x}}{h_+}+\frac{b_+\mathbf{u}}{h_+})\Big]K(\mathbf{u})d\mathbf{u}\\
&\quad -\frac{1}{h_-^p}\int\Big[K(\frac{\mathbf{X}_j-\mathbf{x}}{h_-})+K(\frac{\mathbf{X}_j-\mathbf{x}}{h_-}+\frac{b_+\mathbf{u}}{h_-})\Big]K(\mathbf{u})d\mathbf{u},\\
K*K^{+}_{h_+,h_-}(\mathbf{X}_j-\mathbf{x}) &= \frac{1}{h_+^p}\int\Big[K(\frac{\mathbf{X}_j-\mathbf{x}}{h_+})+K(\frac{\mathbf{X}_j-\mathbf{x}}{h_+}+\frac{b_+\mathbf{u}}{h_+})\Big]K(\mathbf{u})d\mathbf{u}\\
&\quad +\frac{1}{h_-^p}\int\Big[K(\frac{\mathbf{X}_j-\mathbf{x}}{h_-})+K(\frac{\mathbf{X}_j-\mathbf{x}}{h_-}+\frac{b_+\mathbf{u}}{h_-})\Big]K(\mathbf{u})d\mathbf{u}.
\end{aligned}$$

By letting

$$\begin{aligned}
Q_j(\mathbf{x}) &= \frac{1}{8f(\mathbf{x})}K*K^{-}_{h_+,h_-}(\mathbf{X}_j-\mathbf{x})\big[\sigma^2(\mathbf{X}_j)(\varepsilon_j^2-1)+\phi^2(\mathbf{X}_j)(\varepsilon_j^2-1)\big]\\
&\quad +\frac{1}{8f(\mathbf{x})}K*K^{+}_{h_+,h_-}(\mathbf{X}_j-\mathbf{x})\big[2\sigma(\mathbf{X}_j)\phi(\mathbf{X}_j)(\varepsilon_j^2-1)\big],
\end{aligned}$$

and $b_+=o(h_+)$, $b_-=o(h_-)$, we could obtain that

$$\begin{aligned}
&\mathrm{Var}(Q_1(x))\\
&= \frac{1}{4f(\mathbf{x})h_-^p}\gamma^2_{cov}(\mathbf{x})\int\Big[\frac{h_-^p}{h_+^p}K\left(\frac{h_-\mathbf{v}}{h_+}\right)+K(\mathbf{v})\Big]^2d\mathbf{v}\\
&\quad +\frac{1}{16f(\mathbf{x})h_-^p}\gamma^2(\mathbf{x})\int\Big[\frac{h_-^p}{h_+^p}K\left(\frac{h_-\mathbf{v}}{h_+}\right)-K(\mathbf{v})\Big]^2d\mathbf{v}\\
&\quad -\frac{1}{8f(\mathbf{x})h_-^p}\theta(\mathbf{x})\int\Big[\frac{h_-^{2p}}{h_+^{2p}}K^2(\frac{h_-\mathbf{v}}{h_+})-K^2(\mathbf{v})\Big]d\mathbf{v}+o(\frac{1}{h_-^p}),
\end{aligned}$$

where

$$\begin{aligned}
\gamma^2(\mathbf{x}) &= E\Big\{\big[\sigma^2(\mathbf{X}_1)(\varepsilon_1^2-1)+\phi^2(\mathbf{X}_1)(\varepsilon_1^2-1)\big]^2|\mathbf{X}_1=\mathbf{x}\Big\},\\
\gamma^2_{cov}(\mathbf{x}) &= E\Big\{\sigma^2(\mathbf{X}_1)\phi^2(\mathbf{X}_1)(\varepsilon_1^2-1)^2|\mathbf{X}_1=\mathbf{x}\Big\},\\
\theta(\mathbf{x}) &= E\Big\{\big[\sigma^2(\mathbf{X}_1)(\varepsilon_1^2-1)+\phi^2(\mathbf{X}_1)(\varepsilon_1^2-1)\big]\\
&\qquad \big[\sigma(\mathbf{X}_1)\phi(\mathbf{X}_1)(\varepsilon_1^2-1)\big]|\mathbf{X}_1=\mathbf{x}\Big\}.
\end{aligned}$$

