UNIVERSITY OF CALIFORNIA
RIVERSIDE

Essays on Nonparametric and Semiparametric Models and Continuous Time Models

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Yun Wang

June 2012

Dissertation Committee:

Dr. Aman Ullah , Chairperson
Dr. Gloria Gonzalez-Rivera
Dr. Jang-Ting Guo
Dr. Tae-Hwy Lee

The Dissertation of Yun Wang is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would first like to thank my dissertation advisor, Professor Aman Ullah, for his constant support and excellent guidance in developing this thesis. Since my second year at UCR, he was always standing by me and gave me enormous amount of time, patience and assistance in guiding me towards my goals. Whenever I had doubts and questions, Professor Ullah was ready to help me out and encourage me to make further progress. I remember vividly that in spite of his busy schedule, Professor Ullah always sit with me in his office and talk with me for hours in the morning, lunch time, or late afternoon, sometimes even during weekends. Professor Ullah led me into the door of academic career and lightened my research road.

Also, I am indebted to the other members of my dissertation committee, Professor Gloria Gonzalez-Rivera, Professor Jang-Ting Guo, and Professor Tae-Hwy Lee for giving me invaluable suggestions at various stages of my studies at UCR and for all their kind advices and inspirations for the problems I encountered during my stay at Riverside. Whenever I needed, they were ready for helping me and spending their valuable time with me.

I would like to sincerely thank Dr. Liangjun Su, Dr. Yong Bao, and Dr. Jun Yu for their support in developing this thesis. I also thank all professors who taught me during my five-year stay and thank the Economics Department for providing financial and professional assistance to me.

I wish to thank my friends at UCR. It will be too long to list all of them here. However, I would like to give particular thanks to Sarah Lynne Daway. She was always around me and helped me out whenever I needed. My dear friends brought me countless joyance during my stay in California.

Most importantly, I would like to thank my family for the tremendous support, endless care and love. Without you, I would not be able to accomplish all my goals. I wish to thank my parents for everything you've done for me. I want to thank my husband for being so supportive through all my difficulties and struggles. Even though we were apart by the Pacific Ocean, you were always there to encourage me to be strong in hard time and share with me the happiness. Thank you for supporting me and motivating me to do my best.

To my family for all the support

ABSTRACT OF THE DISSERTATION

Essays on Nonparametric and Semiparametric Models and Continuous Time Models

by

Yun Wang

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2012
Dr. Aman Ullah , Chairperson

My dissertation consists of six essays which contribute new theoretical results to two econometrics frontiers: nonparametrics and finite sample econometrics. Chapters 2 to 3 discuss the estimation and inference of the nonparametric and semiparametric models. In chapter 2 an efficient two-step estimator is developed in single nonparametric regression model with a general parametric error covariance. By fully utilizing the information incorporated in the error covariance into estimation, the newly developed method is more efficient compared to the conventional local linear estimator (LLLS) and some other two-step estimator. The corresponding asymptotic theorems are derived. Monte Carlo study shows the relative efficiency gain of the newly proposed estimator. Chapter 3 systematically develops a new set of results for seemingly unrelated regression (SUR) analysis within nonparametric and semiparametric framework. We study the properties of LLLS and local linear weighted least squares (LLWLS) estimators, provide an efficient two-step estimation for the system and establish the asymptotic theorems under both unconditional and conditional error variance-covariance cases. The procedures of estimation for various nonparametric and semiparametric SUR models are proposed. In addition, two nonparametric goodness-of-fit measures for the system are given. Chapter

4 applies the estimation method developed in chapter 2 and 3 to an empirical analysis on return to public capital in U.S.

Chapters 5 to 6 study the finite sample properties of the mean reversion parameter estimator in continuous time models. In chapter 5 we approximate the bias of $\widehat{\kappa}$ for the Lévy-based Ornstein-Uhlenbeck (OU) process, and propose bias corrected estimators of κ. In chapter 6 the exact distribution of the MLE is investigated under different scenarios: known or unknown drift term, fixed or random start-up value, and zero or positive κ. The numerical calculations demonstrate the remarkably reliable performance of the proposed exact approach.

In chapter 7 we study the efficiency of the coefficient of determination based on final prediction error (R^2_{FPE}) and compare it with conventional goodness-of-fit measures (R^2, R^2_a) in linear regression models with both normal and non-normal disturbances. The efficiency results show that R^2_{FPE} has practical use in empirical analysis, for examples, panel data analysis and time series analysis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent decades, nonparametric econometrics is a booming frontier in econometrics analysis. In contrast to traditional parametric econometrics, nonparametric analysis does not require any functional form for underlying regression model, but it is driven by real data. This feature brings to nonparametric analysis an important property of avoiding the unforgiving consequences of parametric misspecification. In other words, nonparametric analysis can lean closer to the reality than parametric analysis does, which encourage me developing new theory within this field. Another field my dissertation focuses on is finite sample econometrics. The use of the asymptotic theory has been popular and well developed in econometric analysis over the past decades. However, asymptotic properties hinge upon an infinite large sample, which generally is not practical in extensive studies, such as economics, psychology, engineering, and sociology and so on. And notice that using asymptotic theory for small or even moderately large samples may cause unpleasant misleading results. Another undesired drawback for the use of asymptotic theory lies in that under some circumstances, various estimators have identical asymptotic distributions. Hence, it is impossible to provide clear preference of one estimator over the other. As a powerful tool to overcome the above shortcomings, finite sample theory becomes one of most important frontiers in modern econometric analysis, which motivates my interest in developing finite sample theory in popularly used econometrics models.

My dissertation mainly covers the following topics: (1) estimation and asymptotic theory in single nonparametric models and a system of multiple equations in nonparametric and semiparametric models; (2) moments approximations and exact distribution of the mean reversion parameter estimator in continuous time models; (3) finite sample properties on coefficient of dertermination based on final prediction errors. The contributions of my work would be twofold. First, it would contribute to theoretical methods. Second, my dissertation research would benefit extensive empirical studies, as it could provide methodological approach to help analyze practical real-world questions. To illustrate the practical use of the newly developed methods, we also conduct an empirical study to address the public capital puzzle.

An interest in the estimation of nonparametric regression relationship by exploring the information in the error covariance has been growing recently. The intuition behind is analogous to the efficient estimators in linear parametric models. As we know, in order to obtain a best linear unbiased estimator (BLUE) in classical linear parametric models, zero conditional mean, uncorrelated relationship, and homoskedasiticity of disturbances are among assumptions to be satisfied. If we suppose that variance-covariance matrix of disturbances $\Omega \neq \sigma^2 I$, we can apply Aitken's generalized least-squares to premultiple the original variables by a matrix P such that $E(P\varepsilon\varepsilon^{'}P^{'}) = P\Omega P^{'} = I$, then a BLUE will yield since the usual assumptions of the least-squares model can be satisfied after this transformation (Zellner, 1962). To obtain more efficient estimators in nonparametric regression analysis, the information enclosed in variance-covariance matrix of disturbances is worth consideration. In the second chapter, we propose a two-step estimator of nonparametric regression function with general parametric error covariance and demonstrate it is more efficient than the usual local linear estimator and some other two-step estimator in literature. This chapter studies the multivariate case for single nonparametric regression, and establish the asymptotic theorem for both mean and slope estimators. A small set of Monte Carlo studies shows the relative efficiency gain of the newly proposed estimator in comparison with LLLS and some other two-step estimator in nonparametric regression with either AR(2) errors or heteroskedastic errors. The theoretical results can be widely applied to a general single nonparametric regression analysis.

Along the line of chapter 2, chapter 3 systematically develops a new set of results for seemingly unrelated regression (SUR) analysis within nonparametric and semiparametric framework. It is well known that the SUR models have been extensively studied in parametric framework and widely used in substantial empirical economic analysis, such as, the wage determinations for different industries, a system of consumer demand equations, and capital asset pricing models, and so on. However, it hasn't been well developed within nonparametric framework. In chapter 3, we study the properties of LLLS and local linear weighted least squares (LLWLS) estimators in nonparametric SUR. To obtain a more efficient estimation, we develop a two-step estimator for the system and establish the corresponding asymptotic theorems under both unconditional and conditional error variance-covariance cases. The procedures of estimation for various nonparametric and semiparametric SUR models are proposed, such as, the NP SUR model with error components, partially linear semiparametric model, model with nonparametric autocorrelated errors, additive nonparametric model, varying coefficient model, and the model with endogeneity. In addition, two nonparametric goodness-of-fit measures for the system are given. To examine the finite sample properties, a small set of Monte Carlo simulations is conducted to compare the newly developed two-step estimator with LLLS, LLWLS estimators, and a class of two-step estimator as well.

To illustrate the practical use of the newly developed methods in chapter 2 and 3, an empirical analysis on return to public capital in U.S. is presented in chapter 4. A panel data for the U.S. 48 contiguous states over the period of 1970-1986 is employed to revisit the public capital puzzle. Is public-sector capital productive? What's the role for public-sector in affecting private economic performance? The debates on these questions have been receiving extensive attention from economists. Under parametric framework, empirical studies reach contrary results by either assuming a particular model specification for the underlying production function or employing various parametric estimation methods. Within the parametric framework constant elasticities of the specified models are assumed across all the states and all the years. The question arises naturally is whether or not the estimates of returns to inputs can be trusted under above settings. As we know, nonparametric method is not only free from the unforgiving misspecification issue, but also provides local estimates so that variety of the estimates of returns to inputs can be observed across all the states and all the years. Upon these properties, nonparametric method would give more precise analysis to address the issue. In this empirical analysis, there are some interesting findings: First, the average returns of public capital on states' private economic growth are statistically significant and positive. In other words, the public capital has positive spill-over effects on average across states, and even though its spill-over effects are smaller than private sector capital stock but still non-negligible. Second, in general, the returns to the public capital are positive. However, a few states, for instances, Wyoming, South Dakota, North Dakota, New Mexico, Montana, have negative returns to the public capital, which are consistent with some recent studies under nonparametric framework. Third, the mean returns to the public capital across all the 48 states changes over the period of 1970-1986. The returns to public capital increased sharply during recessions, started decreasing when the economy stepped into recovering, and fluctuated in small magnitudes during normal time. The reason behind this may be in that when the economy is in recession the private sector becomes weak, so that public sector capital plays a more effective role than normal time. The private sector may gain more benefits from the government investments on the public capital during recessions than the other times.

Chapters 5-7 are developed within the finite sample framework in which we derive and evaluate our moment approximations and exact distribution of the mean reversion parameter estimator ($\widehat{\kappa}$) in continuous time models (also known as diffusion processes). The mean reversion parameter (κ) measures the persistence in the stochastic processes and $1-\kappa$ measures the speed of mean reversion. The smaller value of mean reversion, the higher persistence in the stochastic process, which means the process is more likely to remain in the same state from one observation to the next. In practice, this parameter is of important implications for asset pricing, risk management and forecast. It has been shown in the literature that mean reversion parameter suffers the most serious bias problem among all the parameters in diffusion processes. The difficulty in the estimation of κ is related to the finite sample bias problem well documented in the discrete time literature (see, for instance, Kendall (1954)).

More specifically, chapter 5 considers the bias of the mean reversion estimator ($\widehat{\kappa}$) in the continuous time Lévy processes. In recent years, it has been reported strong evidence of infinite activity jumps in financial variables (see, for example, Aït-Sahalia and Jacod (2008)). It is known that the continuous time Lévy processes not only can capture the infinite activity jumps, but also allow a general form of errors. Due to these features, the Lévy processes have become increasingly popular and various Lévy models have been developed in the asset pricing literature (see for example, Barndorff-Nielsen (1998), Madan, Carr and Chang (1999), Carr and Wu (2003)). Although an extensive literature has developed methods for estimating the parameters in continuous time diffusion models and for approximating the estimation bias, the effect of nonnormality on the estimation has not been studied. The bias of $\widehat{\kappa}$ is approximated and the bias expressions are obtained for the Lévy-based Ornstein-Uhlenbeck (OU) process. The approximate bias of $\widehat{\kappa}$ under normality is also derived as a special case. The bias expressions indicate that both the skewness and the kurtosis of the Lévy measure affect the bias when the time span is not very large and the sampling frequency is not very high. The initial condition, the long term mean (μ), and the volatility parameter (σ) also enter the bias expressions. Bias corrected estimators of κ are proposed. Monte Carlo studies are conducted to examine the performance of the bias corrected estimators.

It has been documented in literature that the maximum likelihood estimator (MLE) of κ tends to over estimate the true value. On one hand, the true distribution of MLE can be severely skewed in finite samples and the asymptotic results in general may provide misleading results. Its asymptotic distribution, on the other hand, depends on how the data are sampled (under expanding, infill, or mixed domain) as well as how we spell out the initial condition. This poses a tremendous challenge to practitioners in terms of estimation and inference. In chapter 6, we investigate the exact distribution of the MLE under different scenarios: known or unknown drift term, fixed or random start-up value, and zero or positive κ. In particular, we employ numerical integration via analytical evaluation of a joint characteristic function. The numerical calculations demonstrate the remarkably reliable performance of the newly proposed exact approach.

Chapter 7 studies the efficiency properties of the coefficient of determination

(R^2_{FPE}) based on final prediction error and compares it with conventional goodness-of-fit measures (R^2, R_a^2) in linear regression models with both normal and non-normal disturbances. The literature has shown that using R-square based on final prediction error as a model selection criterion is perfectly consistent with using AIC and is closest with the criterion BIC than other conventional R-squares. However, there is no theoretical proof on its efficiency properties. Motivated by its good performance and lack of theoretical studies in the literature, I developed this chapter. My theoretical results show it is a useful tool as a model selection and goodness-of-fit measure in both cross-sectional analysis and time series analysis.

Chapter 8 concludes the thesis. The mathematical derivations are provided in the appendix.

Chapter 2

Single Equation Nonparametric Estimation with Non-Scalar Covariance *

2.1 Introduction

As is well known, in order to obtain a best linear unbiased estimator (BLUE) in classical linear parametric models, zero conditional mean, uncorrelated relationship, and homoskedasiticity of disturbances are among assumptions to be satisfied. If we suppose that variance-covariance matrix of disturbances $\Omega \neq \sigma^2 I$, we can apply Aitken's generalized least-squares to premultiple the original variables by a matrix P such that $E(P\varepsilon\varepsilon' P') = P\Omega P' = I$, then a BLUE will yield since the usual assumptions of the least-squares model can be satisfied after this transformation (Zellner, 1962). The intuition behind local linear generalized least squares (LLGLS) estimator is the same as Aitken's generalized least-squares. To consider the contemporaneous correlated disturbances in

*This chapter is a joint work with Dr. Liangjun Su and Dr. Aman Ullah.

nonparametric regression model, the standard procedure is to apply Aitken's generalized least-squares to obtain LLGLS by premultiplying a matrix P such that $E(P\varepsilon\varepsilon' P') = P\Omega P' = I$ on both sides of the weighted equation (Das, 2005). In this spirit, in order to gain the efficiency of estimators, the information enclosed in variance-covariance matrix of disturbances is worth consideration.

Recently more and more interests are growing in the estimation of nonparametric regression relationship by exploring the information in the error covariance. Since conventional local linear least square estimator (LLE) fully ignores the information in the error covariance structure, it is not efficient when the error terms are not independently identically distributed. Ruckstuhl, Welsh and Carroll (2000) considered a semiparametric model as $Y_{ij} = \alpha_i + m(X_{ij}) + \epsilon_{ij}$, $i = 1, \ldots n$, $j = 1, \ldots J$, where α_i and ϵ_{ij} are independent random variables with zero mean and variances $\sigma^2_\alpha > 0$, $\sigma^2_\epsilon > 0$ respectively. Considering the structure of variance component, they obtained a two-step estimator which is more asymptotic efficient than the pooled estimator that ignored the dependence in their semiparametric model. The idea of their two-step estimator is to multiply both sides of the model by the square-root of the inverse covariance matrix to transform the original disturbances to be indepent and identically distributed. Following the basic intuition of Ruckstuhl, Welsh and Carroll (2000) and employing the covariance structure, Su and Ullah (2007) developed a class of two-step estimators in which the regressors were allowed to be a random vector, and two different bandwidth sequences were used in the two steps. In addition, Su and Ullah (2007) also considered a more efficient estimation of the first order derivatives of the nonparametric regression mean. Xiao, Linton, Carroll and Mammen (2003) proposed a kernel-based procedure for local polynomial estimation in Nonparametric regression with autocorrelated errors, which was more efficient than the estimator obtained by ignoring the correlation structure entirely. The intuition behind the procedure proposed by Xiao et al. (2003) is to transform their original model to have uncorrelated disturbances. Unlike Xiao et al. (2003) investigated a stationary case, Linton and Mammen (2008) considered both stationary case and unit root case in disturbances. However, the intuition behind the procedure of Linton and Mammen (2008) was still familiar, that is, they employed a dynamic transformation to make the error term white noise. In addition, when estimating nonparametric function for panel data with measurement error, Lin and Carroll (2000) computed separate regressions at each time period and averaged the weighted resulting estimates, which improved efficiency over a single measurement error analysis by pooling all the panel data. Different with the papers mentioned before where the errors exhibited a parametric correlation structure, Su and Ullah (2006) provided a three-step procedure to let the errors enter the model nonparametrically, and then their baisc model was constructed as $Y_t = m_1(X_t) + m_2(U_{t-1}, \ldots, U_{t-p}) + \varepsilon_t$.

The case considered by Martins-Filho and Yao (2009, MY hereafter) is relatively more general than the above. For nonparametric regression with general parametric error covariance, they have proposed a two-step estimator of nonparametric regression function and demonstrated it is more efficient than the traditional local linear estimator (LLE). Intuitively MY gains the relative efficiency of their estimator over the LLE because the former applies the information in the off-diagonal elements of the error covariance whereas the latter fully ignores the information in the error covariance structure. Nevertheless, MY did not fully explore the information in the diagonal elements of the error covariance. Consequently, if these diagonal elements are not identical across observations (say when the error term is an AR process or heteroskedastic of known form), then their estimator can be further improved.

In this chapter, we propose a modified estimator of MY. We demonstrate clearly that the full use of the error covariance structure can result in an asymptotically more efficient estimator than MY's estimator. The relative efficiency of our estimator over MY's is verified through simulations where the error terms in the nonparametric regression follow an AR(2) process or a heteroskedastic structure. In addition, we extend MY's estimator to the multivariate case, and also establish the asymptotic theorems for the slope estimators which are not studied by MY. To illustrate the applicability of our asymptotic results to popular nonparametric models, we study the asymptotic properties of our two-step estimators for seemingly unrelated regression and clustered/panel data models. Also, the practical use of the newly proposed method is demonstrated within a nonparametric panel regression model with random effects in a real data setting.

The chapter is structured as follows. We introduce the MY's estimator in Section 2 and demonstrate it can easily be improved to achieve a more efficient estimator in Section 3 where the asymptotic bias and variance for the two-step estimator are derived for both seemingly unrelated regression models and clustered/panel data models. A small set of simulations is conducted in Section 4. Finally, the concluding remarks are made in Section 5.

2.2 The MY's estimator

Consider the nonparametric regression model

$$Y_i = m\left(X_i\right) + U_i,\ i = 1, \cdots, n, \tag{2.1}$$

where X_i is a $q \times 1$ vector of exogenous regressors that is continuously distributed, and U_i is an error term such that $E\left(U_i\right) = 0$ and

$$E\left(U_i U_j\right) = \omega_{ij}\left(\theta_0\right) \text{ for some } \theta_0 \in \Re^p,\ i, j = 1, \cdots, n. \tag{2.2}$$

Following MY, we assume for simplicity that $\{U_i\}$ is independent of $\{X_i\}$ but allow for time series structure in either process. In addition, we permit non-identical distributions across i's.

Let $\mathbf{Y} \equiv (Y_1, \cdots, Y_n)'$, $R_{ix} \equiv (1, (X_i - x)')'$, and $\mathbf{R}_x \equiv (R_{1x}, \cdots, R_{nx})'$. Let $\delta(x) \equiv (m(x), \partial m(x)/\partial x')'$. The conventional LLE of $\delta(x)$ is given by

$$\widehat{\delta}_{LL,h_1}(x) = \left(\mathbf{R}_x' \mathbf{K}_{x,h_1} \mathbf{R}_x\right)^{-1} \mathbf{R}_x' \mathbf{K}_{x,h_1} \mathbf{Y} \tag{2.3}$$

where $\mathbf{K}_{x,h_1}$ =diag$(K_{h_1}(X_1 - x), \cdots, K_{h_1}(X_n - x))$, $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1^q$, $K(\cdot)$ is a kernel function, and h_1 is a bandwidth parameter. In particular, the conventional LLE of $m(x)$ is given by

$$\widehat{m}_{LL,h_1}(x) = \mathbf{e}' \left(\mathbf{R}_x' \mathbf{K}_{x,h_1} \mathbf{R}_x\right)^{-1} \mathbf{R}_x' \mathbf{K}_{x,h_1} \mathbf{Y} \tag{2.4}$$

where $\mathbf{e} \equiv (1, 0, \cdots, 0)'$ denotes a $(q+1) \times 1$ vector.

Since $\widehat{m}_{LL,h_1}(x)$ does not explore the information in the error covariance structure, it cannot be asymptotically efficient in any sense. For this reason, MY proposes a two-step estimator of $m(x)$ that applies the information in (5.9). To proceed, let $\Omega(\theta)$ be an $n \times n$ matrix with the (i,j)th element given by $\omega_{ij}(\theta)$. Assume that $\Omega(\theta) = P(\theta) P(\theta)'$ for some square matrix $P(\theta)$. Let $p_{ij}(\theta)$ and $\upsilon_{ij}(\theta)$ denotes the (i,j)th element of $P(\theta)$ and $P(\theta)^{-1}$, respectively. When $\theta = \theta_0$, the true parameter value, we frequently suppress the dependence of these matrices and their elements on θ_0 and, for example, write P for $P(\theta_0)$ and υ_{ij} for $\upsilon_{ij}(\theta_0)$. Let $\mathbf{m} \equiv (m(X_1), \cdots, m(X_n))'$, $\mathbf{U} \equiv (U_1, \cdots, U_n)'$, and $H \equiv$diag$\left(\upsilon_{11}^{-1}, \cdots, \upsilon_{nn}^{-1}\right)$. Define $\mathbf{Z} \equiv HP^{-1}\mathbf{Y} + \left(I_n - HP^{-1}\right)\mathbf{m}$ where I_n is an $n \times n$ identity matrix. Then

$$\mathbf{Z} = \mathbf{m} + \epsilon \text{ with } \epsilon \equiv HP^{-1}\mathbf{U},$$

and it is easy to verify that ϵ has mean 0 and covariance matrix as a diagonal matrix:

$$E\left(\epsilon\epsilon'\right) = H^2 = \text{diag}\left(\upsilon_{11}^{-2}, \cdots, \upsilon_{nn}^{-2}\right). \tag{2.5}$$

The two-step MY's estimators of $\delta(x)$ and $m(x)$ are given by

$$\widehat{\delta}_{MY,h_2}(x) = \left(\mathbf{R}_x'\mathbf{K}_{x,h_2}\mathbf{R}_x\right)^{-1}\mathbf{R}_x'\mathbf{K}_{x,h_2}\widehat{\mathbf{Z}} \tag{2.6}$$

$$\widehat{m}_{MY,h_2}(x) = \mathbf{e}'\left(\mathbf{R}_x'\mathbf{K}_{x,h_2}\mathbf{R}_x\right)^{-1}\mathbf{R}_x'\mathbf{K}_{x,h_2}\widehat{\mathbf{Z}}, \tag{2.7}$$

where $\widehat{\mathbf{Z}} \equiv HP^{-1}\mathbf{Y} + \left(I_n - HP^{-1}\right)\widehat{\mathbf{m}}_{LL,h_1}$, $\widehat{\mathbf{m}}_{LL,h_1} \equiv (\widehat{m}_{LL,h_1}(X_1), \cdots, \widehat{m}_{LL,h_1}(X_n))'$, and the bandwidth h_2 is usually different from h_1. Clearly, here it is assumed that θ_0, and therefore H and P, are known. When θ_0 is unknown but can be estimated by $\widehat{\theta}$ at $\sqrt{n}$-rate, we can replace H and P by $H(\widehat{\theta})$ and $P(\widehat{\theta})$ and it is trivial to show that such a replacement won't affect the first-order asymptotic properties of $\widehat{m}_{MY,h_2}(x)$. Hence it is not restrictive to assume that θ_0 is known.

The following theorem extends the findings in MY to the multivariate case and it also incorporates the asymptotic properties for slope estimators. To proceed with the asymptotic theorem, we first state a list of general assumptions.

Assumption A1. $K(\cdot)$ is a product kernel such that $K(x) = \Pi_{i=1}^q k(x_i)$ where $k(\cdot)$ is a univariate symmetric kernel with compact support S_k satisifying: (i) $\int k(x_i)dx_i = 1$; (ii) $\int x_i k(x_i)dx_i = 0$; (iii) $\int x_i^2 k(x_i)dx_i = \sigma_k^2$; (iv) for all x_i,$x_i^{'} \in S_k$ we have $|k(x_i) - k(x_i')| \leq c|x_i - x_i'|, c \in [0, \infty)$.

Assumption A2. (i) $f_i(x, \theta_0)$ is the marginal density of X_i evaluated at x, with $f_i(x, \theta_0)$ $< c$ for all i; (ii) $\bar{f}(x) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n f_i(x, \theta_0)$, and $0 < \bar{f}(x) < \infty$; (iii) $f_i(x, \theta_0)$ is differentiable, and $|f_i^{(1)}(x, \theta_0)| < c$; (iv) $|f_i(x, \theta_0) - f_i(x', \theta_0)| \leq c|x - x'|$ for all x, x', and θ_0, where θ_0 denotes the true parameters.

Assumption A3. $m^{\alpha}(x) < c$ for all x and $\alpha = 1, 2$, $m^{\alpha}(x)$ is the αth-order derivative of $m(x)$ evaluated at x.

Assumption A4. As $n \to \infty$, $h \to 0, nh^{q+2} \to \infty$ and $nh^{q+6} \to 0$.

Assumption A5. $\overline{\omega}_f(x, \theta_0) \equiv \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \upsilon_{ii}^{-2} f_i(x)$, and $0 < \overline{\omega}_f(x, \theta_0) < \infty$, where υ_{ii} is the diagonal element of H.

Assumption A6. $\omega_f^*(x, \theta_0) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \upsilon_{ii}^2 f_i(x)$, and $0 < \omega_f^*(x, \theta_0) < \infty$.

To compare the asymptotic efficiency among LLE and two-step estimators, we first present the asymptotic distribution of LLE. Let $f_i(x)$ denote the marginal density of X_i, $\overline{f}(x) \equiv \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f_i(x)$, and $\omega_f(x, \theta_0) \equiv \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \omega_{ii}(\theta_0) f_i(x)$.

Theorem 1 *Assume that Assumptions A1-A4 are met, we have*

$$\sqrt{nh_2^q} D_h \left(\widehat{\delta}_{LL,h}(x) - \delta(x) - \begin{pmatrix} \frac{\kappa_{21} h_2^2}{2} \sum_{j=1}^{q} \frac{\partial^2 m(x)}{\partial x_j^2} \\ 0_{q\times 1} \end{pmatrix} \right) \xrightarrow{d} N(0, \Omega_{LL}).$$

where $\Omega_{LL} = \begin{pmatrix} \frac{\omega_f(x,\theta_0)(\kappa_{02})^q}{\overline{f}^2(x)} & 0' \\ 0 & \frac{\omega_f(x,\theta_0)\kappa_{22}(\kappa_{02})^{q-1}}{\overline{f}^2(x)\kappa_{21}^2} I_q \end{pmatrix}$, $D_h = diag(1, h_2, \cdots, h_2)_{(q+1)\times(q+1)}$, *and* $\kappa_{ij} = \int z^i k(z)^j dz$ *for* $i, j = 0, 1, 2.$*where* $\kappa_{ij} = \int z^i k(z)^j dz$ *for* $i, j = 0, 1, 2.$

The asymptotic distribution for the two-step estimator is given in the following Theorem.

Theorem 2 *Assume that Assumptions A1-A5 are met, we have*

$$\sqrt{nh_2^q} D_{h_2} \left(\widehat{\delta}_{MY,h_2}(x) - \delta(x) - B_{MY} \right) \xrightarrow{d} N(0, \Omega_{MY}).$$

where

$$B_{MY} = \begin{pmatrix} \frac{\kappa_{21} h_2^2}{2} \sum_{j=1}^{q} \frac{\partial^2 m(x)}{\partial x_j^2} \\ \mathbf{0}_{q\times 1} \end{pmatrix}, \quad \Omega_{MY} = \begin{pmatrix} \frac{\overline{\omega}_f(x,\theta_0)(\kappa_{02})^q}{\overline{f}^2(x)} & \mathbf{0}_{1\times q} \\ \mathbf{0}_{q\times 1} & \frac{\overline{\omega}_f(x,\theta_0)\kappa_{22}(\kappa_{02})^{q-1}}{\overline{f}^2(x)\kappa_{21}^2} I_q \end{pmatrix},$$

$D_{h_2} = diag(1, h_2, \cdots, h_2)$ *is a* $(q+1) \times (q+1)$ *diagonal matrix, and* $\kappa_{ij} = \int z^i k(z)^j dz$ *for* $i, j = 0, 1, 2$.

The proof of the above theorem follows straightforwardly from that of Theorem 3 in MY, is similar to that of Theorem 3 below and thus omitted. To obtain the above result, a necessary condition on (h_1, h_2) is that $h_1/h_2 \to 0$ in order to eliminate the first order asymptotic bias due to the first stage estimation error. Also, in order for the remainder term from the second-order Taylor expansion of $m(X_i)$ at x vanish asymptotically, we need $\limsup_{n\to\infty} nh_2^{q+4} \to c \in [0, \infty)$.

Let $\widehat{\beta}_{MY,h_2}(x)$ denote the vector of the last q elements of $\widehat{\delta}_{MY,h_2}(x)$. Theorem 2 implies that

$$\sqrt{nh_2^q}\left(\widehat{m}_{MY,h_2}(x) - m(x) - \frac{\kappa_{21}h_2^2}{2}\sum_{j=1}^{q}\frac{\partial^2 m(x)}{\partial x_j^2}\right) \overset{d}{\to} N\left(0, \frac{\overline{\omega}_f(x,\theta_0)(\kappa_{02})^q}{\overline{f}^2(x)}\right),$$

$$\sqrt{nh_2^q}h_2\left(\widehat{\beta}_{MY,h_2}(x) - \frac{\partial m(x)}{\partial x}\right) \overset{d}{\to} N\left(0, \frac{\overline{\omega}_f(x,\theta_0)\kappa_{22}(\kappa_{02})^{q-1}}{\overline{f}^2(x)\kappa_{21}^2}I_q\right).$$

It is easy to see that $\widehat{m}_{MY,h_2}(x)$ shares the same asymptotic bias as the traditional LLE $\widehat{\mathbf{m}}_{LL,h_2}(x)$ but has smaller asymptotic variance than the latter. To see this, note that the asymptotic variance of $\widehat{\mathbf{m}}_{LL,h_2}(x)$ is given by

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\omega_{ii}(\theta_0) f_i(x)(\kappa_{02})^q/\overline{f}^2(x).$$

By the fact that for any nonsingular matrix A with inverse A^{-1}, we have $a_{ii}a^{ii} \geq 1$ $\forall i$ with a_{ii} and a^{ii} being the ith diagonal elements of A and A^{-1} respectively, we can readily show that

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\omega_{ii}(\theta_0) f_i(x) - \overline{\omega}_f(x,\theta_0) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\left(\omega_{ii}(\theta_0) - v_{ii}^{-2}\right) f_i(x) \geq 0.$$

That is, $\widehat{m}_{MY,h_2}(x)$ is asymptotically more efficient than $\widehat{m}_{LL,h_2}(x)$. By the same token, $\widehat{\beta}_{MY,h_2}(x)$ shares the same asymptotic bias as the traditional LLE of $\partial m(x)/\partial x$ but has smaller asymptotic variance than the latter.

2.3 A more efficient two-step estimator

In this section we first demonstrate that the MY's estimator can be improved to obtain a more efficient estimator and then consider applying our estimation method to both seemingly unrelated regression models and panel data models.

2.3.1 A more efficient two-step estimator

As indicated in the introduction, MY's estimator does not fully use the information in the diagonal elements of the error covariance matrix $\Omega(\theta_0)$ into estimation. Note that the transformed errors by MY's method is non-scalar, when there is serial correlation and/or heteroskedasticity in original errors. So it still has a room to improve MY's two-step estimator. Apparently, the cause of the lack of efficiency of MY's estimator is due to the misuse of the diagonal matrix H in the definition of $\mathbf{Z}$. It turns out that we can modify the definition of $\mathbf{Z}$ to obtain a more efficient estimator. Let $\mathbf{Z}^* \equiv H^{-1}\mathbf{Z}$. Then

$$\mathbf{Z}^* = H^{-1}\mathbf{m} + \epsilon^* \text{ with } \epsilon^* \equiv P^{-1}\mathbf{U}. \tag{2.8}$$

Clearly, ϵ^* has mean 0 and covariance matrix as an identity matrix. We can consider the local linear estimation of $\delta(x)$ based on the transformed equation in (2.8).

It is straightforward to verify that our two-step estimator of $\delta(x)$ based on (2.8) is given by

$$\widehat{\delta}_{SUW,h_2}(x) \equiv \left(\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{R}_x^*\right)^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\widehat{\mathbf{Z}}^* \tag{2.9}$$

where $\mathbf{R}_x^* \equiv H^{-1}\mathbf{R}_x$, and $\widehat{\mathbf{Z}}^* \equiv P^{-1}\mathbf{Y} + \left(H^{-1} - P^{-1}\right)\widehat{\mathbf{m}}_{LL,h_1}$. Then we have the following theorem.

Theorem 3 *Assume that Assumptions A1-A4 and A6 are met, we have*

$$\sqrt{nh_2^q}D_{h_2}\left(\widehat{\delta}_{SUW,h_2}(x) - \delta(x) - B_{SUW}\right) \xrightarrow{d} N\left(0, \Omega_{SUW}\right),$$

where

$$B_{SUW} = \begin{pmatrix} \frac{\kappa_{21}h_2^2}{2}\sum_{j=1}^{q}\frac{\partial^2 m(x)}{\partial x_j^2} \\ \\ \mathbf{0}_{q\times 1}\end{pmatrix},\ \Omega_{SUW} = \begin{pmatrix} \frac{(\kappa_{02})^q}{\omega_f^*(x,\theta_0)} & \mathbf{0}_{1\times q} \\ \\ \mathbf{0}_{q\times 1} & \frac{\kappa_{22}(\kappa_{02})^{q-1}}{\omega_f^*(x,\theta_0)\kappa_{21}^2}I_q\end{pmatrix},$$

and $\omega_f^*(x,\theta_0) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\upsilon_{ii}^2 f_i(x)$.

The proof of the above theorem is delegated to the section I of Appendix A. Theorem 3, in conjunction with Theorem 2, implies that $\widehat{\delta}_{SUW,h_2}(x)$ shares the same asymptotic bias as MY's estimator $\widehat{\delta}_{MY,h_2}(x)$. To compare their asymptotic covariances, noting that by the Cauchy-Schwarz inequality, we have

$$\frac{1}{n^{-1}\sum_{i=1}^{n}\upsilon_{ii}^2 f_i(x)} \le \frac{n^{-1}\sum_{i=1}^{n}\upsilon_{ii}^{-2} f_i(x)}{\{n^{-1}\sum_{i=1}^{n} f_i(x)\}^2},$$

which implies that $1/\omega_f^*(x,\theta_0) \le \overline{\omega}_f(x,\theta_0)/\overline{f}^2(x)$. Thus, the asymptotic covariance of $\widehat{\delta}_{SUW,h_2}(x)$ is less than that of $\widehat{\delta}_{MY,h_2}(x)$. That is, our two-stage estimator may have smaller asymptotic variance than MY's if a non-negligible portion of the diagonal elements are distinct from others. In other words, it pays off to explore the information in the diagonal elements of the error covariance matrix.

2.3.2 Two-step estimator for clustered or panel data models

In order to illustrate the applicability of our theorems to popular nonparametric models, we derive the asymptotic bias and variances of our two-step estimator for clustered or panel data models. The panel data model has been studied in MY for the univariate case.

Here we consider the following one-way random effects model

$$Y_{ij} = m(X_{ij}) + \alpha_i + \varepsilon_{ij},\ i = 1, \cdots, n,\ j = 1, \cdots, J,$$

where X_{ij} is a $q \times 1$ vector of exogenous variables, α_i is independent and identically distributed (IID) $(0, \sigma_\alpha^2)$, ε_{ij} is IID $(0, \sigma_\varepsilon^2)$, α_i and ε_{lj} are uncorrelated for all

$i, l = 1, 2, ..., n$, and $m(\cdot)$ is an unknown smooth function. Let $u_{ij} = \alpha_i + \varepsilon_{ij}$, $u_i \equiv (u_{i1}, ..., u_{iJ})'$, and $u \equiv (u_1, ..., u_n)'$. By assumption, we have $\Sigma \equiv E(u_i u_i') = \sigma_\varepsilon^2 I_J + \sigma_\alpha^2 1_J 1_J'$ and $\Omega(\sigma_\varepsilon^2, \sigma_\alpha^2) \equiv E(uu') = I_n \otimes \Sigma$, where 1_J is a $J \times 1$ vector of ones. As in MY, assuming that $\Omega = PP'$ for some square matrix P, then $P^{-1} = I_n \otimes V^{-1/2}$, where $V^{-1/2} = (v_{ij})_{i,j=1,...,J}$ with $v_{ii} \equiv v = \frac{1}{\sigma_\varepsilon} - (1 - \frac{\sigma_\varepsilon}{\sigma_1})\frac{1}{J\sigma_\varepsilon}$ for all $i = 1, ..., J$, $v_{ij} = v_o = -(1 - \frac{\sigma_\varepsilon}{\sigma_1})\frac{1}{J\sigma_\varepsilon}$ for all $i \neq j = 1, ..., J$, and $\sigma_1 = \sqrt{J\sigma_\alpha^2 + \sigma_\varepsilon^2}$. Our two-step estimator is $\widehat{\delta}_{SUW,h_2}(x) = (\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{R}_x^{*\prime})^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\widehat{\mathbf{Z}}^*$, where $\mathbf{R}_x^* \equiv H^{-1}\mathbf{R}_x$, $\mathbf{R}_x \equiv (X_{x,11}, ..., X_{x,1J}, ..., X_{x,n1}, ..., X_{x,nJ})$, $X_{x,ij} = (1, (X_{ij} - x)')'$, $\mathbf{K}_{x,h_2} \equiv \text{diag}(K_{h_2}(X_{11} - x)$, ..., $K_{h_2}(X_{1J} - x)$, ..., $K_{h_2}(X_{n1} - x)$, ..., $K_{h_2}(X_{nJ} - x))$, and $\widehat{\mathbf{Z}}^*$ is analogously defined as in Section 3.1. Then Theorem 3 implies that

$$\sqrt{nh_2^q} D_h \left(\widehat{\delta}_{SUW,h_2}(x) - \delta(x) - B^{(Panel)}\right) \xrightarrow{d} N\left(0, \Omega^{(Panel)}\right) \tag{2.10}$$

where $B^{(Panel)} = \begin{pmatrix} \frac{\kappa_{21}h_2^2}{2}\sum_{s=1}^q \frac{\partial^2 m(x)}{\partial x_s^2} \\ \mathbf{0}_{q\times 1} \end{pmatrix}$, $\Omega^{(Panel)} = \begin{pmatrix} \frac{(\kappa_{02})^q}{v^2\sum_{j=1}^J f_j(x)} & \mathbf{0}_{1\times q} \\ \mathbf{0}_{q\times 1} & \frac{\kappa_{22}(\kappa_{02})^{q-1}}{v^2\sum_{j=1}^J f_j(x)\kappa_{21}^2} I_q \end{pmatrix}$, and $f_j(\cdot)$ denotes the marginal density of X_{ij}.

2.4 Monte Carlo simulations

Now we conduct a small set of Monte Carlo simulations to compare the finite sample performance of our estimator with that of LLE and MY. Consider the following data generating process:

$$Y_i = m(X_i) + U_i,\ i = 1, \cdots, n,$$

where the univariate random variables X_i are first generated independently from $N(0, 1)$ and then truncated at ± 3. We use two specifications for $m(x)$: $0.5 + e^{-4x}/(1 + e^{-4x})$ and $1 - 0.9e^{-2x^2}$, which correspond respectively to $m_2(x)$ and $m_3(x)$ in MY.

For the error terms, we consider two cases. In Case 1, we assume a time series structure for U_i and generate U_i from the following AR(2) process: $U_i = 0.5U_{i-1} - 0.4U_{i-2} + \varepsilon_i$, where ε_i are independently and identically distributed (IID) $N(0, 1)$. In Case 2, we assume that U_i are heteroskedastic but independent of each other, and generate U_i, $i = 1, \cdots, \frac{n}{2}$, as IID from $N(0, 2)$, and U_i, $i = \frac{n}{2} + 1, \cdots, n$, as IID from $N(0, 4)$. In the first case, only the first two diagonal elements in the square root matrix (P) of the covariance matrix (Ω) of $\mathbf{U} \equiv (U_1, \cdots, U_n)'$ are distinct from others, so that the MY and SUW estimators are asymptotically equivalent and we should not observe significant difference in the finite sample performance between the two estimators. In the second case, however, the LLE and MY estimators are asymptotically equivalent and both are dominated by the SUW estimator.

For all estimators, we use the Gaussian kernel. For bandwidth sequences, we use the least-squares cross validation to choose h_2, and set $h_1 = h_2^{5/4}$, where h_1 and h_2 are used in the first and second step estimations, respectively, for both MY and SUW estimators. The one-step LLE estimator uses h_2 in the estimation.

Although we know the covariance matrix Ω of $\mathbf{U}$ in the simulation, we estimate it according to the AR(2) specification in Case 1 and heteroskedastic specification in Case 2. To be specific, we estimate the two autoregressive coefficients in the first case and the two variances in the second case. Based on the estimation of $m(x)$ on all data points $X_1, \cdots, X_n$, we calculate the bias, standard deviation (Std), root mean squared error (RMSE), and mean squared error (MSE) for each estimator and average them across 1000 replications. The sample sizes under our investigation are 100 and 200.

Table 2.1 on page 22 reports the finite sample performance for the three estimators for both $m(x)$ and $\partial m(x)/\partial x$ in the case of AR(2) errors. First, in terms of Std and RMSE (or MSE), both MY and SUW estimators outperform the LLE estimator, and have smaller bias for estimators of $\partial m(x)/\partial x$, but the former tend to have slightly larger biases for estimating $m(x)$. Second, as expected the efficiency gain of the SUW estimator over the MY estimator is tiny and may be ignored in the AR(2) error structure. Noting that the more different diagonal elements in the square root matrix P of Ω, the more efficiency gain we may have, we expect that prominent efficiency gain can be achieved only in AR(p) model with $p \equiv p(n) \to \infty$ as $n \to \infty$ or in ARMA(p, q)-type of models.

Table 2.2 on page 23 compares the three estimators for both $m(x)$ and $\partial m(x)/\partial x$ under the heteroskedastic errors [1]. As expected, the performance of the MY estimator is identical to that of LLE given the fact that they share the same first order asymptotic bias and variance. Obviously, our estimator SUW has improvement over LLE and MY in the sense of having lower Std and RMSE (or MSE). The simulation results provide a strong support that the SUW estimator is more efficient than the LLE and MY estimators by considering heterogeneity in the error structure.

2.5 Concluding Remarks

In this chapter we propose a two-step estimator (SUW) for nonparametric regression with a general parametric error covariance that is more efficient than that of MY's. The results are applied to one-way random effects model. Notice that by the transformation which we employ to obtain our two-step estimator the transformed errors has spherical parametric covariance structure. Therefore, intuitively SUW estimator should outperform those nonparametric regression estimators that fail to fully utilize the information in the error covariance. Simulations confirm the finite sample out-

[1] The results in Table 2.2 are obtained for the heteroskedastic error case with two different variances. We also did the simulations for the case with four different variances, and observed higher relative efficiency gain of SUW over MY compared to the former case.

performance of our estimator over both LLE and MY's under both serial correlation case and heteroskedastic case. Notice that under heteroskedasticity MY's estimator degenerates to LLE as the former fails to incorporate the diagonal information in the error covariance, which is also confirmed in Monte Carlo simulations.

Table 2.1: Comparison of various estimators of $m(x)$ and $\partial m(x)/\partial x$ for AR(2) errors

n	Estimators	Specification 1				Specification 2			
		Bias	Std	RMSE	MSE	Bias	Std	RMSE	MSE
					$m(x)$				
100	LLE	0.0081	0.2422	0.2681	0.0719	0.0193	0.2657	0.3031	0.0919
	MY	-0.0084	0.2141	0.2355	0.0555	0.0221	0.2386	0.2748	0.0755
	SUW	-0.0086	0.2113	0.2307	0.0532	0.0212	0.2372	0.2734	0.0748
200	LLE	-0.0009	0.1825	0.2067	0.0427	0.0231	0.1913	0.2274	0.0517
	MY	-0.0008	0.1619	0.1829	0.0335	0.0253	0.1697	0.2028	0.0411
	SUW	-0.0009	0.1600	0.1793	0.0322	0.0254	0.1689	0.2021	0.0409
					$\partial m(x)/\partial x$				
100	LLE	-0.0136	0.9198	1.0064	1.0129	0.0132	1.0658	1.1774	1.3864
	MY	-0.0119	0.7799	0.8666	0.7510	0.0128	0.9214	1.0287	1.0582
	SUW	-0.0116	0.7799	0.8659	0.7497	0.0118	0.9177	1.0263	1.0532
200	LLE	-0.0151	0.7521	0.8223	0.6762	-0.0056	0.7289	0.8178	0.6689
	MY	-0.0109	0.6421	0.7129	0.5082	-0.0051	0.6099	0.6936	0.4811
	SUW	-0.0108	0.6400	0.7086	0.5022	-0.0052	0.6074	0.6911	0.4777

Table 2.2: Comparison of various estimators of $m(x)$ and $\partial m(x)/\partial x$ for heteroskedastic errors

n	Estimators	Specification 1				Specification 2			
		Bias	Std	RMSE	MSE	Bias	Std	RMSE	MSE
					$m(x)$				
100	LLE/MY	-0.0049	0.5923	0.6708	0.4500	0.0499	0.6032	0.6437	0.4144
	SUW	-0.0119	0.5147	0.6002	0.3602	0.0467	0.5314	0.5796	0.3359
200	LLE/MY	-0.0063	0.4679	0.5489	0.3013	0.0386	0.4736	0.5845	0.3416
	SUW	-0.0078	0.4040	0.4872	0.2373	0.0402	0.4098	0.5261	0.2768
					$\partial m(x)/\partial x$				
100	LLE/MY	-0.0198	1.6925	1.9024	3.6189	0.0055	1.9187	2.0116	4.0465
	SUW	-0.0173	1.4692	1.7182	2.9522	0.0044	1.7060	1.8271	3.3382
200	LLE/MY	0.0051	1.8507	2.1445	4.5991	-0.0227	1.3191	1.4887	2.2163
	SUW	0.0059	1.5985	1.9353	3.7455	-0.0135	1.1327	1.3250	1.7555

Chapter 3

System of Equations and Panel Model Nonparametric and Semiparametric Estimation

3.1 Introduction

The advantages of using nonparametric method lie in that not only it is free from misspecification issue of functional form, but also it gives local estimations which can provide more deep information than parametric method does. It is well known that the weighted least squares (WLS, also known as GLS) estimator in a parametric regression model with a known non-scalar covariance matrix of errors, is the best linear unbiased estimator. This also holds asymptotically for operational WLS estimator in which the non-scalar covariance matrix is replaced by a consistent estimator, see Greene (2007, p.157) and Hayashi (2000, p.138). Further, in small samples it is known to be unbiased for the symmetric errors, see Kakwani (1967), and its efficiency properties are analyzed in Taylor (1977). In the case of nonparametric regression model with a non-scalar covariance, various local linear weighted least squares (LLWLS) estimators have been developed for the pointwise local linear regression and its derivative estimators, see Welsh and Yee (2006), Ullah and Roy (1998), Henderson and Ullah (2005 , 2008), Lin and Carroll (2000), among others. However, it has been shown in Henderson and Ullah (2008), Welsh and Yee (2006), and Lin and Carroll (2000), among others, that such LL-WLS estimators may not be efficient even when the covariance matrix is known. In fact, often they are even beaten by the local linear least squares (LLLS) estimator ignoring the existence of a non-scalar covariance matrix. In view of this Ruckstuhl, Welsh, and Carroll (2000) proposed a two-step estimator in which the dependent variable, which is filtered (transformed), with the mean as the regression function and the non-scalar covariance matrix transformed to a scalar covariance matrix, also see Su and Ullah (2007). Martins-Filho and Yao (2009) estimated the filtered dependent variable with its mean as the regression function but a non-scalar covariance matrix consisting heteroscedasticity. Su, Ullah and Wang (2011) then suggested a new two-step estimator in which the filtered dependent variable has a mean with transformed regression and a scalar covariance matrix. They showed that their two-step estimator is asymptotically more efficient than both the LLLS and the two-step estimator proposed by Martins-Filho and Yao (2009). In a simulation study they also show that their two-step estimator is also more efficient compared to both the LLLS and the Martins-Filho and Yao's two-step estimator.