Therefore, following the similar arguments in the proof of Theorem 3.1 based on the regular conditions and stationarity,

$$\begin{aligned}
&\mathrm{Var}(\frac{1}{n}\sum_{j=1}^{n} Q_j(x))\\
=&\frac{\mathrm{Var}(Q_1(\mathbf{x}))}{n} + o(\frac{1}{nh_-^p})\\
=&\frac{1}{4nf(\mathbf{x})h_-^p}\gamma_{cov}^2(\mathbf{x})\int\left[\frac{h_-^p}{h_+^p}K\left(\frac{h_-\mathbf{v}}{h_+}\right)+K(\mathbf{v})\right]^2 d\mathbf{v}\\
&+\frac{1}{16nf(\mathbf{x})h_-^p}\gamma^2(\mathbf{x})\int\left[\frac{h_-^p}{h_+^p}K\left(\frac{h_-\mathbf{v}}{h_+}\right)-K(\mathbf{v})\right]^2 d\mathbf{v}\\
&-\frac{1}{8nf(\mathbf{x})h_-^p}\theta(\mathbf{x})\int\left[\frac{h_-^{2p}}{h_+^{2p}}K^2(\frac{h_-\mathbf{v}}{h_+})-K^2(\mathbf{v})\right]d\mathbf{v} + o(\frac{1}{nh_-^p}).
\end{aligned}$$

Consequently, by assuming $b_+ = o(h_+)$ and $b_- = o(h_-)$, we could establish the asymptotic distribution of the conditional covariance estimator $\hat{\sigma}_{Y,Z}(\mathbf{x})$ as follows,

$$\begin{aligned}
&\sqrt{nh_-^p}\{\hat{\sigma}_{Y,Z}(\mathbf{x})-\sigma_{Y,Z}(\mathbf{x})-B_{Y,Z}(\mathbf{x})\}\\
&\qquad\xrightarrow{d} N\left(0, \frac{\gamma_{cov}^2(\mathbf{x})\nu_{K,K,+}^2+\frac{\gamma^2(\mathbf{x})\nu_{K,K,-}^2}{4}-\frac{\theta(\mathbf{x})\nu_{K,K,2}}{2}}{4f(\mathbf{x})}\right),
\end{aligned}$$

where

$$\begin{aligned}
B_{Y,Z}(\mathbf{x}) =& \frac{1}{8}h_+^2\mathrm{tr}\left\{\frac{2\mathbf{M}_2^K\nabla\sigma_+^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})}+\mathbf{M}_2^K\mathbf{S}_+(\mathbf{x})\right\}\\
&-\frac{1}{8}h_-^2\mathrm{tr}\left\{\frac{2\mathbf{M}_2^K\nabla\sigma_-^2(\mathbf{x})(\nabla f(\mathbf{x}))^\top}{f(\mathbf{x})}+\mathbf{M}_2^K\mathbf{S}_-(\mathbf{x})\right\},\\
\nu_{K,K,+}^2 =& \int\left[\frac{h_-^p}{h_+^p}K\left(\frac{h_-\mathbf{v}}{h_+}\right)+K(\mathbf{v})\right]^2 d\mathbf{v},\\
\nu_{K,K,-}^2 =& \int\left[\frac{h_-^p}{h_+^p}K\left(\frac{h_-\mathbf{v}}{h_+}\right)-K(\mathbf{v})\right]^2 d\mathbf{v},\\
\nu_{K,K,2} =& \int\left[\frac{h_-^{2p}}{h_+^{2p}}K^2(\frac{h_-\mathbf{v}}{h_+})-K^2(\mathbf{v})\right]d\mathbf{v}.
\end{aligned}$$