The objective of this chapter is to systematically develop the theory and application of two-step estimation in the context of the seemingly unrelated regression (SUR) models. As we know, the SUR models have been extensively studied in parametric framework and widely used in substantial empirical economic analysis, such as, the wage determinations for different industries, a system of consumer demand equations, and capital asset pricing models, and so on. However, it hasn't been well developed within nonparametric framework, although see, for example, Smith and Kohn (2000), and Koop, Poier and Tobias (2005), where nonparametric Bayesian methods is used to estimate multiple equations, Wang, Guo, and Brown (2000) where a penalized spline estimation method is considered, and Welsh and Yee (2006) where LLWLS estimators are used.

This chapter develops a new set of results for SUR regression analysis within nonparametric and semiparametric framework. Specifically, we study the properties of conventional LLLS and LLWLS in nonparametric SUR, and develop efficient two-step estimation for nonparametric SUR following Su, Ullah, and Wang (2011) in the context of single equation model. The corresponding asymptotic theorems under both unconditional and conditional error variance-covariance cases are established. Then we compare its asymptotic properties with the LLLS and LLWLS estimators. The theoretical results show that our two-step estimator is more asymptotically efficient than LLLS. It is known that various nonparametric and semiparametric specifications have been developed and widely used within cross-sectional models or panel data models, and the corresponding estimation and statistical properties have been well discussed in literature. However, these specifications haven't been considered in NP and SP SUR models. It would be also interesting to know the estimation and statistical inference for different specifications within NP and SP SUR system. Hence, the procedures of estimation for various nonparametric and semiparametric SUR models are proposed in the current chapter, such as, the model with error components, partially linear semiparametric model, additive nonparametric model, varying coefficient model, and the model with endogeneity. In addition, two nonparametric goodness-of-fit measures for the system are given as well, which provide a fundamental knowledge that can be used to develop various tests based on R-square for SUR system. To examine the finite sample properties, we conduct a small set of Monte Carlo simulations to compare our two-step estimator with LLLS, LLWLS estimators, and a class of other two-step estimators as well. The latter can be shown as a special case of ours. The simulation results confirm that our two-step estimator outperforms others in the finite sample settings.

The structure of this chapter is as follows. In section 2, we introduce SUR NP estimations including LLLS estimator, a general two-step estimator, and provide their asymptotic distributions under unconditional error variance-covariance. In addition, various LLWLS estimators are discussed. In section 3 we propose the estimation procedures for a variety of popular NP/SP SUR functions, specifically, partially linear semiparametric model, model with NP autocorrelated errors, additive NP models, varying coefficient NP models, varying coefficient IV models, and NP SUR models with error components. Section 4 discusses NP SUR models with conditional error covariance, and its estimation incorporating the conditional covariance. The corresponding asymptotic distribution is also provided. In section 5 we define two types of nonparametric Goodness-of-fit measures in terms of ANOVA decomposition and indicator function. The following section 6 conducts a small set of Monte Carlo simulations to examine the finite sample performance of LLLS, LLWLS, and two-step estimators.

3.2 Nonparametric Seemingly Unrelated Regression System

We start with the following basic nonparametric seemingly unrelated regression models

$$y_{ij} = m_i(X_{ij}) + u_{ij},\ i = 1, ..., M,\ j = 1, \cdots, N. \tag{3.1}$$

The economic variable y_{ij} is the jth observation on the ith cross-sectional unit, X_{ij} is the jth observation on the ith unit. $m_i(\cdot)$ is an unknown function form, which can differ across the cross-sectional units. The observation y_{ij} is related to X_{ij}, a $q_i \times 1$ vector of exogenous regressors, and X_{ij}, $i = 1, \cdots M$, can differ for different regression models. For the present, we assume strict exogeneity of X_{ij}, $E(u_{ij}|X_{ij}) = 0$, and homoscedasticity $Var(u_{ij}|X_{ij}) = \sigma_{ii}^2$ within each equation. Also, we assume that the disturbances are uncorrelated across observations but correlated across equations, i.e, $E(u_{ij}u_{i'j}|X_{ij}, X_{i'j}) = \sigma_{ii'}$ for $i, i' = 1, \cdots, M$ and $i \neq i'$, and $j = 1, \cdots, N$. For simpliticy, the fixed number of observations N is assumed. However, it can be extended to unequal numbers of observations.

The economic examples of such models include: (i) economic growth model in which i stands for different countries, j indexes the time periods, specifically, y_{ij} is the growth variable for the jth time period on the ith country, and X_{ij} is a vector of regressors that affect the economic growth of ith country at the jth period; (ii) regional consumption model in which i denotes the ith cluster, j denotes the jth household; (iii) the wage determination for different industries, in which we can set different equations for different industries, that is, i indexes the ith industry, j is the jth observation; (iv) a system of consumer demand equations on a panel data set, etc. In a special case, $m_i(X_{ij}) = X_{ij}\beta_i$ that is the standard Zellner's (1962) parametric SUR system.

3.2.1 Estimation with Unconditional Error Variance-Covariance Ω

In this section, we introduce the local linear least squares estimator (LLLS), propose a more efficient two-step estimator in a general form, and also discuss the properties of LLWLS estimators within SUR. The asymptotic distributions for both LLLS and two-step estimators are given for multivariate nonparametric SUR models.

3.2.1.1 Local Linear Least Squares Estimator

By first order Taylor expansion, we write

$$\begin{aligned} y_{ij} &= m_i(X_{ij}) + u_{ij} \\ &\simeq m_i(x_i) + (X_{ij} - x_i)m_i^{(1)}(x_i) + u_{ij} \\ &= \begin{pmatrix} 1 & (X_{ij} - x_i)' \end{pmatrix} \begin{pmatrix} m_i(x_i) \\ m_i^{(1)}(x_i) \end{pmatrix} + u_{ij} \\ &= Z_{ij}(x_i)\delta_i(x_i) + u_{ij}, \end{aligned}$$

where $\delta_i(x_i) = \begin{pmatrix} m_i(x_i) & m_i^{(1)\prime}(x_i) \end{pmatrix}'$, which is a $(q_i + 1) \times 1$ vector, and $Z_{ij}(x_i) = \begin{pmatrix} 1 & (X_{ij} - x_i)' \end{pmatrix}$. Let $y_i = (y_{i1}, ...y_{iN})'$, $Z_i(x_i) = \begin{pmatrix} Z_{i1}(x_i), & \cdots, & Z_{iN}(x_i) \end{pmatrix}'$, which has a dimension of $N \times (q_i+1)$, and $u_i = \begin{pmatrix} u_{i1}, & \cdots, & u_{iN} \end{pmatrix}'$. In a vector representation, for each regression i, we can write

$$y_i \simeq Z_i(x_i)\delta_i(x_i) + u_i.$$

Further, one can stack regression $i = 1, \cdots, M$, in a matrix version,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} Z_1(x_1) & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & Z_M(x_M) \end{pmatrix} \begin{pmatrix} \delta_1(x_1) \\ \vdots \\ \delta_M(x_M) \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_M \end{pmatrix},$$

which can be written compactly as

$$\begin{aligned} \mathbf{y} &= \mathbf{m}(X) + \mathbf{u} \\ &\simeq Z(x)\delta(x) + \mathbf{u}, \end{aligned} \tag{3.2}$$

where $\mathbf{y} = \left(y_1^{'}, \ldots, y_M^{'}\right)'$, is a $MN \times 1$ vector, $\mathbf{m}(X) = (\mathbf{m}_1(X_1), ..., \mathbf{m}_M(X_M))'$, $\mathbf{m}_i(X_i) = (m_i(X_{i1}), ..., m_i(X_{iN}))'$, $\mathbf{u} = \left(u_1^{'}, ..., u_M^{'}\right)'$,

$$Z(x) = diag\left(Z_1(x_1), \quad .... \quad Z_M(x_M)\right),$$

which has $MN \times (\Sigma_{i=1}^{M} q_i + M)$ dimension, and $\delta(x) = \left(\delta_1(x_1), \quad .... \quad \delta_M(x_M)\right)$, a $(\Sigma_{i=1}^{M} q_i + M) \times 1$ vector. By the assumption of basic SUR models, we have $E\left(\mathbf{u}|\mathbf{X}\right) = \mathbf{0}_{MN\times 1}$ and $\Omega \equiv$Var$(\mathbf{u}|\mathbf{X}) = \Sigma \otimes I_N$, where Σ is a $M \times M$ matrix with typical diagonal element σ_{ii}^2 and off-diagonal element $\sigma_{ii'}$ for i, $i' = 1, ..., M$. Then the local linear estimator of $\delta(x)$ is obtained by minimizing $\mathbf{u}'K(x)\mathbf{u}$,

$$\hat{\delta}(x) = (Z'(x)K(x)Z(x))^{-1}Z'(x)K(x)\mathbf{y},$$

where $K(x) \equiv diag\left(K_{h_1}(X_1 - x_1), \quad .... \quad K_{h_M}(X_M - x_M)\right)$, is a $MN \times MN$ diagonal matrix, $K_{h_i}(X_i - x_i) \equiv diag\left(K_{h_i}(X_{i1} - x_i), \quad .... \quad K_{h_i}(X_{iN} - x_i)\right)$ and $K_{h_i}(X_{ij} - x_i) = \frac{1}{h_i}k(\frac{X_{ij}-x_i}{h_i})$.

Assumption A1. $K(\cdot)$ is a product kernel such that $K(x_i) = \Pi_{s=1}^{q} k(x_{is})$ where $k(\cdot)$ is a univariate symmetric kernel with compact support S_k satisifying: (i) $\int k(x_{is})dx_{is} = 1$; (ii) $\int x_{is}k(x_{is})dx_{is} = 0$; (iii) $\int x_{is}^2 k(x_{is})dx_{is} = \sigma_k^2$; (iv) for all x_{is},$x_{is}' \in S_k$ we have $|k(x_{is}) - k(x_{is}')| \leq c|x_{is} - x_{is}'|, c \in [0, \infty)$.

Assumption A2. (i) $f_{ij}(x_i, \theta_0)$ is the marginal density of X_{ij} evaluated at x_i, with $f_{ij}(x_i, \theta_0) < c$ for all i, j, and x_i; (ii) $\bar{f}_i\left(x_i, \theta_0\right) = \lim_{N\to\infty} N^{-1}\sum_{j=1}^{N} f_{ij}(x_i, \theta_0)$, and $0 < \bar{f}_i\left(x\right) < \infty$; (iii) $f_{ij}(x_i, \theta_0)$ is differentiable, and $|f_{ij}^{(1)}(x_i, \theta_0)| < c$; (iv)

$|f_{ij}(x_i,\theta_0) - f_{ij}(x'_i,\theta_0)| \leq c|x_i - x_{i'}|$ for all x_i, $x_{i'}$, and θ_0, where θ_0 denotes the true parameters.

Assumption A3. $g_j(x_i, x_{i'})$ denotes a joint density of $\left(X_{ij}, X_{i'j}\right)$ evaluated at $(x_i, x_{i'})$. The partial derivatives of $g_j(x_i, x_{i'})$ exist and are continuous.

Assumption A4. $m(x)$ is two times differentiable.

Assumption A5. As $n \to \infty$, $h_i \to 0, nh_i^{q_i+2} \to \infty$ and $nh_i^{q_i+6} \to 0$.

Theorem 4 *Under the assumptions A1-A5 and the assumptions on the error terms of basic SUR models, we have*

$$D(\hat{\delta}(x) - \delta(x) - B_{LLLS}) \xrightarrow{d} N\left(0, \Omega_{LLLS}\right)$$

where $D \equiv diag\begin{pmatrix} D_1, \cdots, & D_M \end{pmatrix}$, $D_i = \sqrt{Nh_i^{q_i}} D_{h_i}$, $D_{h_i} \equiv diag(1, h_i, ..., h_i)$ *is a* $(1+q_i)\times(1+q_i)$ *diagonal matrix,* $B_{LLLS} = \begin{pmatrix} B_{1,LLLS}, & \ldots, & B_{M,LLLS} \end{pmatrix}'$,
$\Omega_{LLLS} = diag\begin{pmatrix} \Omega_{1,LLLS}, & \ldots, & \Omega_{M,LLLS} \end{pmatrix}$,

$$B_{i,LLLS} = \begin{pmatrix} \frac{k_{21}h_i^2}{2} \sum_{s=1}^{q_i} \frac{\partial^2 m_i(x_i)}{\partial x_{i,s}^2} \\ \mathbf{0}_{q_i \times 1} \end{pmatrix},$$

$$\Omega_{i,LLLS} = \begin{pmatrix} \frac{\sigma_{ii}^2 (\kappa_{02})^{q_i}}{f_i(x_i)} & \mathbf{0}'_{1\times q_i} \\ \mathbf{0}_{q_i\times 1} & \frac{\sigma_{ii}^2 \kappa_{22} (\kappa_{02})^{q_i - 1}}{f_i(x_i)\kappa_{21}^2} I_{q_i} \end{pmatrix},$$

and x_{is} *is the sth element of* x_i *for* $i = 1, \ldots, M$.

Remark 1 We allow the marginal distribution of x differ across equations and across observations in each equation. The average of the densities must converge.

Remark 2 Notice that local linear least squares method doesn't incorporate the covariance into estimation, hence, the asymptotic distribution of LLLS for the whole

SUR system actually is the same as the ones for each eqution regression by LLLS. Since the asymptotic covariances across different equations are smaller order than the asymptotic variances, the off-diagonals in the asymptotic variance-covariance matrix are zero.

3.2.1.2 Two-step Estimator

To utilize the information incorporated in the variance-covariance of errors, we propose the following two-step estimator to improve the estimation. The transformation required for the second step is made as follows

$$\begin{aligned} \mathbf{y} &= \mathbf{m}(X) + \mathbf{u} \\ \Omega^{-1/2}\mathbf{y} + (\mathbf{H}^{-1} - \Omega^{-1/2})\mathbf{m}(X) &= \mathbf{H}^{-1}\mathbf{m}(X) + \Omega^{-1/2}\mathbf{u} \\ \vec{\mathbf{y}} &= \mathbf{H}^{-1}\mathbf{m}(X) + \mathbf{v} \\ &= \mathbf{H}^{-1}Z(x)\delta(x) + \mathbf{v}, \end{aligned} \tag{3.3}$$

where $\vec{\mathbf{y}} \equiv \Omega^{-1/2}\mathbf{y} + (\mathbf{H}^{-1} - \Omega^{-1/2})\mathbf{m}(X)$, $\mathbf{v} \equiv \Omega^{-1/2}\mathbf{u}$. It is clear to see that the transformed errors are now independent and identically distributed. The intuition behind the above transformation is similar with the parametric GLS in which X is standardized by the standard deviation of errors. Here, we use $\mathbf{H}$ to standardize unknown function $\mathbf{m}(X)$. For example, when there is no correlation across errors, i.e., $\Omega = diag\left(\sigma_{ii}^2\right)_{i=1}^{MN}$, then $v_{i,i} = 1/\sigma_{ii}$, and the unknown function $\mathbf{m}(X)$ is standadized by the standard errors. At point X_{ij}, the transformed unknown function is $m_i(X_{ij})/\sigma_{ii}$. If Ω is not a diagonal matrix, $\mathbf{H}$ will take care of both the variance and covariance of errors.

Assume that $\Omega = PP'$ for some $MN \times MN$ matrix P. Let p_{ij} and v_{ij} denote the (i, j)th element of P and P^{-1}, respectively. Let $\mathbf{H} \equiv$diag$(v_{1,1}^{-1}, ..., v_{MN,MN}^{-1})$, $R^*(x) =$

$\mathbf{H}^{-1}Z(x)$, then by minimizing $\mathbf{v}'K(x)\mathbf{v}$ the two-step estimator would be

$$\hat{\delta}_{2-step}(x) = (R^{*\prime}(x)K(x)R^*(x))^{-1}R^{*\prime}(x)K(x)\vec{\mathbf{y}}. \tag{3.4}$$

Even though the two-step estimator described above has the same form as Su, Ullah, and Wang (2011), the interpretaion here is different from that paper. In Su, Ullah, and Wang (2011), the two-step estimator that incorporated a general parametric covariance into estimation is motivated by improving the one proposed by Martins-Filho and Yao (2009) which failed to consider the unconditional heteroskedastic errors. The transformation proposed by Martins-Filho and Yao (2009) is as

$$\mathbf{H}P^{-1}\mathbf{Y} + \left(I - \mathbf{H}P^{-1}\right)\mathbf{m}(X) = \mathbf{m}(\mathbf{X}) + \epsilon,$$

where $\epsilon \equiv \mathbf{H}P^{-1}\mathbf{u}$. Obviously, the covariance matrix of transformed errors is a diagonal matrix: $E\left(\epsilon\epsilon'\right) = \mathbf{H}^2$, which consists heteroskedasticity. Also, it is interesting to notice that if the errors are uncorrelated across equations, and $K(x) \to K(0)$, the nonparametric two-step estimator $\hat{\delta}_{2-step}$ will become the parametric GLS estimator. To derive the asymptotic distribution for the two-step estimator, we need additional assumption.

Assumption A6. $\omega^*_{f,i}\left(x_i, \theta_0\right) = \lim_{N\to\infty} N^{-1}\sum_{j=1}^{N} \upsilon^2_{(i-1)N+j} f_{ij}(x_i)$, and $0 < \omega^*_{f,i}\left(x_i, \theta_0\right) < \infty$, for every x_i, and θ_0, where $\upsilon_{(i-1)N+j}$ is the diagonal element of $\mathbf{H}$.

The asymptotic distribution for the two-step estimator is given in the following Theorem.

Theorem 5 *Under the assumptions A1-A6 and the assumptions on the error terms of basic SUR models, we have*

$$D\left(\hat{\delta}_{2-step}(x) - \delta\left(x\right) - B_{2-step}\right) \overset{d}{\to} N\left(0, \Omega_{2-step}\right)$$

where $D \equiv diag\left(D_1, \cdots, \quad D_M\right)$, $D_i = \sqrt{Nh_i^{q_i}} D_{h_i}$, $D_{h_i} \equiv diag(1, h_i, ..., h_i)$, *a* $(1+q_i) \times (1+q_i)$ *diagonal matrix,* $B_{2-step} = \left(B_{1,2-step}, \quad \ldots, \quad B_{M,2-step}\right)'$,

$$
\begin{aligned}
\Omega_{2-step} &= diag\left(\Omega_{1,2-step}, \quad \ldots, \quad \Omega_{M,2-step}\right), \\
B_{i,2-step} &= \begin{pmatrix} \frac{k_{21}h_i^2}{2} \sum_{s=1}^{q_i} \frac{\partial^2 m_i(x_i)}{\partial x_{is}^2} \\ \mathbf{0}_{q_i \times 1} \end{pmatrix}, \\
\Omega_{i,2-step} &= \begin{pmatrix} \frac{(\kappa_{02})^{q_i}}{\omega_{f,i}^*(x_i,\theta_0)} & \mathbf{0}_{1 \times q_i} \\ \mathbf{0}_{q_i \times 1} & \frac{\kappa_{22}(\kappa_{02})^{q_i - 1}}{\omega_{f,i}^*(x_i,\theta_0)\kappa_{21}^2} I_{q_i} \end{pmatrix},
\end{aligned}
$$

and x_{is} *is the sth element of* x_i *for* $i = 1, \ldots, M$.

To compare the efficiency of the two-step estimator with LLLS, we need compare $\frac{1}{\omega_{f,i}^*(x_i,\theta_0)}$ with $\frac{\sigma_{ii}^2}{f_i(x_i,\theta_0)}$. By the fact that for any nonsingular matrix A with inverse A^{-1}, we have $a_{ii}a^{ii} \geq 1 \forall i$, where a_{ii} and a^{ii} are the $i$the diagonal elements of A and A^{-1} respectively, we can readily show that

$$\lim_{N \to \infty} N^{-1} \sum_{j=1}^{N} (\upsilon_{(i-1)N+j,(i-1)N+j}^2 - \sigma_{ii}^{-2}) f_{ij}(x_i) \geq 0.$$

That is, $\hat{\delta}_{2-step}(x)$ is asymptotically more efficient than $\widehat{\delta}_{LLLS}(x)$.

Remark 3 So far it is assumed that Ω, H and P, are known. When Ω, H and P, are unknown but can be estimated at $\sqrt{N}-$rate we can replace them by $\hat{\Omega}$, $\hat{H}$ and $\hat{P}$ and it is trivial to show that such a replacement won't affect the first-order asymptotic properties of those above estimators. Hence it is not restrictive to assume that Ω, H and P, are known.

Two-step Estimator for SUR model with M =2 To simplify the notation and give a more specific two-step estimator, we will focus on the case with $J = 2$. Let

$\mathbf{y} \equiv \left(y^{(1)\prime}, y^{(2)\prime}\right)'$, $X_{i,x_j} \equiv (1, (X_i^{(j)} - x_j)')'$, $\mathbf{X}_{x_j}^{(j)} \equiv \left(X_{1,x_j}, ..., X_{n,x_j}\right)'$,

$$\mathbf{X}_x^* \equiv \begin{pmatrix} \mathbf{X}_{x_1}^{(1)} & \mathbf{0}_{n\times(q_2+1)} \\ \mathbf{0}_{n\times(q_1+1)} & \mathbf{X}_{x_2}^{(2)} \end{pmatrix},$$

and $\Sigma \equiv \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{22}\rho \\ \sigma_{11}\sigma_{22}\rho & \sigma_{22}^2 \end{pmatrix}$. As before, we can obtain the conventional LLE as $\hat{\delta}_{LL} = (\mathbf{X}_x^{*\prime}\mathbf{K}\mathbf{X}_x^*)^{-1}\mathbf{X}_x^{*\prime}\mathbf{K}\mathbf{y}$, where $\mathbf{K}$ =diag$(\mathbf{K}_1, \mathbf{K}_2)$, and $\mathbf{K}_j =$ diag$(K_{h_{1j}}(X_1^{(j)} - x_j), \cdots K_{h_{1j}}(X_n^{(j)} - x_j))$ for $j = 1, 2$. Similarly, assume that $\Omega = PP'$ for some $2n \times 2n$ matrix P. Let p_{ij} and v_{ij} denote the (i, j)th element of P and P^{-1}, respectively. Let $H \equiv$diag$(v_{1,1}^{-1}, ..., v_{2n,2n}^{-1})$. By Cholesky decomposition we have

$$P^{-1} = \Omega^{-1/2} = \begin{pmatrix} \left(\sigma_{11}\sqrt{1-\rho^2}\right)^{-1} I_n & -\rho\left(\sigma_{22}\sqrt{1-\rho^2}\right)^{-1} I_n \\ \mathbf{0}_{n\times n} & \sigma_{22}^{-1} I_n \end{pmatrix},$$

i.e., $v_{ii} = 1/\left(\sigma_{11}\sqrt{1-\rho^2}\right)$ and $v_{n+i,n+i} = 1/\sigma_{22}$ for $i = 1, \cdots, n$.

Let $\delta(x) = (m_1(x_1), \partial m_1(x_1)/\partial x_1', m_2(x_2), \partial m_2(x_2)/\partial x_2')'$ where x is a disjoint union of x_1 and x_2. And $\mathbf{K}$ =diag$(\mathbf{K}_1, \mathbf{K}_2)$, and $\mathbf{K}_j$ =diag$(K_{h_{2j}}(X_1^{(j)} - x_j), \cdots K_{h_{2j}}(X_n^{(j)} - x_j))$ for $j = 1, 2$. Notice that the bandwidth h_{2j} is used in the second step. Applying our two-step estimator to the seemingly unrelated regression models yields the following estimator of $\delta(x)$:

$$\widehat{\delta}_{SUW}(x) = \left(\mathbf{R}_x^{*\prime}\mathbf{K}\mathbf{R}_x^*\right)^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}\widehat{\mathbf{Z}}^* \tag{3.5}$$

where $\mathbf{R}_x^*$ =diag$(H_1^{-1}\mathbf{X}_{x_1}^{(1)}, H_2^{-1}\mathbf{X}_{x_2}^{(2)})$, H_1 =diag$\left(v_{11}^{-1}, ..., v_{nn}^{-1}\right)$,
H_2 =diag$\left(v_{n+1,n+1}^{-1}, ..., v_{2n,2n}^{-1}\right)$, and $\widehat{\mathbf{Z}}^* \equiv P^{-1}\mathbf{Y} + \left(H^{-1} - P^{-1}\right)\widehat{\mathbf{m}}_{LL,h_1}$. Then we have

$$\widetilde{D}\left(\widehat{\delta}_{SUW}(x) - \delta(x) - B^{(SUR)}\right) \overset{d}{\to} N\left(0, \Omega^{(SUR)}\right) \tag{3.6}$$

where $\widetilde{D} \equiv$diag$\left(\widetilde{D}_{h_{21}}, \widetilde{D}_{h_{22}}\right)$, $\widetilde{D}_{h_{2j}} = \sqrt{nh_{2j}^q}$diag$(1, h_{2j}, ..., h_{2j})$ is a $(1 + q_j) \times (1 + q_j)$

diagonal matrix, $B^{(SUR)} = \begin{pmatrix} B_1^{(SUR)} \\ B_2^{(SUR)} \end{pmatrix}$, $\Omega^{(SUR)} = \begin{pmatrix} \Omega_1^{(SUR)} & \mathbf{0}_{(1+q_1)\times(1+q_2)} \\ \mathbf{0}_{(1+q_2)\times(1+q_1)} & \Omega_2^{(SUR)} \end{pmatrix}$,

$B_j^{(SUR)} = \begin{pmatrix} \frac{k_{21}h_{2j}^2}{2}\sum_{s=1}^{q_j}\frac{\partial^2 m_j(x_j)}{\partial x_{js}^2} \\ \mathbf{0}_{q_j\times 1} \end{pmatrix}$, $\Omega_j^{(SUR)} = \begin{pmatrix} \frac{(\kappa_{02})^{q_j}}{\omega_{f,j}^*(x,\theta_0)} & \mathbf{0}_{1\times q_j} \\ \mathbf{0}_{q_j\times 1} & \frac{\kappa_{22}(\kappa_{02})^{q_j-1}}{\omega_{f,j}^*(x,\theta_0)\kappa_{21}^2} I_{q_j} \end{pmatrix}$, $\omega_{f,1}^*(x,\theta_0) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \upsilon_{ii}^2 \times f_i(x_1)$, and $\omega_{f,2}^*(x,\theta_0) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \upsilon_{n+i,n+i}^2 f_i(x_2)$, and x_{js} is the sth element of x_j for $j = 1$ and 2.

A Special Case of Two-step Estimator Ruckstuhl, Welsh and Carroll (2000) proposed a class of two-step estimator for nonparametric panel data models with random effects as follows

$$\begin{aligned} \mathbf{y} &= \mathbf{m}(X) + \mathbf{u} \\ \tau\Omega^{-1/2}\mathbf{y} + (\mathbf{I}-\tau\Omega^{-1/2})\mathbf{m}(X) &= \mathbf{m}(X) + \tau\Omega^{-1/2}\mathbf{u} \qquad (3.7) \\ \mathbf{y}^* &= \mathbf{m}(X) + \mathbf{u}^*, \mathbf{u}^* \equiv \tau\Omega^{-1/2}\mathbf{u}. \end{aligned}$$

By minimizing $\mathbf{u}^{*\prime}K(x)\mathbf{u}^*$, their two-step estimator can be obtained as

$$\hat{\delta}_\tau(x) = (Z'(x)K(x)Z(x))^{-1}Z'(x)K(x)\mathbf{y}^*.$$

Su and Ullah (2007) follow the same idea of Ruckstuhl, Welsh, and Carroll (2000), propose the above $\tau-$type two-step estimator and provide its asymptotic normality and the optimal τ. Notice that the class of two-step estimator in Ruckstuhl, Welsh, and Carroll (2000) and Su and Ullah (2007) is a special case of ours. Let $\mathbf{H} = \tau I$, I is an indentity matrice, then our method in (3.3) can be written as

$$\Omega^{-1/2}\mathbf{y} + (\tau^{-1}\mathbf{I}-\Omega^{-1/2})\mathbf{m}(X) = \tau^{-1}\mathbf{m}(X) + \Omega^{-1/2}\mathbf{u}.$$

We multiply τ on both sides of the above equation, then it becomes

$$\tau\Omega^{-1/2}\mathbf{y} + (\mathbf{I}-\tau\Omega^{-1/2})\mathbf{m}(X) = \mathbf{m}(X) + \tau\Omega^{-1/2}\mathbf{u}.$$

Notice that $\mathbf{H} = \tau I$ implies that all the diagonal elements in $\Omega^{-1/2}$ contain identical information, that is $v_{ii} = \tau^{-1}$ for $i = 1, \ldots, MN$. However, by our settings, $\mathbf{H}$ can incorporate both heteroskedastic and correlation infromation in errors. Hence, our method actually generalizes the class of $\tau-$type two-step estimator. The corresponding asymptotic properties of $\hat{\delta}_\tau(x)$ for nonparametric SUR models can be modified from the ones of $\hat{\delta}_{2-step}(x)$, and it is given in the following theorem.

Theorem 6 *Under the same conditions as stated in Theorem 5, we have*

$$D(\hat{\delta}_\tau(x) - \delta(x) - B_\tau) \xrightarrow{d} N\left(0, \Omega_\tau\right),$$

where $D \equiv diag\left(D_1, \cdots, \quad D_M\right)$, $D_i = \sqrt{Nh_i^{q_i}} D_{h_i}$, $D_{h_i} \equiv diag(1, h_i, ..., h_i)$ *is a* $(1+q_i) \times (1+q_i)$ *diagonal matrix,* $B_\tau = \left(B_{1,\tau}, \quad \ldots, \quad B_{M,\tau}\right)'$, $\Omega_\tau = diag\left(\Omega_{1,\tau}, \quad \ldots, \quad \Omega_{M,\tau}\right)$,

$$B_{i,\tau} = \begin{pmatrix} \frac{k_{21} h_i^2}{2} \sum_{s=1}^{q_i} \frac{\partial^2 m_i(x_i)}{\partial x_{i,s}^2} \\ \mathbf{0}_{q_i \times 1} \end{pmatrix},$$

$$\Omega_{i,LLLS} = \begin{pmatrix} \frac{\tau^2 (\kappa_{02})^{q_i}}{f_i(x_i,\theta_0)} & \mathbf{0}'_{1\times q_i} \\ \mathbf{0}_{q_i \times 1} & \frac{\tau^2 \kappa_{22} (\kappa_{02})^{q_i - 1}}{f_i(x_i,\theta_0)\kappa_{21}^2} I_{q_i} \end{pmatrix},$$

and x_{is} *is the sth element of* x_i *for* $i = 1, \ldots, M$.

The optimal τ can be obtained by minimizing the mean squared error of $\hat{\delta}_\tau(x)$. To compare the efficiency of this class of two-step estimator with LLLS, we need compare $\frac{\tau^2}{f_i(x_i,\theta_0)}$ with $\frac{\sigma_{ii}^2}{f_i(x_i,\theta_0)}$. As long as $\tau^2 \leq \sigma_{ii}^2$ for $i = 1, \ldots, M$, the two-step estimator is more efficient than LLLS. Since this class of $\tau-$type two-step estimator is a special case of ours, we will just focus on the generalized two-step estimator $\hat{\delta}_{2-step}(x)$ in the remainder of the present chapter.

Operational two-step estimator The two-step estimator proposed in the previous sections is infeasible, since $\overrightarrow{\mathbf{y}}$ is unobservable, and Ω and $\mathbf{H}$ are unknown. In this section, we introduce the operational two-step estimator for nonparametric SUR models in (3.6). The procedure is described as follows:

(1) First, obtain a preliminary consistent estimator of m_i by first-order local polynomial smoothing y_{ij} on X_{ij} for each equation i. Denote $\hat{u}_{ij} = y_{ij} - \hat{m}_i(X_{ij})$.

(2) Second, we can obtain a consistent estimator of $\hat{\Omega}$, $\hat{H}$ by estimating

$$\begin{aligned}\hat{\sigma}_{ii'} &= \frac{1}{N-1}\sum_{j=1}^{N}\left(\hat{u}_{ij} - \overline{\hat{u}}_{ij}\right)\left(\hat{u}_{i'j} - \overline{\hat{u}}_{i'j}\right), \\ \hat{\sigma}_{ii}^2 &= \frac{1}{N-1}\sum_{j=1}^{N}\left(\hat{u}_{ij} - \overline{\hat{u}}_{ij}\right)^2.\end{aligned}$$

Further we can obtain the feasible $\overrightarrow{\mathbf{y}} = \hat{\Omega}^{-1/2}\mathbf{y} + (\hat{\mathbf{H}}^{-1} - \hat{\Omega}^{-1/2})\hat{\mathbf{m}}(X)$.

(3) Third, by first-order local polynomial soomthing feasible $\overrightarrow{\mathbf{y}}$ on X, obtain the two-step estimator $\hat{\delta}_{2-step}(x) = (R^{*\prime}(x)K(x)R^*(x))^{-1}R^{*\prime}(x)K(x)\overrightarrow{\mathbf{y}}$.

So far, our estimator is based on unconditional covariance. The two-step estimator can also be extended to the nonparametric SUR models with the conditional covariance. Later on we will discuss the estimation of nonparametric SUR with conditional covariance, and the method to obtain a conditional covariance matrix within this framework.

3.2.1.3 Local Linear Weighted Least Squares Estimator

Another popular class of local linear estimator in nonparametric literature is called local linear weighted least squares (LLWLS) estimator. By minimizing the following weighted sum of squared residuals

$$(\mathbf{y} - Z(x)\delta(x))'W_r(x)(\mathbf{y} - Z(x)\delta(x)),$$

the LLWLS can be obtained as

$$\hat{\delta}_r(x) = (Z'(x)W_r(x)Z(x))^{-1}Z'(x)W_r(x)\mathbf{y},$$

where $W_r(x)$ is kernel based weight matrix. For $r = 1, 2, 3, 4$,

$$W_1(x) = K^{1/2}(x)\Omega^{-1}K^{1/2}(x),$$

$$W_2(x) = \Omega^{-1}K(x),$$

$$W_3(x) = K(x)\Omega^{-1},$$

$$W_4(x) = \Omega^{-1/2}K(x)\Omega^{-1/2}.$$

$W_1(x)$ and $W_2(x)$ are given in Lin and Carroll (2000) for nonparametric panel data models with random effect, $W_4(x)$ is discussed in Ullah and Roy (1998) for fixed effect models. Henderson and Ullah (2005) considered $W_1(x)$, $W_2(x)$ and $W_4(x)$ for nonparametric random effect model, and proposed the corresponding feasible estimators. Welsh and Yee (2006) give all these four types of LLWLS estimators, but only study the bias and variance of LLWLS estimator $\hat{\delta}_1(x)$ with weight $W_1(x)$ for a SUR with $M = 2$ for both unconditonal and conditional variance-covariance of errors.

Comparing Welsh and Yee (2006)'s SUR model with ours, there are two differences on the assumptions. One difference is that the SUR model considered by Welsh and Yee assumes heteroskedastic errors in each equation, but we assume the homoskedastic errors in each equation which is the assumption made by extensive literature on classic SUR models. The other different assumption is that they assume an independent and identically joint distribution of (X_{1j}, X_{2j}) across observations. However, we allow different marginal density across equations and observations, and different independent joint distribution across equations. In addition, Welsh and Yee (2006) doesn't give the asymptotic distribution for LLWLS estimators. Hence, it is not appropriate to directly compare the bias and variance of LLWLS in their paper with those properties of our two-step estimator. However, by further examining the bias and variance of their estimator, it is not difficult to see that the bias of the first derivatives estimator $\hat{m}^{(1)}(x)$

in $\hat{\delta}_1(x)$ would have smaller order than the bias of $\hat{m}(x)$, and the covariance of $\hat{\delta}_1(x)$ would be smaller order than the variance of $\hat{m}(x)$ and $\hat{m}^{(1)}(x)$ when we consider the corresponding converge rates, that is, $\sqrt{Nh^q}$ for $\hat{m}(x)$, and $\sqrt{Nh^q}h$ for $\hat{m}^{(1)}(x)$. Also, we notice that the LLWLS estimator $\hat{\delta}_1(x)$ gains no efficiency compared to the LLLS when the errors are modified to satisfy the conventional assumption in SUR models, which is also pointed out in Remark 3 in their paper. Therefore, we can expect the efficiency gains of our two-step estimator over the LLWLS estimators under the conventional assumptions in SUR models. Henderson and Ullah (2008) compare the efficiency among LLWLS estimators and various two-step estimators by simulations, and also find that the latter outperform the former. When the assumptions on errors allow heteroskedasticity within each equation and different correlation across equations, it is difficult to compare the efficiency of LLWLS with LLLS for both conditional variance-covariance and unconditional case. As Welsh and Yee (2006) mentioned, the LLWLS estimator may be less efficient than LLLS in this scenario.

As we know, within parametric framework, if there is no correlation across equations, the generalized least squares estimator (GLS) doesn't have efficiency gain over least squares estimator (LS), even though the heteroskedasticity exists across the equations. In the following, we examine whether or not this result holds for nonparametric SUR models. Suppose there is no correlation across regressions, we can write $\Omega = diag\begin{pmatrix}\Omega_1 & ,\ldots, & \Omega_M\end{pmatrix}$, $\Omega_i = \sigma_i^2 I_N$, for $i = 1, \ldots, M$. That is, we allow heteroskedasticity across the regressions, but no correlations are allowed. Since now the variance-covariance matrix is a diagonal matrix, then $W_1(x) = W_2(x) = W_3(x) = W_4(x)$, four estimators are equivalent. For simplicity, we take $W_1(x) = K^{1/2}(x)\Omega^{-1}K^{1/2}(x)$ as an example to show the relationship between LLWLS and LLLS under the above settings. Since all the matices are diagonal in $\hat{\delta}_1(x)$, we can write the LLWLS separately for each regression. The LLWLS for the ith regression is

$$\begin{aligned}
\hat{\delta}_{1,i}(x) &= (Z_i'(x)W_{1,i}(x)Z_i(x))^{-1}Z_i'(x)W_{1,i}(x)\mathbf{y}_i \\
&= (Z_i'(x)K_i^{1/2}(x)\Omega_i^{-1}K_i^{1/2}(x)Z_i(x))^{-1}Z_i'(x)K_i^{1/2}(x)\Omega_i^{-1}K_i^{1/2}(x)\mathbf{y}_i \\
&= (Z_i'(x)K_i(x)Z_i(x))^{-1}Z_i'(x)K_i(x)\mathbf{y}_i \\
&= \hat{\delta}_{LLLS,i}(x),
\end{aligned}$$

where $\hat{\delta}_{LLLS,i}(x)$ is the LLLS estimator for the ith regression. The above equation gives mathematical equivalance between LLLS and LLWLS when there is no correlation across regressions in nonparametric SUR system. The simulation results conducted later in the present chapter also confirms this equivalence between LLWLS and LLLS. Furthermore, when Ω is an identity matrix, obviously four LLWLS estimators become LLLS estimator, that is, $\hat{\delta}_r(x) = \hat{\delta}(x)$ for $r = 1, 2, 3, 4$.

In addition, like parametric SUR models, if the equations have identical explanatory variables, i.e., $X_i = X_j$, then LLLS and LLWLS are identical. The following examines $W_1(x) = K^{1/2}(x)\Omega^{-1}K^{1/2}(x)$ case. The cases with $W_2(x) = \Omega^{-1}K(x)$, $W_3(x) = K(x)\Omega^{-1}$, and $W_4(x) = \Omega^{-1/2}K(x)\Omega^{-1/2}$, follow the similar proof. Now let $X_i = X_j = X$, hence, $K_i = K_j = \overline{K}$, $Z_i(x_i) = Z_j(x_j) = \overline{Z}$. Then the LLWLS can be written as

$$
\begin{aligned}
\hat{\delta}_1(x) &= (Z'(x)K^{1/2}(x)\Omega^{-1}K^{1/2}(x)Z(x))^{-1}Z'(x)K^{1/2}(x)\Omega^{-1}K^{1/2}(x)\mathbf{y} \\
&= \begin{pmatrix} \sigma^{11}\overline{Z}'\overline{KZ} & \cdots & \sigma^{1M}\overline{Z}'\overline{KZ} \\ \vdots & \ddots & \vdots \\ \sigma^{M1}\overline{Z}'\overline{KZ} & \cdots & \sigma^{MM}\overline{Z}'\overline{KZ} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{M} \sigma^{1j}\overline{Z}'\overline{K}y_j \\ \vdots \\ \sum_{j=1}^{M} \sigma^{Mj}\overline{Z}'\overline{K}y_j \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11}\left(\overline{Z}'\overline{KZ}\right)^{-1} & \cdots & \sigma_{1M}\left(\overline{Z}'\overline{KZ}\right)^{-1} \\ \vdots & \ddots & \vdots \\ \sigma_{M1}\left(\overline{Z}'\overline{KZ}\right)^{-1} & \cdots & \sigma_{MM}\left(\overline{Z}'\overline{KZ}\right)^{-1} \end{pmatrix} \begin{pmatrix} \left(\overline{Z}'\overline{KZ}\right)\sum_{l=1}^{M} \sigma^{1l}\hat{\delta}_{LLLS,l}(x) \\ \vdots \\ \left(\overline{Z}'\overline{KZ}\right)\sum_{l=1}^{M} \sigma^{Ml}\hat{\delta}_{LLLS,l}(x) \end{pmatrix},
\end{aligned}
$$

where σ^{ij} is the (i,j)th element in Σ^{-1}, and σ_{ij} is the (i,j)th element in Σ. We can have the ith LLWLS $\hat{\delta}_1(x)$ is

$$
\begin{aligned}
\hat{\delta}_{i,1}(x) &= \sum_{j=1}^{M}\sigma_{ij}\sum_{l=1}^{M}\sigma^{jl}\hat{\delta}_{LLLS,l}(x) \\
&= \hat{\delta}_{LLLS,1}(x)\sum_{j=1}^{M}\sigma_{ij}\sigma^{j1} + \cdots + \hat{\delta}_{LLLS,M}(x)\sum_{j=1}^{M}\sigma_{ij}\sigma^{jM} \\
&= \hat{\delta}_{LLLS,i}(x).
\end{aligned}
$$

The last equality holds since $\Sigma\Sigma^{-1} = I$.

3.2.2 Alternative Specifications of NP/SP SUR Models

Up to now all estimators are discussed for the basic NP SUR models. In reality, we may have various specifications for the system. For examples, partially linear semiparametric model, additive nonparametric model, varying coefficient model, model with endogeneity, and error components models, etc. These models are well discussed in either cross-sectional or panel data framework. However, within SUR system framework, they haven't been studied. Since all these specifications have practical use in extensive empirical analysis, it is worth to provide theoretical results for these models within

SUR framework. This section is devoted to propose an efficient estimation for various specifications of NP/SP SUR models.

3.2.2.1 Partially Linear Semiparametric SUR Models

We consider the partially linear semiparametric SUR models

$$y_{ij} = m_i(X_{ij}) + Z_{ij}\delta_i + u_{ij},\ i = 1, ..., M,\ j = 1, \cdots, N, \tag{3.8}$$

the assumptions on errors remain the same as in (3.6). One way to estimate (3.8) is using profile least squares method as the following:

$$\begin{aligned} y_{ij} - Z_{ij}\delta_i &= m_i(X_{ij}) + u_{ij} \\ y^*_{ij} &\equiv y_{ij} - Z_{ij}\delta_i \\ y^*_{ij} &= m_i(X_{ij}) + u_{ij}. \end{aligned}$$

Let $y^*_i = (y^*_{i1}, ...y^*_{iN})'$, $\mathbf{y}^* = (y^*_1, ...y^*_M)'$, we stack the models into

$$\mathbf{y}^* = \mathbf{m}(X) + u.$$

By the first order Taylor expansion, we local linearize the function and write the model as

$$\mathbf{y}^* \simeq \chi(x)\gamma(x) + u,$$

where $\gamma(x) = \begin{pmatrix}\gamma_1(x_1), & ..., & \gamma_M(x_M)\end{pmatrix}$, $\gamma_i(x_i) = \begin{pmatrix} m_i(x_i) & m_i^{(1)'}(x_i)\end{pmatrix}'$. Then the local linear least squares estimator of $\gamma(x)$ is

$$\hat{\gamma}(x) = (\chi'(x)K(x)\chi(x))^{-1}\chi'(x)K(x)\mathbf{y}^*. \tag{3.9}$$

Stack the model (3.8) into a matrix form as

$$\begin{aligned} \mathbf{y} &= \mathbf{m}(X) + Z\delta + \mathbf{u} \qquad (3.10) \\ &\simeq \chi(x)\gamma(x) + Z\delta + \mathbf{u}, \end{aligned}$$

where $Z = diag\left(Z_1, \quad \ldots \quad Z_M\right)$, $Z_i = \left(Z_{i1}, \quad \cdots, \quad Z_{iN}\right)'$,

$\delta = \left(\delta_1 \iota_N, \quad \ldots \quad \delta_M \iota_N\right)'$, and ι_N is a vector of ones. Substitute (3.9) into (3.10), then the estimator of δ can be obtained by

$$\begin{aligned}\mathbf{y} &= \chi(x)\hat{\gamma}(x) + Z\delta + u \\ &= \chi(x)(\chi'(x)K(x)\chi(x))^{-1}\chi'(x)K(x)(\mathbf{y} - Z\delta) + Z\delta + u \\ &= S(x)\mathbf{y} - S(x)Z\delta + Z\delta + u, \;\; S(x) \equiv \chi(x)(\chi'(x)K(x)\chi(x))^{-1}\chi'(x)K(x).\end{aligned}$$

Reorganizing the above, one can have

$$[I - S(x)]\,\mathbf{y} = [I - S(x)]\,Z\delta + u.$$

By the LS method, the estimator of δ is

$$\hat{\delta} = \left[Z'\left(I - S'(x)\right)\left(I - S(x)\right)Z\right]^{-1} Z'\left(I - S'(x)\right)\left(I - S(x)\right)\mathbf{y}. \tag{3.11}$$

Since all the information incorporated in the estimator of δ in (3.11) are known, we can estimate δ by (3.11) first, then substitute it into (3.9) to get

$$\hat{\gamma}(x) = (\chi'(x)K(x)\chi(x))^{-1}\chi'(x)K(x)(\mathbf{y} - Z\hat{\delta}).$$

Alternatively, we can estimate (3.8) using the idea of Robinson (1988) as the followings:

(i) Taking the conditional expectation of (3.8) leads to

$$E(y_{ij}|X_{ij}) = m_i(X_{ij}) + E(Z_{ij}|X_{ij})\delta_i. \tag{3.12}$$

(ii) Subtracting the above from (3.8) we have

$$y_{ij} - E(y_{ij}|X_{ij}) = (Z_{ij} - E(Z_{ij}|X_{ij}))\delta_i + u_{ij}, \;\; i = 1, ..., M, \;\; j = 1, \cdots, N. \tag{3.13}$$

Also, one can rewrite (3.12) as

$$m_i(X_{ij}) = E(y_{ij}|X_{ij}) - E(Z_{ij}|X_{ij})\delta_i. \tag{3.14}$$

(iii) The left hand side of (3.13) can be treated as the residuals ($\hat{u}_{ij,yx}$) obtained by regressing y_{ij} on X_{ij}, and similarly the right hand side of (3.13) $Z_{ij} - E(Z_{ij}|X_{ij})$ can be treated as the residuals ($\hat{u}_{ij,zx}$) obtained by regression Z_{ij} on X_{ij}. To estimate both regressions, we can just use local constant estimation. Hence, further we can estimate δ_i by OLS estimator of regressing $\hat{u}_{ij,yx}$ on $\hat{u}_{ij,zx}$, denoted as $\hat{\delta}_i = \left(\sum \hat{u}^2_{ij,zx}\right)^{-1} \sum \hat{u}_{ij,zx}\hat{u}_{ij,yx}$.