□

CHAPTER 4

Conclusion and Future work

In Chapter 2, we investigated the estimation problem of conditional covariance matrix from the perspective of the behaviour of dynamic conditional correlation coefficients. In the beginning, we obtained the nonparametric estimators of unknown conditional means and variances by local linear smoothing. Next, the resulting estimators for pairwise conditional correlations were also derived through smoothing techniques based on preliminary estimates. Since factor models serve as an effective tool in dimension reduction, we introduced a reduced rank model for the conditional correlation coefficients to characterize the variation pattern, by regarding $F_1(u), \cdots, F_p(u)$ as functional common factors. Our estimation of common functional factors and coefficients relies on nonparametric smoothing, thus it is model free and allows much more flexibility. In addition, a detailed theoretical discussion of the estimators of conditional correlation coefficients, common functional factors and loadings was presented under some regularity conditions. Moreover, as indicated in our empirical analysis, it is worth mentioning that our proposed approach could successfully describe the movement of pairwise correlations and explain the asymmetric effect of returns on the conditional correlations through estimated common factors.

In Chapter 3 of this thesis, the estimation problem of conditional variance functions was examined. Instead of relying on estimates of the conditional means, we proposed a new approach by incorporating the smoothing techniques into the difference-based methods. Therefore, we could take advantage of difference-based approaches by omitting a complicated construction of difference sequences, but providing estimates of local variations by kernel smoothing at the first stage. Asymptotic properties of our conditional variance (covariance) estimator were examined under some mild conditions. Finally, a simulation study was conducted for the purpose of comparing the performance of our variance estimator with the residual-based estimator of Fan and Yao (1998) and the difference-based estimator of Cai, Levine and Wang (2009). By investigating various sets of conditional mean functions and variance functions, our developed estimator seems superior to other two estimators in most situations.

There are some open problems for future research.

1. In Chapter 2, we only imposed a factor model structure on all pairwise conditional correlations, and focused on the problem about identifying common functional factors. Actually, we could introduce the factor model representation into the error terms, and study the estimation problem of conditional covariance matrix directly. Specifically, let $Z = (\mathbf{z}_1, ..., \mathbf{z}_p)^\top \in \mathcal{R}^p$ be a p-dimensional random vector, and $\mu(u) = (\mu_1(u), ..., \mu_p(u))^\top$ be a vector of measurable functions of u on an interval. For a random variable U, we consider the following model framework for $\mathbf{z}_k$,

$$\mathbf{z}_k = \mu_k(U) + \mathbf{x}_k, \quad k = 1, ..., p, \tag{4.1}$$

where $\mu_k(u) = E(\mathbf{z}_k|U = u)$ is the conditional mean function of $\mathbf{z}_k$, $k = 1, ..., p$, and $X = (\mathbf{x}_1, \cdots, \mathbf{x}_p)^\top$ is defined to be a zero-mean random vector. We introduce the following conditional factor model

$$\mathbf{x}_k = \ell_{k1}(U)\epsilon_1 + ... + \ell_{kq}(U)\epsilon_q + \sigma_k\eta_k, \quad k = 1, ..., p. \tag{4.2}$$

where $\epsilon_1, ..., \epsilon_q$ and $\eta_1, ..., \eta_p$ are IID random variables with mean 0 and variance 1. Subsequently, it follows that

$$Var(\mathbf{z}_k|U) = \ell_{k1}^2(U) + ... + \ell_{kq}^2(U) + \sigma_k^2, \quad k = 1, ..., p,$$

and

$$Cov(\mathbf{z}_k, \mathbf{z}_j|U) = \ell_{k1}(U)\ell_{j1}(U) + ... + \ell_{kq}(U)\ell_{jq}(U), \quad k \neq j.$$

Rewrite (4.2) as

$$X = \mathbf{L}(U)\epsilon + \eta, \tag{4.3}$$

with $\epsilon = (\epsilon_1, ..., \epsilon_q)^\top$, $\eta = (\eta_1, ..., \eta_p)^\top$, and

$$\mathbf{L}(U) = \begin{pmatrix} \ell_{11}(U) & \cdots & \ell_{1q}(U) \\ \vdots & & \vdots \\ \ell_{p1}(U) & \cdots & \ell_{pq}(U) \end{pmatrix},$$

then $\epsilon_1, ..., \epsilon_q$ in (4.3) are actually latent common factors, and $\ell_{kj}(u), k = 1, ..., p, j = 1, ..., q$ are functional factor loadings. Therefore, the main focus will be to construct effective approaches to estimate the conditional variance-covariance matrix $\mathbf{\Sigma}(U)$ given by

$$\mathbf{\Sigma}(U) = E(XX^\top|U) = \mathbf{L}(U)\mathbf{L}^\top(U) + \mathbf{\Sigma}_\eta, \tag{4.4}$$

where $\mathbf{\Sigma}_\eta = \mathrm{diag}(\sigma_1^2, ..., \sigma_p^2)$.