(iv) Further, one can regress $y^*_{ij} = y_{ij} - Z_{ij}\hat{\delta}_i$ on X_{ij} to obtain the LLLS estimator

$$\hat{\gamma}(x) = (\chi'(x)K(x)\chi(x))^{-1}\chi'(x)K(x)\mathbf{y}^*.$$

Notice that the above procedures do not incorporate the variance-covariance of the errors in the system. The following gives a more efficient estimation.

We can apply two-step estimator to the model (3.8). Combined with the profile least squares method mentioned earlier, our two-step estimator under this model can be derived as the following

$$\begin{aligned}
\mathbf{y}^* &= \mathbf{m}(X) + \mathbf{u}, \mathbf{y}^* = \mathbf{y} - Z\delta \\
\Omega^{-1/2}\mathbf{y}^* + (\mathbf{H}^{-1} - \Omega^{-1/2})\mathbf{m}(X) &= \mathbf{H}^{-1}\mathbf{m}(X) + \mathbf{v} \\
\mathbf{y}^{**} &= \chi^*(x)\gamma(x) + \mathbf{v}, \chi^*(x) = \mathbf{H}^{-1}\chi(x) \\
\hat{\gamma}(x) &= (\chi^{*\prime}(x)K(x)\chi^*(x))^{-1}\chi^{*\prime}(x)K(x)\mathbf{y}^{**}. \quad (3.15)
\end{aligned}$$

By substituting $\hat{\gamma}(x)$ into model (3.10), we have

$$\begin{aligned}
\mathbf{y} &= \chi(x)\hat{\gamma}(x) + Z\delta + \mathbf{u} \\
&= \chi(x)(\chi^{*\prime}(x)K(x)\chi^*(x))^{-1}\chi^{*\prime}(x)K(x)\Big[\Omega^{-1/2}(\mathbf{y} - Z\delta) \\
&\quad + (\mathbf{H}^{-1} - \Omega^{-1/2})\hat{\mathbf{m}}(x)\Big] + Z\delta + \mathbf{u} \\
&= S^*(x)[\Omega^{-1/2}(\mathbf{y} - Z\delta) + (\mathbf{H}^{-1} - \Omega^{-1/2})\hat{\mathbf{m}}(x)] + Z\delta + \mathbf{u},
\end{aligned}$$

where $S^*(x) \equiv \chi(x)(\chi^{*\prime}(x)K(x)\chi^*(x))^{-1}\chi^{*\prime}(x)K(x)$, the above can be rewritten as

$$[I - S^*(x)\Omega^{-1/2}]\mathbf{y} - S^*(x)(\mathbf{H}^{-1} - \Omega^{-1/2})\hat{\mathbf{m}}(x) = [I - S^*(x)\Omega^{-1/2}]Z\delta + \mathbf{u}.$$

Let $\tilde{\mathbf{y}} \equiv [I - S^*(x)\Omega^{-1/2}]\mathbf{y} - S^*(x)(\mathbf{H}^{-1} - \Omega^{-1/2})\hat{\mathbf{m}}(x)$, $\tilde{Z} \equiv [I - S^*(x)\Omega^{-1/2}]Z$, the GLS estimator of δ is

$$\hat{\delta}_{SP} = (\tilde{Z}'\Omega^{-1}\tilde{Z})^{-1}\tilde{Z}'\Omega^{-1}\tilde{\mathbf{y}}. \tag{3.16}$$

By substituting the above $\hat{\delta}_{SP}$ into (3.15), we can obtain a more efficient two-step estimator of $\gamma(x)$, $\hat{\gamma}_{SP}(x) = (\chi^{*\prime}(x)K(x)\chi^*(x))^{-1}\chi^{*\prime}(x)K(x)\mathbf{y}^{**}$. Ω can be estimated from the first step profile least squares. Then we will have an operational two-step estimator. Here, we introduce our two-step estimation by combining with the profile least squares. The alternative method combing with the idea of Robinson (1988) can be also proposed in a similar manner.

3.2.2.2 Additive NP models

As we know, the additive model is useful to conquer the notorious "curse of dimension" issue in nonparametric literature. In this section, we consider the following additive NP models within SUR system,

$$\begin{aligned} y_{ij} &= m_i(X_{ij,1}, ..., X_{ij,d}) + \varepsilon_{ij} \\ &= c_i + \sum_{\alpha=1}^{d} m_{i,\alpha}(X_{ij,\alpha}) + \varepsilon_{ij},\ i = 1, ..., M,\ j = 1, \cdots, N, \end{aligned}$$

where $X_{ij,\alpha}$ is the αth regressor.

To stack the regression models into one, we have

$$y = c + \sum_{\alpha=1}^{d} m_\alpha(X_\alpha) + \varepsilon, \tag{3.17}$$

where $y = \begin{pmatrix} y_{11}, & \ldots, & y_{MN} \end{pmatrix}$, $m_\alpha(X_\alpha) = \begin{pmatrix} m_{1,\alpha}(X_{1,\alpha}), & \ldots, & m_{M,\alpha}(X_{M,\alpha}) \end{pmatrix}'$, $\varepsilon = (\varepsilon_{11}, \ldots, \varepsilon_{MN})$. To estimate the above additive NP regression model, we use the marginal integration method. The similar idea can be found in Yang, Hardle, and Nielson (1999) for a single nonparametric autoregression model.

The estimation procedure of the marginal integration is as the following:

(1) Let $\overline{X}_{ij,\alpha}$ denote the vector that consists of all the remaining $X_{ij,\beta}, 1 \leq \beta \leq d$ and $\beta \neq \alpha$. We can estimate each component of mean function by $\hat{m}_{i,\alpha}(x_\alpha) = \frac{1}{N}\sum_{l=1}^{N} \hat{m}_i(x_{i,\alpha}, \overline{X}_{il,\alpha})$, which is based on the sample version of marginal integration $\int m_i(x_{i,\alpha}, \overline{X}_{il,\alpha}) dF(\overline{X}_{il,\alpha})$. And $\hat{m}_i(x_{i,\alpha}, \overline{X}_{il,\alpha})$ can be estimated by pth order local polynomial smoothing y_i on $(x_{i,\alpha}, \overline{X}_{il,\alpha})$ as $\hat{m}_i(x_{i,\alpha}, \overline{X}_{il,\alpha}) = e_0'(Z_i'W_{i,l}Z_i)^{-1}Z_i'W_{i,l}y_i$, where $Z_i \equiv \{(X_{ij,\alpha} - x_{i,\alpha})^\lambda\}_{N\times(p+1)}$, $\lambda = 0, \cdots, p$, $W_{i,l} = diag\{\frac{1}{N}K_{h_i}(X_{ij,\alpha} - x_{i,\alpha})L_{g_i}(\overline{X}_{ij,\alpha} - \overline{X}_{il,\alpha})\}_{j=1}^{N}$, L is a kernel function that has the same properties as K.

(2) Obtain $\hat{m}_i(x_i) = \hat{c}_i + \sum_{\alpha=1}^{d} \hat{m}_{i,\alpha}(x_{i,\alpha})$, where $\hat{c}_i = \frac{1}{N}\sum_{j=1}^{N} y_{ij}$.

(3) Further, we can obtain the estimated residuals $\hat{\varepsilon}_{ij} = y_{ij} - \hat{m}_i(x_i)$ to estimate Ω, H, and P.

By applying the transformation proposed in two-step estimation, we can transfer (3.17) into

$$\begin{aligned}
\Omega^{-1/2}\mathbf{y} + (\mathbf{H}^{-1} - \Omega^{-1/2})\left(c + \sum_{\alpha=1}^{d} m_\alpha(X_\alpha)\right) &= \mathbf{H}^{-1}\left(c + \sum_{\alpha=1}^{d} m_\alpha(X_\alpha)\right) + v \\
\vec{\mathbf{y}} &= \mathbf{H}^{-1}c + \mathbf{H}^{-1}\sum_{\alpha=1}^{d} m_\alpha(X_\alpha) + v \\
&= c^* + \sum_{\alpha=1}^{d} m_\alpha^*(X_\alpha) + v
\end{aligned}$$

Then employing the procedure proposed above, we can estimate the transformed model to obtain $\hat{m}_{\alpha,2-step}(X_\alpha)$. Specifically, the feasible transformed response variable can be obtained from the previous results as

$$\vec{\mathbf{y}} = \hat{\Omega}^{-1/2}\mathbf{y} + (\hat{\mathbf{H}}^{-1} - \hat{\Omega}^{-1/2})\left(\hat{c} + \sum_{\alpha=1}^{d} \hat{m}_\alpha(X_\alpha)\right).$$

The two-step estimator of $m_{i,\alpha}(x_{i,\alpha})$ is obtained as follows:

$$\hat{m}_{i,\alpha}(x_{i,\alpha}) = \frac{1}{N}\sum_{l=1}^{N}\hat{m}_{i,2-step}(x_{i,\alpha}, \overline{X}_{il,\alpha}),$$

$$\hat{m}_{i,2-step}(x_{i,\alpha}, \overline{X}_{il,\alpha}) = e_0'(Z_i^{*\prime}W_{i,l}Z_i^*)^{-1}Z_i^{*\prime}W_{i,l}\overrightarrow{\mathbf{y}}_i,$$

where $Z_i^* \equiv \{\hat{H}_i^{-1}(X_{ij,\alpha} - x_{i,\alpha})^{\lambda}\}_{N\times(p+1)}, \lambda = 0, \cdots, p$, $\hat{H}_i$ is the ith submatrix in $\hat{H}$.

3.2.2.3 Varying Coefficient NP models

Varying coefficient NP models are practically useful in applied works. The procedure for estimating this kind of model in a single equation has been extensively discussed in the literature. See Pagan and Ullah (1999) and Li and Racine (2007) for details. In this section, we consider the following varying coefficient NP model within SUR framework,

$$y_{ij} = \beta_i(Z_{ij})X_{ij} + \varepsilon_{ij}, i = 1, ..., M,\ j = 1, \cdots, N. \tag{3.18}$$

By local linearizing the coefficient, we have

$$\begin{aligned} y_{ij} &= \left[\beta_i(z_i) + (Z_{ij} - z_i)\beta_i^{(1)}(z_i)\right]X_{ij} + u_{ij} \\ &= \begin{pmatrix} 1 & (Z_{ij} - z_i)\end{pmatrix} X_{ij}\begin{pmatrix}\beta_i(z_i) \\ \beta_i^{(1)}(z_i)\end{pmatrix} + u_{ij} \\ &= \chi_{ij}(Z_{ij}, z_i, X_{ij})\delta_i(z_i) + u_{ij}, \end{aligned}$$

where $\beta_i^{(1)}(z_i) \equiv \partial\beta_i(z_i)/\partial z_i$, $\chi_{ij}(Z_{ij}, z_i, X_{ij}) \equiv \begin{pmatrix} 1 & (Z_{ij} - z_i)\end{pmatrix} X_{ij}$,

$$\chi_i(Z_i, z_i, X_i) = \begin{pmatrix}\chi_{i1}(Z_{i1}, z_i, X_{i1}), & \ldots, & \chi_{iN}(Z_{iN}, z_i, X_{iN})\end{pmatrix}',$$

which has dimension $N \times (q_i + 1)$. Stack the above models $j = 1, \cdots, M$, in a matrix form as

$$\begin{aligned} \mathbf{y} &= \beta(Z)X + \mathbf{u} \\ &= \chi(z)\delta(z) + u, \end{aligned}$$

where

$$\chi(z) = diag\begin{pmatrix}\chi_1(Z_1, z_1, X_1) & ,\ldots, & \chi_M(Z_M, z_M, X_M)\end{pmatrix},$$
$$\delta(z) = \begin{pmatrix}\delta_1(z_1) & ,\ldots, & \delta_M(z_M)\end{pmatrix}.$$

The local linear least squares estimator for the varying coefficient NP models in (3.18) is

$$\hat{\delta}(z) = (\chi'(z)\mathbf{K}(z)\chi(z))^{-1}\chi'(z)\mathbf{K}(z)\mathbf{y}.$$

Then we apply the two-step estimator as follows

$$\begin{aligned}\Omega^{-1/2}\mathbf{y} + (\mathbf{H}^{-1} - \Omega^{-1/2})\beta(Z)X &= \mathbf{H}^{-1}\beta(Z)X + v \\ \vec{\mathbf{y}}_{VF} &= \mathbf{H}^{-1}\beta(Z)X + v.\end{aligned}$$

The corresponding two-step estimator can be written as

$$\hat{\delta}_{2-step}(z) = (\chi^{*\prime}(z)\mathbf{K}(z)\chi^*(z))^{-1}\chi^{*\prime}(z)\mathbf{K}(z)\vec{\mathbf{y}}_{VF}, \tag{3.19}$$

where $\chi^*(\mathbf{z}) = \mathbf{H}^{-1}\chi(\mathbf{z})$. To obtain the operational estimator, in the first step, we can estimate each equation by local linear least squares to get residuals. Then use the residuals to get a consistent estimator of covariance, further, obtain the feasible $\vec{\mathbf{y}}_{VF} = \hat{\Omega}^{-1/2}\mathbf{y} + (\hat{\mathbf{H}}^{-1} - \hat{\Omega}^{-1/2})\chi(z)\hat{\delta}(z)$. In the second step, we regress the feasible $\vec{\mathbf{y}}_{VF}$ on $\mathbf{H}^{-1}\beta(Z)X$ to get the two-step estimator.

3.2.2.4 Varying Coefficent IV Models

In the previous section, we have discussed varying coefficient NP models with exogenous variables. In this section, we further consider the varying coefficient model with endogenous variables in SUR system. We extend the method proposed by Su, Murtazashvili, and Ullah (2011) for varying coefficient IV models within cross-sectional framework to semiparametric SUR system. Specifically, the model considered here is

$$\begin{aligned} y_{ij} &= \beta_i\left(U_{ij}\right)X_{ij} + \varepsilon_{ij}, i = 1,...,M,\ j = 1,\cdots,N \\ E(\varepsilon_{ij}|Z_{ij},U_{ij}) &= 0 \text{ almost surely (a.s.)} \end{aligned}$$

where X_{ij} is an endogenous regressor, U_{ij} denotes a $q_i \times 1$ vector of continuous exogenous regressors, and Z_{ij} is a $p_i \times 1$ vector of instrument variables.

The orthogonality condition $E(\varepsilon_{ij}|Z_{ij},U_{ij}) = 0$ a.s. provides the intuition that the unknown functional coefficients can be estimated by nonparametric generalized method of moments (NPGMM). Let $V_{ij} = (Z_{ij}', U_{ij})'$, we can write the orthogonality condition as

$$E[Q_{u_i}(V_{ij})\varepsilon_{ij}|V_{ij}] = E[Q_{u_i}(V_{ij})\{y_{ij} - \chi_{ij}(U_{ij},u_i,X_{ij})\delta_i(u_i)\}|U_{ij}] = 0,$$

where $\chi_{ij}(U_{ij},u_i,X_{ij}) = \begin{pmatrix} 1 & (U_{ij}-u_i)' \end{pmatrix} X_{ij}$, $\delta_i(u_i) = \begin{pmatrix} \beta_i\left(u_i\right) & \beta_i^{(1)'}\left(u_i\right) \end{pmatrix}'$. Following the idea of local linear GMM esitmaiton proposed by Su, Murtazashvili, and Ullah (2011), we can choose $Q_{u_i}(V_{ij}) = \begin{pmatrix} Z_{ij}^a \\ Z_{ij}^a \otimes \left(U_{ij}-u_i\right)/h_i \end{pmatrix}$, which is a $p_i(q_i+1)\times 1$ vector, where Z_{ij}^a is a $p_i \times 1$ vector of "global" instruments. The above conditional moment can be approximated by its sample analogue

$$\begin{aligned} g_i\left(u_i\right) &= \frac{1}{N}\sum_{j=1}^{N} Q_{u_i}(V_{ij})[y_{ij} - \chi_{ij}(U_{ij},u_i,X_{ij})\delta_i(u_i)]K_{h_i}(U_{ij}-u_i) \\ &= \frac{1}{N}Q_{u_i}(V_i)'K_{h_i}(u_i)\left[\mathbf{y}_i - \chi_i(u_i)\delta_i(u_i)\right], \end{aligned}$$

where $g_i\left(u_i\right)$ is a $k_i \times 1$ vector, $k_i = p_i(q_i+1)$,

$$Q_{u_i}(V_i)_{N\times p_i(q_i+1)} = \begin{pmatrix} Q_{u_i}(V_{i1}), & \ldots, & Q_{u_i}(V_{iN}) \end{pmatrix}',$$

ι_i is a $p_i(q_i+1)\times 1$ vector with unit elements, and

$$\begin{aligned} K_{h_i}(u_i) &= diag\begin{pmatrix} K_{h_i}(U_{i1}-u_i), & \ldots, & K_{h_i}(U_{iN}-u_i) \end{pmatrix} \\ &= diag\begin{pmatrix} \frac{1}{h_i}K(\frac{U_{i1}-u_i}{h_i}), & \ldots, & \frac{1}{h_i}K(\frac{U_{iN}-u_i}{h_i}) \end{pmatrix}. \end{aligned}$$

Define

$$\begin{aligned} g(u) &= \Big(g_1(u), \quad \ldots, \quad g_M(u)\Big)' \\ &= \frac{1}{N} Q(u)' K_h(u) \left[y - \chi(u)\delta(u)\right]. \end{aligned}$$

The dimension of $g(u)$ is $\sum_{i=1}^{M} k_i \times 1$, $Q(u) \equiv diag\left(Q_{u_1}(V_1), \quad \ldots, \quad Q_{u_M}(V_M)\right)$, which has dimension $MN \times \left(\sum_{i=1}^{M} k_i\right)$. To obtain $\delta(u)$, we can minimize the following local linear GMM criterian function

$$g(u)' \, \Psi(u)^{-1} g(u),$$

where

$$\begin{aligned} \Psi(u) &= E(g(u)\, g(u)') \\ &= \frac{1}{N^2} E\left(Q(u)' K_h(u)\, \varepsilon\varepsilon' K_h(u)\, Q(u)\right) \\ &= \frac{1}{N^2} Q(u)' K_h(u)\, \Omega K_h(u)\, Q(u), \end{aligned}$$

which is a symmetric $\sum_{i=1}^{M} k_i \times \sum_{i=1}^{M} k_i$ weight matrix that is positive definite. The above function can be written as

$$\left[Q(u)' K(u) \left(\mathbf{y} - \chi(u)\delta(u)\right)\right]' \Psi(u)^{-1} \left[Q(u)' K(u) \left(\mathbf{y} - \chi(u)\delta(u)\right)\right].$$

Then the local linear GMM estimator of $\delta(u)$ is given by $\hat{\delta}_{GMM}(u)$ as

$$\begin{aligned} \hat{\delta}_{GMM}(u) &= \left[\chi(u)' K(u) Q(u) \Psi(u)^{-1} Q(u)' K(u) \chi(u)\right]^{-1} \\ &\quad \chi(u)' K(u) Q(u) \Psi(u)^{-1} Q(u)' K(u) \mathbf{y} \\ &= \left\{\chi(u)' K(u) Q(u) \left[Q(u)' K_h(u)\, \Omega K_h(u)\, Q(u)\right]^{-1} Q(u)' K(u) \chi(u)\right\}^{-1} \\ &\quad \chi(u)' K(u) Q(u) \left[Q(u)' K_h(u)\, \Omega K_h(u)\, Q(u)\right]^{-1} Q(u)' K(u) \mathbf{y}. \end{aligned} \tag{3.20}$$

To obtain the optimal choice of weight matrix, we can first get the preliminary estimator $\tilde{\delta}_{GMM}(u)$ of $\delta_{GMM}(u)$ by setting $\Psi(u)$ as an identity matrix. Then we define the local residual $\tilde{\varepsilon}_{ij}(u_i) = y_{ij} - \chi_{ij}(U_{ij}, u_i, X_{ij})\tilde{\tilde{\delta}}_{GMM,i}(u_i)$. Using

$$\tilde{g}_i\left(u_i\right) = \frac{1}{N}\begin{pmatrix} \sum_{j=1}^{N} Z_{ij} K_{hi}(U_{ij} - u_i)\tilde{\varepsilon}_{ij}(u_i) \\ \sum_{j=1}^{N}\left(Z_{ij} \otimes \left(U_{ij} - u_i\right)/h_i\right) K_{h_t}(U_{ij} - u_t)\tilde{\varepsilon}_{ij}(u_i) \end{pmatrix}$$

to estimate $g_i\left(u_i\right)$, we can obtain the optimal choice of weight matrix

$$\tilde{\Psi}(u) = \begin{pmatrix} \tilde{g}_1\left(u_1\right)\tilde{g}_1\left(u_1\right)' & \tilde{g}_1\left(u_1\right)\tilde{g}_2\left(u_2\right)' & \cdots & \tilde{g}_1\left(u_1\right)\tilde{g}_M\left(u_M\right)' \\ \tilde{g}_2\left(u_2\right)\tilde{g}_1\left(u_1\right)' & \tilde{g}_2\left(u_2\right)\tilde{g}_2\left(u_2\right)' & \cdots & \tilde{g}_2\left(u_2\right)\tilde{g}_M\left(u_M\right)' \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{g}_M\left(u_M\right)\tilde{g}_1\left(u_1\right)' & \tilde{g}_M\left(u_M\right)\tilde{g}_1\left(u_1\right)' & \cdots & \tilde{g}_M\left(u_M\right)\tilde{g}_M\left(u_M\right)' \end{pmatrix}.$$

Alternatively, we can directly estimate the local variance-covariance matrix Ω by $\hat{\Omega}(u) = \hat{\Sigma}(u) \otimes I_N$. $\sigma_{ii'}$, the (i, i')th element of Σ, can be estimated by

$$\hat{\sigma}_{ii'} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\tilde{\varepsilon}_{ij}(u_i) - \overline{\tilde{\varepsilon}}_i(u_i)\right)\left(\tilde{\varepsilon}_{i'j}(u_{i'}) - \overline{\tilde{\varepsilon}}_{i'}(u_{i'})\right),$$

where $\overline{\tilde{\varepsilon}}_i(u_i) = \frac{1}{N}\sum_{j=1}^{N}\tilde{\varepsilon}_{ij}(u_i)$, $i, i' = 1, \ldots, M$. Then the feasible local linear GMM estimator is given by

$$\begin{aligned} \hat{\delta}_{GMM}(u) &= \left\{\chi(u)'K(u)Q(u)\left[Q(u)'K_h\left(u\right)\hat{\Omega}(u)K_h\left(u\right)Q(u)\right]^{-1}Q(u)'K(u)\chi(u)\right\}^{-1} \\ &\quad \chi(u)'K(u)Q(u)\left[Q(u)'K_h\left(u\right)\hat{\Omega}(u)K_h\left(u\right)Q(u)\right]^{-1}Q(u)'K(u)\mathbf{y}. \end{aligned} \quad (3.21)$$

3.2.2.5 NP SUR Models with Error Components

It has been widely recognized that the model combing cross-section with time-series data has many advantages and applications. Hence, it is also interesting and meaningful to introduce the nonparametric panel models into SUR system. This section considers nonparametric seemingly unrelated regressions with two or three error components structure. The NP SUR models with three error components is given as

$$y_{it,j} = m_j(X_{it,j}) + u_{it,j},\ j = 1, ..., M,\ i = 1, \cdots, N, t = 1, ..., T, \quad (3.22)$$

$$u_{it,j} = \mu_{i,j} + v_{t,j} + \varepsilon_{it,j}, \quad (3.23)$$

where j is the equation index, i indexes the individuals, and t indexes the time periods. It is the nonparametric analogue to the parametric SUR models considered by Avery (1977).

To derive the covariance of the errors for two observations we have

$$\begin{aligned} E\left(u_{it,j}u_{i't',j'}\right) &= E\left(\mu_{i,j}+v_{t,j}+\varepsilon_{it,j}\right)\left(\mu_{i',j'}+v_{t',j'}+\varepsilon_{i't',j'}\right) \\ &= E\left(\mu_{i,j}\mu_{i',j'}\right)+E\left(v_{t,j}v_{t',j'}\right)+E\left(\varepsilon_{it,j}\varepsilon_{i't',j'}\right). \end{aligned} \tag{3.24}$$

The second equality is based on the assumption of independent non-corresponding components. Following the notation in Avery (1977) and the standard assumptions on errors for parametric SUR, we have

$$\begin{aligned} E\left(\mu_{i,j}\mu_{i',j'}\right) &= \sigma^2_{\mu_{jj'}}, i=i', \\ &= 0, i\neq i', \\ E\left(v_{t,j}v_{t',j'}\right) &= \sigma^2_{v_{jj'}}, t=t', \\ &= 0, t\neq t', \\ E\left(\varepsilon_{it,j}\varepsilon_{i't',j'}\right) &= \sigma^2_{\varepsilon_{jj'}}, i=i' \text{ and } t=t', \\ &= 0, i\neq i' \text{ or } t\neq t', \end{aligned}$$

and the covariance of the errors for two observations is given as

$$\begin{aligned} E\left(u_{it,j}u_{i't',j'}\right) &= \sigma^2_{\mu_{jj'}}+\sigma^2_{v_{jj'}}+\sigma^2_{\varepsilon_{jj'}} \\ &= \sigma^2_{jj'}, \text{ if } i=i' \text{ and } t=t'. \end{aligned}$$

We stack the models (3.22) into a matrix form

$$\mathbf{y}=\mathbf{m}(X)+U,$$

where $U=\left(U_1',...,U_M'\right)'$, $\mathbf{y}\equiv\begin{pmatrix} y_{11,1}, & \ldots, & y_{NT,M}\end{pmatrix}'$,
$\mathbf{m}(X)=(\mathbf{m}_1(X_1),...,\mathbf{m}_M(X_M))'$, and $\mathbf{m}_j(X_j)=(m_j(X_{11,j}),...,m_j(X_{NT,j}))'$ which is

a $NT \times 1$ vector. Following Avery (1977), we express the variance-covariance matrix of the residuals for the entire system in (3.22) as

$$E\left(UU'\right) = \Omega = \begin{pmatrix} \sigma_{11}^2 \Sigma_{11} & \cdots & \sigma_{1J}^2 \Sigma_{1J} \\ \vdots & \ddots & \vdots \\ \sigma_{J1}^2 \Sigma_{J1} & \cdots & \sigma_{JJ}^2 \Sigma_{JJ} \end{pmatrix}, \tag{3.25}$$

where $\Sigma_{jj'} = \left(1 - \rho_{jj'} - \varpi_{jj'}\right) I_{NT} + \rho_{jj'} \left(I_N \otimes \iota_T \iota_T'\right) + \varpi_{jj'} \left(\iota_N \iota_N' \otimes I_T\right)$, $\rho_{jj'} \equiv \frac{\sigma_{\mu_{jj'}}^2}{\sigma_{jj'}^2}$, $\varpi_{jj'} \equiv \frac{\sigma_{v_{jj'}}^2}{\sigma_{jj'}^2}$.

Now it is ready to apply our two-step estimator into the model (3.22). The procedure is described as the followings:

(i) We can estimate each equation by LLLS to obtain the pooled LLLS estimator $\hat{m}_{j,LLLS}(x_j)$. Define $\hat{U}_j \equiv y_j - \hat{m}_{j,LLLS}(x_j)$. Then we can estimate the unconditional covariance component of errors by the two-way analysis of variance method described in Avery (1977) within the parametric framework. Hence, the estimated covariance $\hat{\Omega}$ can be obtained.

(ii) By using $\hat{\Omega}$, we can obtain the estimated $\hat{\mathbf{H}}$ and the feasible $\overrightarrow{\mathbf{y}}$ defined in our two-step estimator (3.3), further, the two-step estimator in the form of (3.4) for the NP SUR with error component models.

The above provides a general nonparametric framework for SUR with three component error structure. If one is interested in random effects SUR model with two error components structure, then the model (3.22) can be simplified to

$$y_{it,j} = m_j(X_{it,j}) + \varepsilon_{it,j} + u_{i,j},\ j = 1, ..., M,\ i = 1, \cdots, N, t = 1, ..., T. \tag{3.26}$$

Each equation j is a one-way random effect model, and $u_{it,j} = \varepsilon_{it,j} + \mu_{i,j}$. In model (3.26), the (j, j')th element in the variance-covariance matrix would be $\sigma_{jj'}^2 \Sigma_{jj'} = \sigma_{jj'}^2 \left(1 - \rho_{jj'}\right) I_{NT} + \sigma_{jj'}^2 \rho_{jj'} \left(I_N \otimes \iota_T \iota_T'\right)$.

The fixed effects SUR model can be written as

$$y_{it,j} = c_{i,j} + m_j(X_{it,j}) + \varepsilon_{it,j},\ j = 1, ..., M,\ i = 1, \cdots, N, t = 1, ..., T, \tag{3.27}$$

where $c_{i,j}$ is correlated with the included variables, i.e., $E(c_{i,j}|X_{i,j}) = h(X_{i,j})$, and the assumptions on $\varepsilon_{it,j}$ remain the same as random effects SUR models. The estimation of the above fixed effects SUR model is a straightforward extension of the method proposed in the section (3.2.2.1) for partially linear semiparametric SUR models.

3.2.3 Estimation with Conditional Error Variance-Covariance $\Omega(x)$

All the aforementioned estimations are based on the parametric variance covariance. This section provides the asymptotic theorems for local linear least squares estimator and our two-step estimator for the NP SUR regressions with conditional error variance-covariance. We consider the SUR model in (3.6) but with the conditional variance-covariance of errors. Now we assume that $E(u_{ij}|X_{ij}) = 0$, and homoscedasticity $Var(\varepsilon_{ij}|X_{ij}) = \sigma_{ii}^2(X_{ij})$ for each equation. Also, we assume that the disturbances are uncorrelated across observations but correlated across equations, i.e, $E(\varepsilon_{ij}\varepsilon_{i'j}|X_{ij}, X_{i'j}) = \sigma_{ii'}(X_{ij}, X_{i'j})$ for $i, i' = 1, \cdots, M$ and $i \neq i'$, and $j = 1, \cdots, N$. In a matrix form, the conditional variance-covariance is $\Omega(x) \equiv \Sigma(x) \otimes I$ for a given evaluated point x.

Assumption A7. $\Sigma(x)$ is a nonsingular matrix, and the partial derivatives of the components of $\Sigma(x)$ exist in a neighbourhood of x, i.e., $\partial_s\sigma_{ii}^2(x)$, $\partial_s\sigma_{ii'}(x)$ exist, $s = 1, \ldots M$.

Assumption A8. $\Omega(x) = P(x)P(x)'$ for some $MN \times MN$ matrix $P(x)$, $p_{ij}(x)$ and $v_{ij}(x)$ denote the (i, j)th element of $P(x)$ and $P^{-1}(x)$, respectively, and $\mathbf{H}(x) \equiv$diag

$(v_{1,1}^{-1}(x), ..., v_{MN,MN}^{-1}(x))$. And the partial derivatives of $v_{ij}(x)$ exist in a neighbourhood of x.

Assumption A9. $\bar{f}_i(x_i) = \lim_{N\to\infty} N^{-1} f_{ij}(x_i)$, and $0 < \bar{f}_i(x_i) < \infty$; $\omega_{f,i}^*(x_i) = \lim_{N\to\infty} N^{-1}\sum_{j=1}^{N} \upsilon_{(i-1)N+j}^2(x_i) f_{ij}(x_i)$, and $0 < \omega_{f,i}^*(x_i) < \infty$.

Theorem 7 *Under the assumptions A1-A9 and the assumptions on the error terms of basic SUR models, we have*

$$D(\hat{\delta}(x) - \delta(x) - B_{LLLS}) \xrightarrow{d} N(0, \Omega_{LLLS}),$$

where $D \equiv diag\begin{pmatrix} D_1, \cdots, & D_M \end{pmatrix}$, $D_i = \sqrt{Nh_i^{q_i}} D_{h_i}$, $D_{h_i} \equiv diag(1, h_i, ..., h_i)$ *is a* $(1+q_i)\times(1+q_i)$ *diagonal matrix,* $B_{LLLS} = \begin{pmatrix} B_{1,LLLS}, & \ldots, & B_{M,LLLS} \end{pmatrix}'$,
$\Omega_{LLLS} = diag\begin{pmatrix} \Omega_{1,LLLS}, & \ldots, & \Omega_{M,LLLS} \end{pmatrix}$,

$$B_{i,LLLS} = \begin{pmatrix} \frac{\kappa_{21} h_i^2}{2} \sum_{s=1}^{q_i} \frac{\partial^2 m_i(x_i)}{\partial x_{i,s}^2} \\ \mathbf{0}_{q_i \times 1} \end{pmatrix},$$

$$\Omega_{i,LLLS} = \begin{pmatrix} \frac{\sigma_{ii}^2(x_i)(\kappa_{02})^{q_i}}{f_i(x_i)} & \mathbf{0}'_{1\times q_i} \\ \mathbf{0}_{q_i\times 1} & \frac{\sigma_{ii}^2(x_i)\kappa_{22}(\kappa_{02})^{q_i-1}}{f_i(x_i)\kappa_{21}^2} I_{q_i} \end{pmatrix},$$

and x_{is} *is the sth element of* x_i *for* $i = 1, \ldots, M$.

Theorem 8 *Under the assumptions A1-A9 and the assumptions on the error terms of basic SUR models, we have*

$$D\left(\hat{\delta}_{2-step}(x) - \delta(x) - B_{2-step}\right) \xrightarrow{d} N(0, \Omega_{2-step}(x)),$$

where $D \equiv diag\begin{pmatrix} D_1, \cdots, & D_M \end{pmatrix}$, $D_i = \sqrt{Nh_i^{q_i}} D_{h_i}$, $D_{h_i} \equiv diag(1, h_i, ..., h_i)$ *is a* $(1+q_i)\times$

$(1+q_i)$ *diagonal matrix,* $B_{2-step} = \left(B_{1,2-step}, \quad \ldots, \quad B_{M,2-step} \right)'$,

$$\Omega_{2-step}(x) = diag\left(\Omega_{1,2-step}(x), \quad \ldots, \quad \Omega_{M,2-step}(x) \right),$$

$$B_{i,2-step} = \begin{pmatrix} \frac{k_{21}h_i^2}{2} \sum_{s=1}^{q_i} \frac{\partial^2 m_i(x_i)}{\partial x_{is}^2} \\ \mathbf{0}_{q_i \times 1} \end{pmatrix},$$

$$\Omega_{i,2-step}(x_i) = \begin{pmatrix} \frac{(\kappa_{02})^{q_i}}{\omega_{f,i}^*(x_i)} & \mathbf{0}_{1\times q_i} \\ \mathbf{0}_{q_i \times 1} & \frac{\kappa_{22}(\kappa_{02})^{q_i-1}}{\omega_{f,i}^*(x_i)\kappa_{21}^2} I_{q_i} \end{pmatrix},$$

and x_{is} *is the sth element of* x_i *for* $i = 1, \ldots, M$.

Remark In both Theorem 7 and 8, the off-diagonals of asymptotic variance-covariance matrix are zero because the off-diagonals are smaller order than the diagonals. Also, the conditional correlation doesn't enter the asymptotic variance-covariance since the terms incoporating the correlations are all smaller order than those terms with conditional variances.

To obtain a feasible two-step estimation in this scenario, the estimated conditional variance-covariance is required. We can estimate the conditional covariance as the following

$$\hat{\sigma}_{ii}^2(x) = \frac{\frac{1}{N}\sum_{j=1}^{N} K_{\mathbf{h}}(x_i - X_{ij})\varepsilon_{ij}^2}{\frac{1}{N}\sum_{j=1}^{N} K_{\mathbf{h}}(x_i - X_{ij})}, \text{ for } i = 1, \ldots, M$$

$$\hat{\sigma}_{ii'}(x) = \widehat{Cov}(\varepsilon_{ij}, \varepsilon_{i'j}) = \frac{\frac{1}{N}\sum_{j=1}^{N} K_{\mathbf{h}}(x - X_j)\varepsilon_{ij}\varepsilon_{i'j}}{\frac{1}{N}\sum_{j=1}^{N} K_{\mathbf{h}}(x - X_j)}, \text{ for } i, i' = 1, \ldots, M \text{ and } i \neq i',$$

where $X_j \in \Re^d$ is a disjoint union of $\{X_{ij}\}$, $\mathbf{h} = diag(h_1, \ldots, h_q)$, $K_{\mathbf{h}}(x - X_j) = |\mathbf{h}|^{-1} K\left(\mathbf{h}^{-1}(x - X_j)\right)$, and $K_{\mathbf{h}}(x_i - X_{ij}) = |\mathbf{h}|^{-1} K\left(\mathbf{h}^{-1}(x_i - X_{ij})\right)$.

3.3 Goodness-of-fit

Another important issue about nonparametric regression analysis is goodness-of-fit. This section proposes two alternative ways to measure the fit of NP SUR models. Consider the stacked SUR model as the following

$$\begin{aligned}\mathbf{y} &= \mathbf{m}(X) + \varepsilon \\ &= Z(x)\delta(x) + u.\end{aligned}$$

For each equation, we can construct variance decomposition within nonparametric framework as follows

$$\begin{aligned}y_{ij} &= \hat{m}_i(x_i) + (X_{ij} - x_i)\hat{m}_i^{(1)}(x_i) + \hat{u}_{ij} \\ y_{ij} - \bar{y} &= \hat{m}_i(x_i) + (X_{ij} - x_i)\hat{m}_i^{(1)}(x_i) - \bar{y} + \hat{u}_{ij} \\ (y_{ij} - \bar{y})\sqrt{K_{h_i}(X_{ij} - x_i)} &= \left[\hat{m}_i(x_i) + (X_{ij} - x_i)\hat{m}_i^{(1)}(x_i) - \bar{y}\right]\sqrt{K_{h_i}(X_{ij} - x_i)} \\ &\quad + \hat{u}_{ij}\sqrt{K_{h_i}(X_{ij} - x_i)}.\end{aligned}$$

If we can show that the above cross-product term equals zero, then the ANOVA decomposition can be applied. Notice that the cross-product term

$$\begin{aligned}&\sum_{i=1}^{M}\sum_{j=1}^{N}\left[\hat{m}_i(x_i) + (X_{ij} - x_i)\hat{m}_i^{(1)}(x_i) - \bar{y}\right]K_{h_i}(X_{ij} - x_i)\,\hat{u}_{ij} \\ =\; &\sum_{i=1}^{M}\sum_{j=1}^{N}\hat{m}_i(x_i)K_{h_i}(X_{ij} - x_i)\,\hat{u}_{ij} + \sum_{i=1}^{M}\sum_{j=1}^{N}(X_{ij} - x_i)\hat{m}_i^{(1)}(x_i)K_{h_i}(X_{ij} - x_i)\,\hat{u}_{ij} \\ =\; &0.\end{aligned}$$

since $\sum_{j=1}^{N}\hat{u}_{ij}K_{h_i}(X_{ij} - x_i) = 0$ and $\sum_{j=1}^{N}\hat{u}_{ij}(X_{ij} - x_i)K_{h_i}(X_{ij} - x_i) = 0$ by the local linear least squares method. Then the local ANOVA decomposition for nonparametric regression can be written as

$$TSS(x) = ESS(x) + RSS(x),$$

specifically,

$$\sum_{i=1}^{M}\sum_{j=1}^{N}(y_{ij}-\bar{y})^2 K_{h_i}(X_{ij}-x_i) = \sum_{i=1}^{M}\sum_{j=1}^{N}\left[\hat{m}_i(x_i)+(X_{ij}-x_i)\hat{m}_i^{(1)}(x_i)-\bar{y}\right]^2 \cdot K_{h_i}(X_{ij}-x_i)+\sum_{i=1}^{M}\sum_{j=1}^{N}\hat{u}_{ij}^2 K_{h_i}(X_{ij}-x_i).$$

Therefore, a local nonparametric goodness-of-fit for SUR system by LLLS can be defined as

$$R^2(x) = 1-\frac{RSS(x)}{TSS(x)}=\frac{ESS(x)}{TSS(x)} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[\hat{m}_i(x_i)+(X_{ij}-x_i)\hat{m}_i^{(1)}(x_i)-\bar{y}\right]^2 K_{h_i}(X_{ij}-x_i)}{\sum_{i=1}^{M}\sum_{j=1}^{N}(y_{ij}-\bar{y})^2 K_{h_i}(X_{ij}-x_i)}.$$

For each i regression, we have

$$R_i^2(x_i)=\frac{\sum_{j=1}^{N}\left[\hat{m}_i(x_i)+(X_{ij}-x_i)\hat{m}_i^{(1)}(x_i)-\bar{y}\right]^2 K_{h_i}(X_{ij}-x_i)}{\sum_{j=1}^{N}(y_{ij}-\bar{y})^2 K_{h_i}(X_{ij}-x_i)}.$$

Similarly, we can define the local nonparametric goodness-of-fit for our two-step estimator

$$\hat{\delta}_{2-step}(x)=(R^{*\prime}(x)K(x)R^*(x))^{-1}R^{*\prime}(x)K(x)\overrightarrow{\mathbf{y}}$$

as

$$R_{2-step}^2(x)=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[\upsilon_{(i-1)N+j}^2\left(\hat{m}_i(x_i)+(X_{ij}-x_i)\hat{m}_i^{(1)}(x_i)\right)-\overline{\overrightarrow{y}}\right]^2 K_{h_i}(X_{ij}-x_i)}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\overrightarrow{y}_{ij}-\overline{\overrightarrow{y}}\right)^2 K_{h_i}(X_{ij}-x_i)},$$

where $\upsilon_{(i-1)N+j}$ is the diagonal element of P^{-1}, and $\hat{m}_i(x_i)$ and $\hat{m}_i^{(1)}(x_i)$ are obtained by our two-step estimator.

Based on local nonparametric goodness-of-fit, the global nonparametric goodness-of-fit can be written as

$$R^2=1-\frac{RSS}{TSS}=\frac{ESS}{TSS}.$$

For example, for LLLS estimator, the global ANOVA decompostition incorporates

$$
\begin{aligned}
TSS &\equiv \int_{\chi} TSS(x)dx = \sum_{i=1}^{M}\sum_{j=1}^{N}(y_{ij}-\bar{y})^2 \int K_{h_i}(X_{ij}-x_i)dx_i, \\
ESS &\equiv \int_{\chi} ESS(x)dx \\
&= \sum_{i=1}^{M}\sum_{j=1}^{N}\int \left[\hat{m}_i(x_i)+(X_{ij}-x_i)\hat{m}_i^{(1)}(x_i)-\bar{y}\right]^2 K_{h_i}\left(X_{ij}-x_i\right)dx_i, \\
RSS &\equiv \int_{\chi} RSS(x)dx = \sum_{i=1}^{M}\sum_{j=1}^{N}\int \hat{u}_{ij}^2 K_{h_i}\left(X_{ij}-x_i\right)dx_i.
\end{aligned}
$$

The corresponding global R^2 for the two-step estimator can be obtained in a similar manner.

Notice that both local and global R^2s defined above for the two-step estimator are acturally goodness-of-fit measures for the corresponding transformed models. In other words, R^2 defined for the above two-step estimator measures the fraction of variation in transformed dependent variables that can be explained by the transformed model. Also, notice that ANOVA decompositions do not hold for $\mathbf{y} = Z(x)\hat{\delta}_{2-step}(x)+\hat{u}$. Therefore, R^2 based on ANOVA decomposition is not available for the fitted model, $\hat{\mathbf{y}} = Z(x)\hat{\delta}_{2-step}(x)$.

To propose a goodness-of-fit measure that can explain the fit of original model estimated by our two-step estimation, we define the following nonparametric R^2 estimator based on an indicator function as

$$
\begin{aligned}
\hat{R}^2_{I,2-step}(x) &= \left[1-\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(y_{ij}-Z_{ij}(x_i)\hat{\delta}_{2-step}(x_i)\right)^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(y_{ij}-\bar{y}\right)^2}\right]\cdot \\
&\quad I\left(\sum_{i=1}^{M}\sum_{j=1}^{N}(y_{ij}-\bar{y})^2 \geq \sum_{i=1}^{M}\sum_{j=1}^{N}\left(y_{ij}-Z_{ij}(x_i)\hat{\delta}_{2-step}(x_i)\right)^2\right),
\end{aligned}
$$

where $I(\cdot)$ is the indicator function, which makes sure that $\hat{R}^2_{I,2-step}(x)$ takes value in $[0,1]$. Here, the idea of using an indicator function follows Yao and Ullah (2011).

This $\hat{R}^2_{I,2-step}(x)$ gives the measure for how well the fitted model $Z(x)\hat{\delta}_{2-step}(x)$ can explain the variation in y. The above two types of goodness-of-fit measure provides a fundamental knowledge that can be used to develop various test based on R^2 for NP/SP SUR system.

3.4 Simulation

In this section, we conduct a small set of Monte Carlo simulations to study the finite sample properties of LLLS, LLWLSs, $\tau-$type two-step estimator, and our two-step estimator. For LLWLS, we examine two types of weights, $W_1(x) = K^{1/2}(x)\Omega^{-1}K^{1/2}(x)$ and $W_4(x) = \Omega^{-1/2}K(x)\Omega^{-1/2}$, which are commonly used in literature.

We first generate data from the following data generating processes (DGPs):

$$DGP1 : \begin{cases} Y_{1i} = \phi\left(X_{1i}, 0.5, 0.15\right) + U_{1i} \\ Y_{2i} = \phi\left(X_{2i}, 0.5, 0.15\right) + U_{2i} \end{cases}$$

where $\phi\left(x, a, b\right)$ is the normal density function with mean 0.5 and standard deviation 0.15. We set $\{X_{1i}\}$ and $\{X_{2i}\}$ are mutually independent iid $U(0,1)$, and

$$\begin{pmatrix} U_{1i} \\ U_{2i} \end{pmatrix} \sim iid\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 2\rho \\ 2\rho & 4 \end{pmatrix}\right).$$

The process $\{X_i = (X_{1i}, X_{2i})\}$ is independent of the process $\{U_{1i}, U_{2i}\}$. We consider different choices of $\rho : 0, \pm 0.5, \pm 0.9$. DGP 1 is designed to compare the finite sample performance of LLS, LLWLSs, and two-step estimators under the specification of $y = m(X) + u$ with unconditional error variance-covariance.

$$DGP2 : \begin{cases} Y_{1i} = X_{1i}^2 + U_{1i} \\ Y_{2i} = \exp(X_{2i})/\left(1 + \exp(X_{2i})\right) + U_{2i} \end{cases}$$

where $\{X_i = X_{1i} = X_{2i}\}$ are standard normal random variables $N(0,1)$ but with compact support $[-12, 12]$. The disturbances are generated as

$$\begin{pmatrix} U_{1i} \\ U_{2i} \end{pmatrix} |X_i \sim iid\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2(X_i)\begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}\right),$$

where $\sigma^2(X_i) = 0.25 + X_i^2$, and $\rho_i = \rho \exp(X_i)/(1 + \exp(X_i))$. We consider different choices of $\rho : 0, \pm 0.5, \pm 0.9$. DGP 2 is designed to compare the finite sample performance of LLS, two-step estimators, and LLWLS estimators under the specification of $y = m(X) + u$ with conditional error variance-covariance.