2. A much more convenient approach to modelling conditional variance (covariance) functions has been constructed in Chapter 3, and reveals better performance in the simulation study. Therefore, the estimated conditional variances (covariances) could serve as raw estimates of the elements in the conditional covariance matrix. Most importantly, further studies about some specific structures of conditional covariance matrices (such as sparsity), could also be conducted based on those raw estimators.

Bibliography

Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation, *Journal of Business & Economic Statistics* **31**(3): 282–299.

Alexander, C. and Chibumba, A. M. (1997). Multivariate orthogonal factor GARCH. Mimeo, University of Sussex.

Amira, K., Taamouti, A. and Tsafack, G. (2011). What drives international equity correlations? volatility or market direction?, *Journal of International Money and Finance* **30**(6): 1234–1263.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001). The distribution of realized exchange rate volatility, *Journal of the American statistical association* **96**(453): 42–55.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts, *Review of Financial studies* **15**(4): 1137–1187.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios, *Journal of Financial Economics* **63**(3): 443–494.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.

Bickel, P. J. and Levina, E. (2008a). Covariance regularization by thresholding, *Annals of statistics* **36**(6): 2577–2604.

Bickel, P. J. and Levina, E. (2008b). Regularized estimation of large covariance matrices, *Annals of Statistics* **36**(1): 199–227.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *The Review of Economics and Statistics* **72**(3): 498–505.

Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances, *The Journal of Political Economy* **96**(1): 116–131.

Boneva, L., Linton, O. and Vogt, M. (2015). A semiparametric model for heterogeneous panel data with fixed effects, *Journal of Econometrics* **188**(2): 327–345.

Brown, L. D. and Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method, *The Annals of Statistics* **35**(5): 2219–2232.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation, *Journal of the American Statistical Association* **106**(494): 672–684.

Cai, T. T., Levine, M. and Wang, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design, *Journal of Multivariate Analysis* **100**(1): 126–136.

Chen, Z. and Leng, C. (2015). Dynamic covariance models, *Journal of the American Statistical Association* (just-accepted): 1–55.

Chesnay, F. and Jondeau, E. (2001). Does correlation between stock returns really increase during turbulent periods?, *Economic Notes* **30**(1): 53–80.

Diebold, F. X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: A multivariate latent factor ARCH model, *Journal of Applied Econometrics* **4**(1): 1–21.

Doz, C. and Renault, E. (2004). Conditionally heteroskedastic factor models: identification and instrumental variables estimation. CIRANO working paper.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business & Economic Statistics* **20**(3): 339–350.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH, *Econometric theory* **11**(01): 122–150.

Engle, R. F., Ng, V. K. and Rothschild, M. (1990). Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills, *Journal of Econometrics* **45**(1): 213–237.

Erb, C. B., Harvey, C. R. and Viskanta, T. E. (1994). Forecasting international equity correlations, *Financial analysts journal* **50**(6): 32–45.

Fan, J. and Gijbels, I. (1996). Local polynomial modelling and its applications. London: Chapman and Hall.

Fan, J., Wang, M. and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components, *Journal of the Royal Statistical Society Series B* **70**(4): 679–702.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression, *Biometrika* **85**(3): 645–660.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*, Springer Science & Business Media.

Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(2): 303–322.

Fiorentini, G., Sentana, E. and Shephard, N. (2004). Likelihood-based estimation of latent generalized ARCH structures, *Econometrica* **72**(5): 1481–1517.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *The journal of finance* **48**(5): 1779–1801.

Hall, P. and Carroll, R. J. (1989). Variance function estimation in regression: the effect of estimating the mean, *Journal of the Royal Statistical Society. Series B. Methodological* **51**(1): 3–14.