For DGP 1, we use cross-validation least squares bandwidth calculated by LLLS for the estimations of LLLS, LLWLS estimators, and the second step of two-step estimators. We choose the first step bandwidth for both two-step estimators as one third of the second step bandwidth. In the first step, a smaller bandwidth than the one in the second step should be used to eliminate the bias raised from the first step estimation.

The bandwidths used for estimations in DGP 2 are the same as in DGP 1. The difference from DGP 1 is that in DGP 2 we need choose a bandwidth to estimate the conditional covariance. Here we use the rule of thumb, i.e., $c = 2.12$, as the bandwidth to estimate the conditional covariance. We use 1,000 replications for $n = 50$, and 500 replications for $n = 100$ for each DGP.

Table 3.1 on page 66 and Table 3.2 on page 67 list the results for DGP 1 for $n = 50, 100$, separately. Both tables report the estimation with true parameters and with estimated parameters in variance-covariance matrix. The finite sample performance is evaluated in terms of absolute bias and mean squared errors. In DGP 1 we assume homoskedasticity within each equation, and a correlation across two equations. The findings in Table 1 are summarized as the followings. First, we can obviously see that our two-step estimator generally outperforms other types of estimators in the sense of having smallest absolute bias and mean squared errors, only except for the cases with $\rho = \pm 0.9$, and the case when $\rho = 0.5$ and using true parameters in estimation, in which τ two-step estimator performs the best. Second, the two types of two-step estimator perform better than LLLS and two different weighted LLWLS. Third, comparing LLWLS with LLLS, we find that these two commonly used LLWLS estimators do not have consistent efficiency gain over LLLS in these settings. It is worth mention that when there is no correlation across equations, local linear weighted estimators have the same performance as LLLS when using the true variance-covariance in estimation and have lager absolute bias and MSE when using estimated variance-covariance. This simulation result is consistent with the theoretical findings in Welsh and Yee (2006), and also it is consistent with the situation within the parametric SUR framework. As we know, under the parametric SUR system, if there is no correlation across equations, the GLS estimator doesn't gain any efficiency over the least squares. That is, without correlation, estimating SUR jointly is equal to estimating each equation separately in parametric framework.

Table 3.2 on page 67 also gives the simulation results for DGP 1 with sample size $n = 100$. First, we can observe that larger sample size better performances for all estimators, by comparing Table 3.2 with Table 3.1. Second, we find that LLWLS with two different weights tend to outperform over LLLS in the sense of having lower MSE for most of cases. For the case with no correlation and using true variance-covariance, we can still see the consistent result as Table 3.1, i.e., LLLS and two types of weighted local linear estimators have the exactly same performance. Also, LLWLS with $W_4(x) = \Omega^{-1/2}K(x)\Omega^{-1/2}$ has better performance than LLWLS with $W_1(x) = K^{1/2}(x)\Omega^{-1}K^{1/2}(x)$. Third, when using estimated variance-covariance in estimation, our two-step estimator always has the best performance among these five estimators in terms of both absolute bias and MSE. When using the true variance-covariance in estimation, $\tau-$type two-step estimator works best for the case with $\rho = \pm 0.9$. Under other cases, our two-step estimator still beats other estimators. Fourth, similarly as Table 3.1 shows, these two types of two-step estimators have better performance than LLLS and two different weighted LLWLS.

In order to compare the performance of all these five estimators for the case with conditional covariance, we propose DGP 2. The results for $n = 50, 100$, and $\rho = 0, \pm 0.5, \pm 0.9$ are reported in Table 3.3 on page 68. We find that our two-step estimator has the smallest absolute biases and lowest MSEs for all cases, and $\tau-$two-step performs better than LLLS, and LLWLSs. An interesting finding here is that LLLS, and LLWLSs with different weights have the same performance. The reason behind is that all nonparametric estimators give local estimates for given evaluated point x. By the design of DGP 2, the conditional variance-covariance will become homoskedastic given particular x. As Welsh and Yee (2006) indicated, if $\rho(x) = \rho$, estimating the system jointly doesn't have any efficiency gain over estimating the system marginally. Also, in section (3.2.1.3), we have shown that if the system has identical explanatory variables, LLWLS is equivalent to LLLS. Hence, it is not surprised to see that LLLS and LLWLS perform the exact same in DGP 2. According to the simulation results in Table 3.1, Table 3.2, and Table 3.3, two-step estimators outperform LLLS, and LLWLSs. And our two-step estimator generally has the best performance.

3.5 Concluding Remarks

The aim of this chapter is to contribute the theoretical advances for nonparametric and semiparametric SUR system. The main contributions include the followings. First, we study the asymptotic properties of LLLS and our two-step estimator for both unconditional error variance-covariance and conditional error variance-covariance cases, and also further discuss the properties of different types of LLWLS proposed in the literature. Second, we introduce various popular nonparametric or semiparametric models in cross-sectional or panel data framework into SUR system, and provide efficient estimation procedures for these various specifications. Third, the nonparametric goodness-of-fit measures are defined for the nonparametric SUR models, which can be used as a fundamental knowledge to develop a series of hypothesis testing based on R^2. The current chapter doesn't give the asymptotic properties for the proposed estimation of various popular specifications. The related works are worth being developed in the future.

Table 3.1: DGP 1 with $n = 50$

sample size $n = 50$		No Estimation		With Esitmation	
Correlation size (ρ)	Estimators	Abs. Bias	MSE	Abs. Bias	MSE
0	LLLS	4.2098	1.8972	4.1602	1.8370
	LLWLS($W_1(x)$)	4.2098	1.8972	4.1827	1.8605
	LLWLS($W_4(x)$)	4.2098	1.8972	4.1824	1.8593
	$\tau-$two-step	4.1822	1.8600	4.1264	1.7975
	two-step	**4.1762**	**1.8504**	**4.1168**	**1.7867**
0.5	LLLS	4.1831	1.8649	4.2144	1.8972
	LLWLS($W_1(x)$)	4.2010	1.8930	4.2049	1.8881
	LLWLS($W_4(x)$)	4.1993	1.8883	4.2017	1.8814
	$\tau-$two-step	**4.1408**	1.8213	4.1632	1.8396
	two-step	4.1417	**1.8181**	**4.1587**	**1.8320**
0.9	LLLS	4.1932	1.8950	4.1327	1.8399
	LLWLS($W_1(x)$)	4.2013	1.8911	4.1142	1.8034
	LLWLS($W_4(x)$)	4.1909	1.8702	4.1058	1.7872
	$\tau-$two-step	**4.1197**	**1.8095**	**4.0616**	**1.7618**
	two-step	4.1393	1.8241	4.0736	1.7726
-0.5	LLLS	4.1841	1.8622	4.1546	1.8304
	LLWLS_W1	4.3190	2.0149	4.1482	1.8186
	LLWLS_W2	4.3194	2.0135	4.1461	1.8141
	$\tau-$two-step	4.1954	1.8612	4.1102	1.7740
	two-step	**4.1905**	**1.8534**	**4.1054**	**1.7699**
-0.9	LLLS	4.1561	1.8494	4.2411	1.9442
	LLWLS_W1	4.1356	1.8230	4.1812	1.8841
	LLWLS_W2	4.1282	1.8079	4.1744	1.8711
	$\tau-$two-step	**4.0887**	**1.7573**	**4.1694**	**1.8513**
	two-step	4.1093	1.7753	4.1812	1.8629

Table 3.2: DGP 1 with $n = 100$

sample size $n = 100$		No Estimation		With Esitmation	
Correlation size (ρ)	Estimators	Abs. Bias	MSE	Abs. Bias	MSE
0	LLLS	3.9918	1.6571	4.0114	1.6806
	LLWLS_W1	3.9918	1.6571	4.0170	1.6885
	LLWLS_W2	3.9918	1.6571	4.0149	1.6856
	$\tau-$two-step	3.9378	1.5975	3.9478	1.6131
	two-step	**3.9189**	**1.5772**	**3.9217**	**1.5868**
0.5	LLLS	3.9294	1.5943	3.9444	1.5981
	LLWLS_W1	3.9522	1.6149	3.9609	1.6159
	LLWLS_W2	3.9321	1.5891	3.9420	1.5917
	$\tau-$two-step	3.8696	1.5384	3.8760	1.5327
	two-step	**3.8594**	**1.5269**	**3.8584**	**1.5150**
0.9	LLLS	3.9841	1.6490	3.9923	1.6593
	LLWLS_W1	3.9918	1.6605	3.9631	1.6243
	LLWLS_W2	3.9318	1.5818	3.9070	1.5532
	$\tau-$two-step	**3.9004**	**1.5659**	3.8874	1.5550
	two-step	3.9098	1.5700	**3.8816**	**1.5469**
-0.5	LLLS	4.0090	1.6915	3.9761	1.6426
	LLWLS($W_1(x)$)	3.9629	1.6383	3.9608	1.6275
	LLWLS($W_4(x)$)	3.9536	1.6253	3.9518	1.6154
	$\tau-$two-step	3.9539	1.6210	3.9091	1.5671
	two-step	**3.9468**	**1.6119**	**3.8898**	**1.5487**
-0.9	LLLS	4.0086	1.6815	3.9933	1.6783
	LLWLS($W_1(x)$)	3.9603	1.6378	3.9317	1.6117
	LLWLS($W_4(x)$)	3.9258	1.5881	3.9024	1.5710
	$\tau-$two-step	**3.9235**	**1.5774**	3.9093	1.5711
	two-step	3.9280	1.5818	**3.9005**	**1.5610**

Table 3.3: DGP 2 with $n = 50, 100$

		n=50		n=100	
Correlation size (ρ)	Estimators	Abs. Bias	MSE	Abs. Bias	MSE
0	LLLS	2.7383	1.1676	2.5591	1.0625
	LLWLS($W_1(x)$)	2.7387	1.1676	2.5591	1.0625
	LLWLS($W_4(x)$)	2.7387	1.1676	2.5591	1.0625
	$\tau-$two-step	2.7203	1.1613	2.5369	1.0557
	two-step	**2.6942**	**1.1532**	**2.5059**	**1.0477**
0.5	LLLS	2.7192	1.1497	2.5506	1.0476
	LLWLS($W_1(x)$)	2.7193	1.1497	2.5506	1.0476
	LLWLS($W_4(x)$)	2.7192	1.1497	2.5506	1.0476
	$\tau-$two-step	2.6997	1.1424	2.5276	1.0399
	two-step	**2.6699**	**1.1332**	**2.4949**	**1.0304**
0.9	LLLS	2.7486	1.1751	2.5544	1.0490
	LLWLS($W_1(x)$)	2.7486	1.1751	2.5544	1.0490
	LLWLS($W_4(x)$)	2.7486	1.1751	2.5544	1.0490
	$\tau-$two-step	2.7321	1.1667	2.5323	1.0390
	two-step	**2.6987**	**1.1542**	**2.4936**	**1.0253**
-0.5	LLLS	2.7299	1.1613	2.5632	1.0575
	LLWLS($W_1(x)$)	2.7300	1.1614	2.5632	1.0575
	LLWLS($W_4(x)$)	2.7300	1.1613	2.5632	1.0575
	$\tau-$two-step	2.7090	1.1540	2.5419	1.0501
	two-step	**2.6800**	**1.1454**	**2.5119**	**1.0413**
-0.9	LLLS	2.7681	1.1880	2.5553	1.0435
	LLWLS($W_1(x)$)	2.7681	1.1880	2.5553	1.0435
	LLWLS($W_4(x)$)	2.7681	1.1880	2.5553	1.0435
	$\tau-$two-step	2.7472	1.1791	2.5295	1.0327
	two-step	**2.7125**	**1.1669**	**2.4891**	**1.0191**

Chapter 4

Return on Public Capital in U.S.: An Application

4.1 Introduction

In order to provide a practical example for applying our newly proposed method to a real data setting, this section is devoted to revisit the relationship between public capital and regional economic performance. The debate on the role of public capital has caused extensive attentions from economists. There are substantial works studying the relationship between public capital and regional economic performance within the United States. To sum up, the empirical works have reached three different conclusions. Some scholars conclude that the public capital played a positive and significant role in effecting the regional productivity, see, for example, Munnell (1990). Some economists hold an opposite conclusion that that the public infrastructure had significant but negative effects on private productivity (see, e.g., Evans and Karras (1994)). The third type of argument is that the contribution of the public infrastructure to private sector is statistically insignificant (see, e.g., Holtz-Eakin (1994) and Baltagi and Pinnoi (1995)).

Notice that the extensive empirical analysis is conducted within the parametric framework by assuming a particular production function and constant elasticity across all the states and all the years.

As we know, nonparametric method has two big advantages compared with parametric regression analysis. One is that the former is free from the notorious function misspecification issue. The other is that nonparametric regression estimation provides local estimates so that we can clearly examine the changing pattern in returns to inputs across all states and years. Even though there are some big advantages in using nonparametric analysis, very few works on this topic employ nonparametric methods, see, for instance, Henderson and Ullah (2008). By utilizing the discussed nonparametric techniques in the previous chapters, we examine the role of public capital in affecting the regional economic performance. The data employed here is the widely used data set from Munnell (1990), which incorporates a panel data of U.S. 48 contiguous states over the period of 1970-1986.

4.2 Model Specification and Estimation

To provide detailed analysis, we adopt the following three nonparametric models:

1. Model 1 is conventional nonparametric one-way random effect model for all U.S. 48 contiguous states, which actually is a special case of the NP SUR model with error components discussed in chapter 3 with $M = 1$, and two error components. It is specified as

$$Y_{it} = m(KG_{it}, KPR_{it}, L_{it}, UNEM_{it}) + \alpha_i + \varepsilon_{it},\ i = 1, \cdots, 48,\ t = 1, \cdots, 17,$$

where Y_{it} denotes the GDP of state i in period t, KG denotes public capital,

KPR is the private capital stock, L is employment, and $UNEM$ stands for the unemployment rate used to control for business cycle effects as commonly used in the literature. All variables in the model except for $UNEM$ are measured in logarithms. This model is an analogue to the parametric setting in Baltagi and Pinnoi (1995).

2. Model 2 is NP SUR model with three error components with $M = 2$. The model has the following form

$$\begin{aligned} Y_{it,j} &= m_j(KG_{it,j}, KPR_{it,j}, L_{it,j}, UNEM_{it,j}) + \alpha_{i,j} + v_{t,j} + \varepsilon_{it,j}, \\ i &= 1, \cdots, 24,\ t = 1, \cdots, 17,\ j = 1, 2. \end{aligned}$$

   In order to apply this model to the data set, we divide 48 states into two regions, low productivity region ($j = 1$) and high productivity region ($j = 2$), according to the states' ranking in terms of 1986 gross state product. By doing so, each region has 24 states. The states within the same productivity region may have similar behavior. Under these settings, we estimate these two groups jointly.

3. Model 3 is estimating nonparametric one-way random effect model for each region separately. The model specification is as model 1. The only difference is that we are using 24 states for each region, and estimate each group separately.

To compare with parametric results, this chapter also estimates the following parametric random effect model which follows Baltagi and Pinnoi (1995),

$$Y_{it} = \beta_1 KG_{it} + \beta_2 KPR_{it} + \beta_3 L_{it} + \beta_4 UNEM_{it} + \alpha_i + \varepsilon_{it},\ i = 1, \cdots, 48,\ t = 1, \cdots, 17.$$

All variables in the above parametric and nonparametric models are measured in logarithms with only one exception of $UNEM$. Therefore, the coefficients of KG, KPR, and L are elasticities in parametric settings. The derivatives of KG, KPR, and L in nonparametric settings are local elasticities.

We use our two-step estimation to estimate the above three nonparametric models. The second order Epanechnikov kernel is used in nonparametric random effect analysis. The second order Gaussian kernel is used in the first step of NP SUR analysis, and in the second step, we still use the second order Epanechnikov kernel. The rule-of-thumb bandwidth is used throughout the nonparametric analysis. As we know, in nonparametric analysis, kernel function doesn't play an important role, but the bandwidth does. So the latter should be chosen cautiously. All the bootstrapped standard errors are calculated from 500 repetitions which are given in parenthesis under the estimates. In model 1 and model 3, σ_α^2 and σ_ε^2 are estimated by using the consistent estimators proposed in Ruckstuhl, Welsh, and Carroll (2000 p. 61). For model 2, the variance-covariance is estimated by the method in Avery (1977).

4.3 Empirical Analysis

The estimation results for mean and median elasticity across all 48 states over the time period are reported in Table 4.1 on page 79. The estimated mean and median elasticity for low productivity region and high productivity region are listed in Table 4.2 on page 79.

From Table 4.1, we have the following findings: (1) All three models give statistically significant public capital elasticity, private capital elasticity, and labor elasticity. Although the magnitude of public capital elasticity is smaller than the private capital, it is nonnegligible. (2) Model 2 gives very similar results to those obtained by model 1, and model 3 tends to give smaller public capital elasticity and larger private capital elasticity than the other two models. (3) By adding up the elasticity of public capital, private capital, and labor for all three models, the sum of elasticity is a bit larger than 1, which might suggest increasing returns to scale. In addition, the estimated correlations for the error components across two regions in model 2 are $\widehat{corr}(\alpha_{i,1}, \alpha_{i,2}) = 0.2897$, $\widehat{corr}(v_{t,1}, v_{t,2}) = 0.8576$, and $\widehat{corr}(\varepsilon_{it,1}, \varepsilon_{it,2}) = -0.0129$, for $i = 1, \cdots, 24$, $t = 1, \cdots, 17$. These results suggest that the correlations across two regions for individual error component and time error component should be considered. Hence, it should be more efficient to use the NP SUR model (model 2) to estimate these two regions jointly than using NP random effect model to estimate each region separately. This may also give the explanation that model 1 and model 2 have very close results, but model 3 has somehow discrepancy. Also, comparing the bootstrapped standard errors in the parenthesis under the estimates, we can find that in general model 2 has the smallest bootstrapped standard errors, and model 3 has the largest standard errors. The explanation behind this perhaps is that model 2 and model 3 divide 48 states into two regions according to productivity, and the states may have more common behaviors with the ones in the same region than those in the other region. And model 2 estimates the two regions jointly by using NP SUR with error components, which not only incorporate the common behavior of the states within each region, but also consider the associations across two regions. The estimated correlations for the error components across two regions in model 2 also confirm that the associations across two regions shouldn't be ignored.

Table 4.2 compares the mean and median elasticity across low productivity region and high productivity region. First, from model 2, we can obviously see that the high productivity region has larger mean/median elasticity of public capital, larger mean/median elasticity of labor, and smaller mean/median elasticity of private capital than the low productivity region. These results are consistent with that public capital and labor are complements, but public capital and private capital are substitutes. Second, all the mean/median elasticities of KG, KPR, and L are statistically significant at 10% level with the only exception of the mean elasticity of KG for high productivity region in model 3. Third, similarly with the results given in Table 4.1, here model 3 doesn't give very close results to model 2. Comparing the estimates for low productivity region and the ones for high productivity region in model 3, we can still observe that higher mean/median elasticity of public capital in high productivity region, higher mean/median labor elasticity, and lower mean/median private capital elasticity, which is consistent with Table 4.1 and the literature. Additionally, generally we can see the lower bootstrapped standard errors in model 2 compared to model 3, which implies that incorporating the correlations across two regions into estimation improves the efficiency of estimator.

To compare with parametric analysis, we also estimate the parametric random effect model. The estimated elasticities of KG, KPR, and L are 0.0044, 0.3105, and 0.7297, respectively. The standard deviation of the estimated elasticity of KG is 0.0234, which shows KG is statistically insignificant. The standard deviations of the estimated elasticity of KPR and L are 0.0198 and 0.0249, respectively. KPR and L are statistically significant. Obviously, all the nonparametric results are different from the parametric results. First, according to nonparametric analysis the public capital has positive and significant effect on state GDP. However, the parametric results show the tiny magnitude of the estimated elasticity of public capital and its insignificance. Second, the estimated elasticities of KPR and L obtained in the parametric regression are larger than those obtained in nonparametric analysis. Third, note that by the parametric random effect model, we assume the constant elasticity across all the states and over the entire period, which is irrealistic in reality. However, the nonparametric analysis provides local estimation. Hence, the latter can reveal more information than the former does.

In order to picture the changes in the elasticity of public capital over the time period and the differences among states, we plot figure Figure 4.1- 4.4 on pages 80 - 83 to help us know deep information behind the regression estimations. Figure 4.1 plot the mean elasticity of public capital across states over 1970-1986 obtained from three models. The solid line is for model 1, the line with dots denotes model 2, and the line with triangles is the elasticity from model 3. The shaded area indicates the recession periods according to NBER. From Figure 4.1, we can see that the mean elasticity of public capital changes over the time, and three models give a consistent result that elasticity increases during recessions, and decreases when the economy started recovering. In addition, Figure 4.1 clearly shows that model 1 and model 2 give very similar results, but model 3 has discrepancy from them. As mentioned earlier, the reason behind maybe in that model 1 and model 2 estimate 48 states jointly, i.e., the connections among these 48 states are incorporated into the estimation of these two models. However, model 3 estimates two regions separately, so the associations among the states from two different regions are ignored. Since model 1 and model 2 give very close results, we only report the estimations obtained from model 1 in the following figures 4.2- 4.4.

Figure 4.2 plots the mean elasticity of public capital for California, New York, South Dakota, and Wyoming. First, Figure 4.2 shows that the elasticity is not constant over the time period of 1970-1986. Second, Figure 4.2 presents the same pattern as Figure 4.1 shows, that is, during the recession periods these four states all have increasing public capital elasticity. This same pattern revealed by Figure 4.1 and Figure 4.2 implies that the government investment on infrastructure during recessions has positive effect on economic productivity. In other words, the fiscal policy indeed plays a positive role in spurring the economy during recession periods. Also, we notice that shortly after each recession period, the elasticity of public capital falls. This may because that after the economy steps out of the recession, the private sector becomes strong, and private capital increases. As a result, the public capital elasticity decreases, which may due to the substitutes effect between public capital and private capital. Third, it is clear to see that both California and New York have positive returns over the studied period. However, South Dakota has consistent negative returns over 1970-1986, and Wyoming has negative returns in most of years.

Figure 4.3 presents the mean elasticity of public capital over 1970-1986 for all 48 states. From Figure 4.3, Montana, New Mexico, North Dakota, South Dakota, and Wyoming have negative returns to public capital on average. North Dakota and South Dakota are from plains, Montana and Wyoming are from rocky mountain region, and New Mexico is from southwest, according to BEA regions. To explore the reason behind these negative returns, we further examine the original data set. From the original data, we calculate the ratio of public capital to GSP and the public capital per labor. We find that the ratios of public capital to GSP for these states with negative returns are higher than the average of the ratios across states, with the only exception of Wyoming, which has a little bit lower ratio (0.42) than the average of ratios (0.45). And South Dakota has the highest ratio (0.65) among all 48 states. By examining the public capital per labor, these five states all have higher levels than the average of public capital per labor across states (15.3). Specifically, Wyoming (26.0) has the highest level, and South Dakota (21.8) is the second to the highest level. The literature has shown the public capital and labor are complements. Hence, the higher levels in public capital per labor for those states with negative public returns may imply the less efficiency in utilizing the public capital. The information from the data gives an explanation that Wyoming has the highest negative returns to public capital among 48 states, and South Dakota has the second highest negative returns to public capital.

Figure 4.4 provides the picture for the elasticity of public capital for all 48 states in 1970, 1982, and 1986. The line with one dot in-between denotes the year 1970, the solid line indicates the year 1982, and the line with two dots in-between stands for the year 1986. Figure 4 tells that for most of states the elasticity of public capital in year 1982 is higher than in the years 1970 and 1986. One may be curious on what happened in the U.S. economy around 1982. In the early 1980s, the U.S. economy suffered a severe recession, which had being widely characterized as a "W-shaped" (also known as "double dip") recession. According to NBER, there were two recessions recorded in the early 1980s. The U.S. economy first stepped into recession in January 1980, followed by a short period of recovering from August 1980 to June 1981, and then dipped back into a severer recession for the period of July 1981-November 1982.[1] Combined Figure 4.4 with Figure 4.1 in which the mean elasticity across states reached the highest level in 1982, our estimation results suggest that the public capital played the most efficient role for the majority of states in the most serious recession during the examined years.