Hall, P., Kay, J. and Titterinton, D. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika* **77**(3): 521–528.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data, *Econometric Theory* **24**(3): 726–748.

Härdle, W. and Marron, J. (1990). Semiparametric comparison of regression curves, *The Annals of Statistics* **18**(1): 63–89.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression, *Journal of Econometrics* **81**(1): 223–242.

James, G. M., Hastie, T. J. and Sugar, C. A. (2000). Principal component models for sparse functional data, *Biometrika* **87**(3): 587–602.

Kaplanis, E. C. (1988). Stability and forecasting of the comovement measures of international stock market returns, *Journal of International Money and Finance* **7**(1): 63–75.

Kring, S., Rachev, S. T., Hochstotter, M. and Fabozzi, F. J. (2007). Composed and factor composed multivariate garch models. University of Karlsruhe Technical Report.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation., *Annals of statistics* **37**(6B): 4254.

Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor garch model, *Journal of Business & Economic Statistics* **25**(1): 61–75.

Levina, E., Rothman, A. and Zhu, J. (2008). Sparse estimation of large covariance matrices via a nested lasso penalty, *The Annals of Applied Statistics* **2**(1): 245–263.

Li, Y., Wang, N. and Carroll, R. J. (2013). Selecting the number of principal components in functional data, *Journal of the American Statistical Association* **108**(504): 1284–1294.

Longin, F. and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960–1990?, *Journal of international money and finance* **14**(1): 3–26.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, *Journal of Finance* **56**(2): 649–676.

Mack, Y. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **61**(3): 405–415.

Magnus, J. R. (1985). On differentiating eigenvalues and eigenvectors, *Econometric Theory* **1**(2): 179–191.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2003). Estimating the error variance in nonparametric regression by a covariate-matched U-statistic, *Statistics: A Journal of Theoretical and Applied Statistics* **37**(3): 179–188.

Munk, A., Bissantz, N., Wagner, T. and Freitag, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(1): 19–41.

Munk, A. and Dette, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory, *The Annals of Statistics* **26**(6): 2339–2368.

Pelletier, D. (2006). Regime switching for dynamic correlations, *Journal of econometrics* **131**(1): 445–473.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation, *Biometrika* **86**(3): 677–690.

Ramchand, L. and Susmel, R. (1998). Volatility and cross correlation across major stock markets, *Journal of Empirical Finance* **5**(4): 397–416.

Rice, J. (1984). Bandwidth choice for nonparametric regression, *Annals of statistics* **12**(4): 1215–1230.

Ruppert, D., Wand, M. P., Holst, U. and Hössjer, O. (1997). Local polynomial variance-function estimation, *Technometrics* **39**(3): 262–273.

Santos, A. A. and Moura, G. V. (2014). Dynamic factor multivariate GARCH model, *Computational Statistics & Data Analysis* **76**(C): 606–617.

Sentana, E. and Fiorentini, G. (2001). Identification, estimation and testing of conditionally heteroskedastic factor models, *Journal of econometrics* **102**(2): 143–164.

Sheppard, K. and Xu, W. (2014). Factor high-frequency based volatility (heavy) models, *Working paper, University of Oxford* .

Silvennoinen, A. and Teräsvirta, T. (2015). Modeling conditional correlations of asset returns: a smooth transition approach, *Econometric Reviews* **34**(1-2): 174–197.

Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, *Journal of Business & Economic Statistics* **20**(3): 351–362.

van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* **17**(5): 549–564.

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance, *The Annals of Mathematical Statistics* **12**(4): 367–395.

Vrontos, I. D., Dellaportas, P. and Politis, D. N. (2003). A full-factor multivariate GARCH model, *The Econometrics Journal* **6**(2): 312–334.

Wang, L., Brown, L. D., Cai, T. T. and Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression, *The Annals of Statistics* **36**(2): 646–664.

Wu, W. B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data, *Biometrika* **90**(4): 831–844.

Yao, Q. and Tong, H. (1994). Quantifying the influence of initial values on nonlinear prediction, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **56**(4): 701–725.

Yin, J., Geng, Z., Li, R. and Wang, H. (2010). Nonparametric covariance model, *Statistica Sinica* **20**: 469–479.