4.4 Concluding Remarks

In this section, we apply the newly developed two-step nonparametric estimation method in chapter 2 and chapter 3 to a real data setting. The relationship between public capital and regional economic performance in U.S are analyzed by using nonparametric random effect model and SUR model with error components. To sum up, there are three interesting findings. The last two findings can only obtained by nonparametric method. The first finding is that the public capital plays an effective role in economic

[1]Source: NBER

performance because the elasticity of public capital in US is positive and significant. The last two findings are: one is that the elasticity of public capital across states increases when the U.S. economy stepped into the recessions, and decreases when the economy started booming. This finding implies the government plays a more effective role during the recession periods than the normal periods. The other is that five states in U.S. have negative returns on public capital. By further examining the data, I find that the ratio of public capital to state GDP and the public capital per labor for these five states are much higher than the rest of states. These imply that these five states overinvest on the infrastructure and inefficiency in using the public capital. Nonparametric can provide us these stories as it gives local estimates for elasticity of public capital for each state at any particular year, whereas parametric estimation cannot.

Table 4.1: Mean and median elasticity across states

	KG	KPR	L	$UNEM$
Model 1				
mean elasticity	0.1314	0.2852	0.6326	-0.0041
	(0.0510)	(0.0265)	(0.0420)	(0.0036)
median elasticity	0.1550	0.2742	0.6501	-0.0027
	(0.0433)	(0.0257)	(0.0339)	(0.0040)
Model 2				
mean elasticity	0.1313	0.2836	0.6342	-0.0042
	(0.0328)	(0.0181)	(0.0280)	(0.0027)
median elasticity	0.1530	0.2696	0.6523	-0.0030
	(0.0271)	(0.0183)	(0.0219)	(0.0030)
Model 3				
mean elasticity	0.1049	0.3094	0.6589	-0.0055
	(0.0447)	(0.0481)	(0.0846)	(0.0125)
median elasticity	0.1358	0.3646	0.5902	-0.0059
	(0.0723)	(0.0524)	(0.0280)	(0.0033)

Table 4.2: Mean and median elasticity for two regions

	KG	KPR	L	$UNEM$
Low Productivity Region				
Model 2				
mean elasticity	0.0911	0.3083	0.6274	-0.0035
	(0.0539)	(0.0227)	(0.0315)	(0.0033)
median elasticity	0.1214	0.2899	0.6457	-0.0021
	(0.0341)	(0.0207)	(0.0236)	(0.0034)
Model 3				
mean elasticity	0.1051	0.3812	0.5604	-0.0040
	(0.0538)	(0.0398)	(0.0415)	(0.0043)
median elasticity	0.1267	0.3839	0.5422	-0.0039
	(0.0528)	(0.0330)	(0.0398)	(0.0047)
High Productivity Region				
Model 2				
mean elasticity	0.1715	0.2590	0.6409	-0.0048
	(0.0395)	(0.0249)	(0.0452)	(0.0036)
median elasticity	0.1765	0.2579	0.6599	-0.0040
	(0.0372)	(0.0244)	(0.0329)	(0.0039)
Model 3				
mean elasticity	0.1047	0.2376	0.7574	-0.0070
	(0.0797)	(0.0909)	(0.1629)	(0.0251)
median elasticity	0.1543	0.2786	0.6610	-0.0080
	(0.0461)	(0.0352)	(0.0424)	(0.0045)

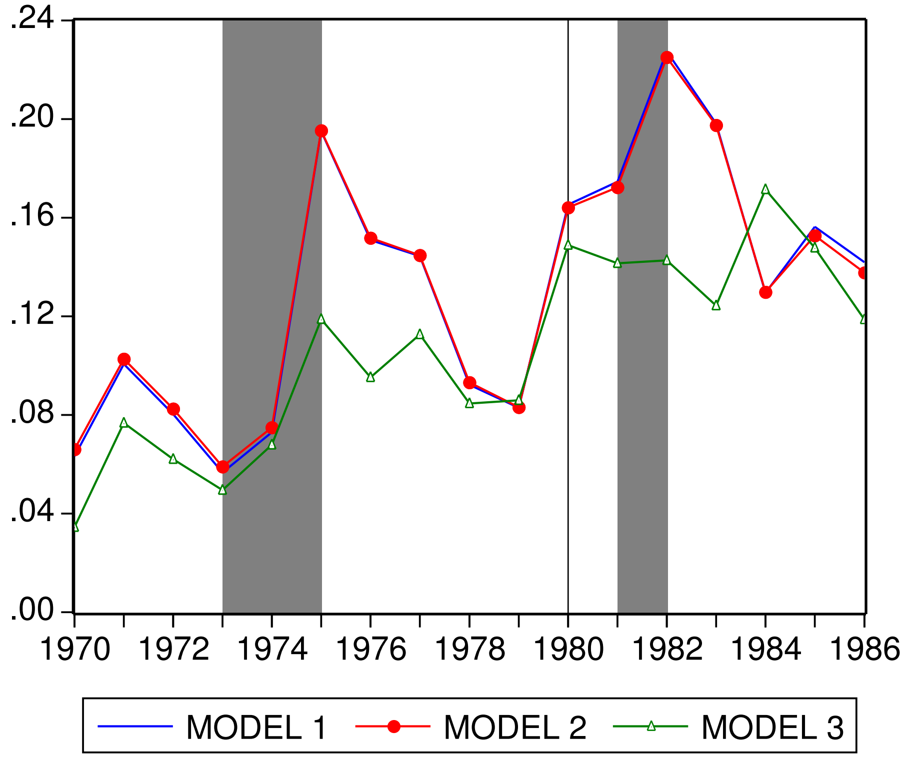

Figure 4.1: Mean Elasticity of Public Capital over 1970-1986

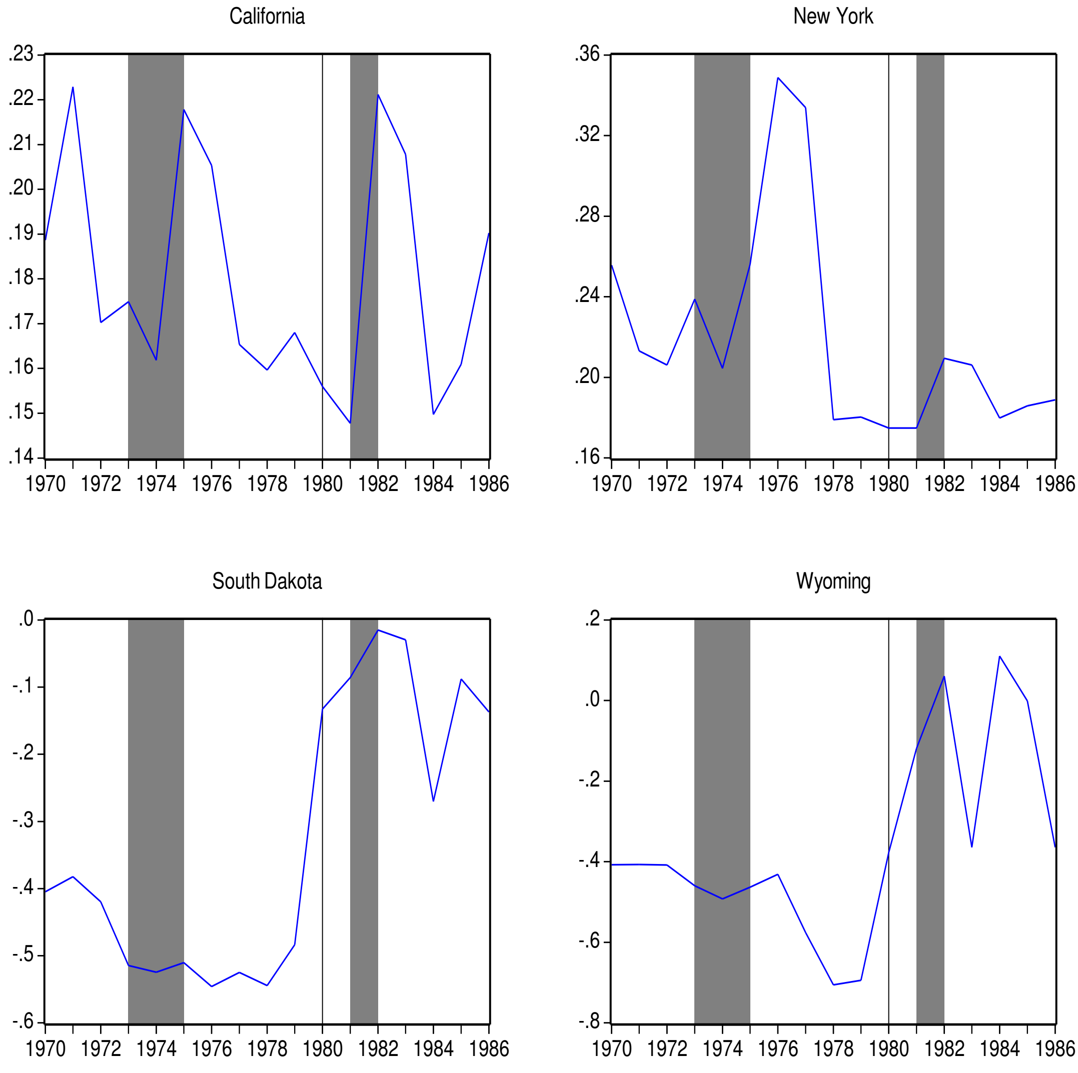

Figure 4.2: Elasticity of Public Capital over 1970-1986

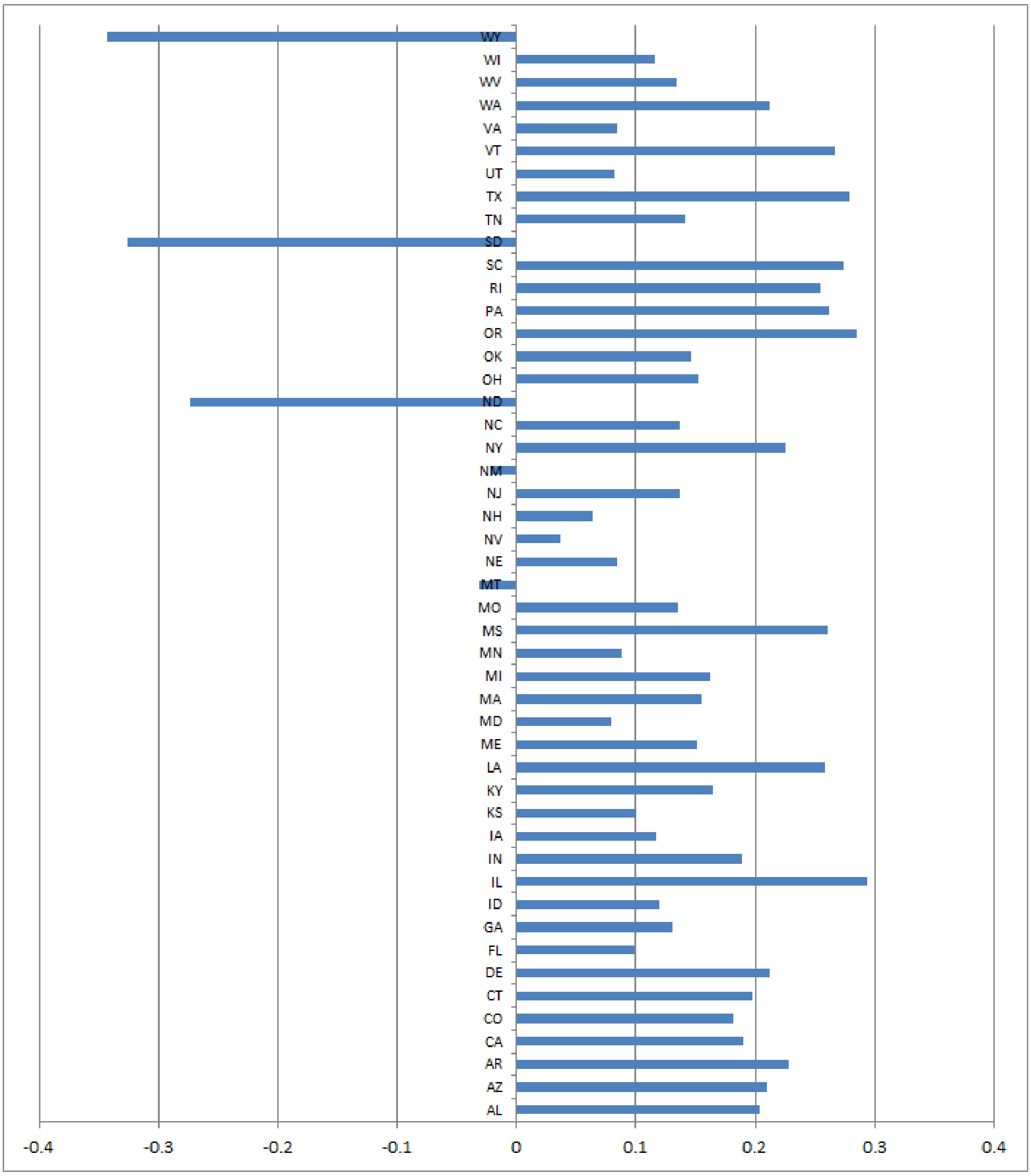

Figure 4.3: Elasticity of Public Capital across States

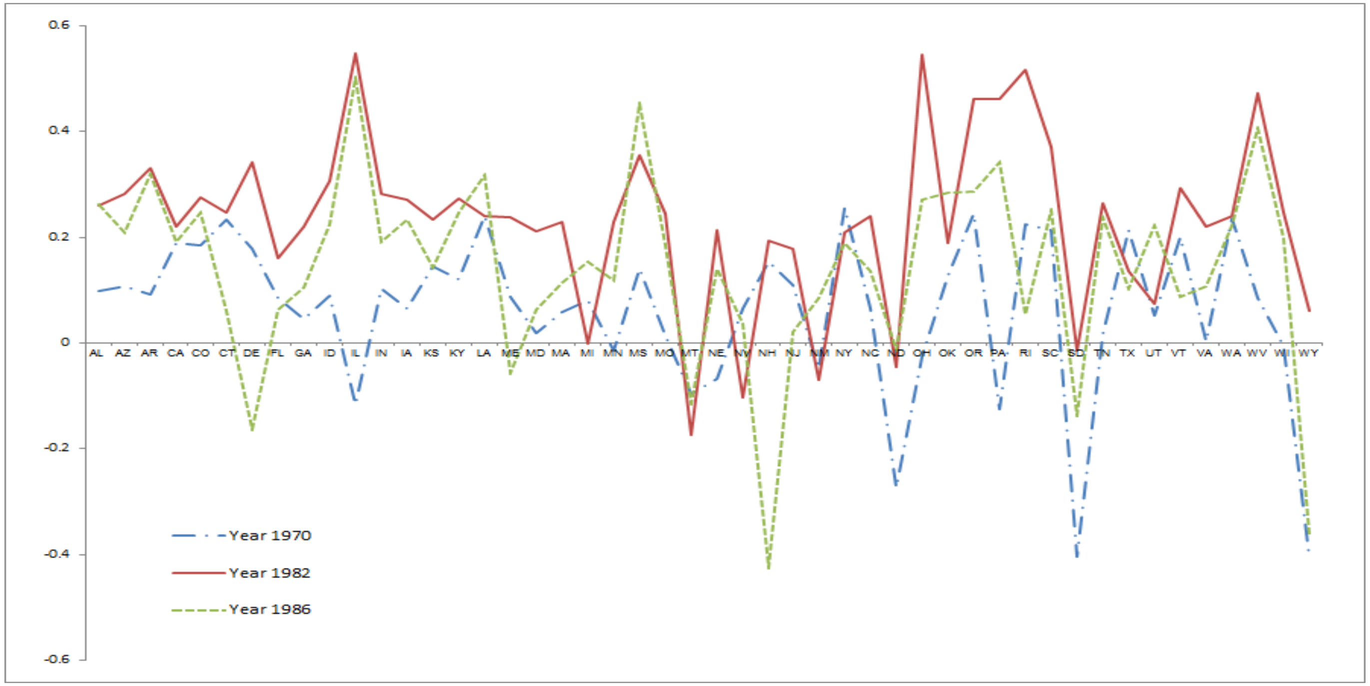

Figure 4.4: Elasticity of Public Capital for Selected Years

Chapter 5

Approximate Moments of Mean Reversion Parameter Estimator in Continuous Time Gaussian and Lévy Processes *

5.1 Introduction

In recent years, an extensive literature has developed on using diffusion processes to model the dynamic behavior of financial securities. For example, Vasicek (1977) used the following Ornstein-Uhlenbeck (OU) process to model the spot interest rate,

$$dX_t = \kappa(\mu - X_t)dt + \sigma dB_t, \tag{5.1}$$

where B_t is a standard Brownian motion. This is a Gaussian Markov process and possesses a stationary distribution when $\kappa > 0$. In this case, κ captures the rate of

*This chapter is a joint work with Dr. Aman Ullah and Dr. Jun Yu convergence towards its long term mean, μ. Tang and Chen (2009) considered a more general form of a Brownian motion based continuous time model (i.e. diffusion process),

$$dX_t = \kappa(\mu - X_t)dt + \sigma(X_t; \theta)dB_t, \tag{5.2}$$

where $\sigma(X_t; \theta)$ is the diffusion function of X_t at time t. If $\sigma(X_t; \theta) = \sigma\sqrt{X_t}$, the diffusion process becomes the CIR model (Cox, Ingersoll, and Ross, 1985). A even more general diffusion process is given by,

$$dX_t = \mu(X_t; \theta)dt + \sigma(X_t; \theta)dB_t, \tag{5.3}$$

with a general drift function $\mu(X_t; \theta)$. An important special case is when $\mu(X_t; \theta) = \mu X_t$ and $\sigma(X_t; \theta) = \sigma X_t$. Black and Scholes (1973) used it to model the spot price of a stock.

All these processes are Brownian-motion based. Under some smoothness conditions on the drift function and the diffusion function, the sample path generated from X_t is continuous everywhere. In recent years, however, it has been reported strong evidence of infinite activity jumps in financial variables; see, for example, Aït-Sahalia and Jacod (2008). To capture the infinite activity jumps, continuous time Lévy processes have become increasingly popular and various Lévy models have been developed in the asset pricing literature (see for example, Barndorff-Nielsen (1998), Madan, Carr and Chang (1999), Carr and Wu (2003)).

In practice, one can only obtain the observations at discrete time points from a finite time span. Let T be the time span, h the sampling interval, and n $(= T/h)$ the number of observations. Hence, $T < \infty$ and $h > 0$. Based on discrete time observations, different methods have been used to estimate the continuous time models. Phillips and Yu (2009c) provided an overview of some widely used estimation methods. When the drift function is linear and slowly mean reverting, it is found that there is serious estimation bias in the mean reversion parameter, κ, by almost all the methods.

Because this parameter is of important implications for asset pricing, risk management and forecasting, how to accurately estimate this parameter has received considerable attentions in the literature. For example, Yu (2009) approximates the bias of the maximum likelihood estimator (MLE) of κ when the long run mean is known and the initial condition is the marginal distribution under the Gaussion OU process. Tang and Chen (2009) approaximates the bias of MLE of κ when the long run mean is unknown under the Gaussion OU process and the CIR model. To reduce the estimation bias of κ, Phillips and Yu (2005) proposed the jackknife method. While the jackknife increases the variance, a carefully designed jackknifing procedure can offer substain improvement in reducing the bias, leading to a decrease in the root mean square errors (RMSE). To further reduce RMSE, Phillips and Yu (2009b) proposed the indirect inference method while Tang and Chen (2009) proposed a parametric bootstrapping method. These two methods are simulation-based and hence numerically more demanding.

The difficulty in the estimation of $\widehat{\kappa}$ is not unexpected because it is related to the finite sample bias problem well documented in the discrete time literature; see, for example, Kendall (1954). However, the magnitude of the bias in $\widehat{\kappa}$ is very large in practically relevant cases to the U.S. data so that the implications for the bias become very important. For example, Phillips and Yu (2005) showed that the bias of maximum likelihood estimator for κ in the CIR model can be over 200% even though 25 years of data were used (regardless of the sample frequency). They further reported evidence that the bias in the drift term estimation are even worse than that caused by the discretization and even that caused by a misspecification of the diffusion function. The simulation results of Phillips and Yu (2005) and Tang and Chen (2009) show that the bias of the long run mean (μ) and parameters in the diffusion function are virtually zero. For instance, in the stationary Vasicek model, as Tang and Chen (2009) showed, the bias of $\widehat{\kappa}$ is up to $O(T^{-1})$, while the bias of σ and μ is $O(n^{-1})$ and $O(n^{-2})$, respectively, as $T \to \infty$ with h is fixed.

While the bias in $\widehat{\kappa}$ has been well studied in the continuous time diffusion process, to the best of our knowledge, nothing has been reported on the bias in $\widehat{\kappa}$ in the continuous time Lévy process. The objective of this chapter is to approximate the bias of $\hat{\kappa}$ under the Lévy measure, then study the effects of nonnormality on the estimation bias. Quasi maximum likelihood (QML)/OLS is used to estimate κ which makes it feasible the analytical expression for $\widehat{\kappa}$. We present the results on the bias under the assumption where the error term follows a non-Gaussian distribution with finite first eight moments. It is found that the kurtosis has a negative effect on the bias of $\hat{\kappa}$. The skewness has a positive (positive) effect on the bias of $\hat{\kappa}$ if the distribution has negative (positive) skewness. In addition, under the Gaussian OU process the initial condition has non-monotonic effect on the bias of $\hat{\kappa}$ and the bias of $\hat{\kappa}$ is a monotonically increasing function of the diffusion parameter, σ. A bias corrected estimator of $\hat{\kappa}$ is proposed. The simulation results show that our proposed estimator generally performs well in terms of bias and the root mean square error, especially, when κ is small. Small values of κ correspond to the near unit root situation and is empirically relevant for financial variables in the U.S., such as interest rates and volatility.

The structure of this chapter is as follows. In Section 2, we introduce a continuous time Lévy process and derive the bias of $\hat{\kappa}$. Section 3 derives the bias of $\hat{\kappa}$ in a higher order term. Section 4 reports the simulation results. Section 5 concludes.

5.2 Parameter Estimation for Lévy Processes

5.2.1 Continuous Time Lévy Process

As argued before, while the diffusion processes are very useful, empirical evidence of infinitely active jumps has been found in data. In this chapter we extend the Gaussian OU model of Vasicek to a Lévy-based OU model:

$$dX_t = \kappa (\mu - X_{t-}) \, dt + \sigma dL_t, \tag{5.4}$$

where $(L_t)_{t\geq 0}$ is a Lévy process defined on $(\Omega, \Theta, \{\Theta\}, P)$ with $L_0 = 0$ and satisfies the following three properties:

1. Independent increments: for every increasing sequence of times $t_0, \ldots, t_n$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent;
2. Stationary increment: the law of $X_{t+h} - X_t$ is independent of t;
3. Stochastic continuity: for all $\varepsilon > 0$, $\lim_{h\to 0} P(|X_{t+h} - X_t| \geq \varepsilon) = 0$. For a given t, the probability of seeing a jump at t is zero. In other words, jumps happen at random times.

Obvisouly, the Brownian motion is a special case of the Lévy process and, hence, the Vasicek model is a special case of Model (5.4). Other well known examples include the Poisson process, the gamma process, the variance gamma process, and the α-stable process. While the Brownian motion has a continuous sample path, it does not allow for any jumps. The Poisson process allows for jumps. However, the jump is of finite activity. General Lévy processes allow an infinite number of jumps within any time interval. Also, general Lévy processes allow non-Gaussian increments.

Protter (1990, Theorem 7) showed that the unique solution exists for Model (5.4). If $\mu = 0$ and is known a priori, Model (5.4) becomes

$$dX_t = -\kappa X_{t-}dt + \sigma dL_t. \tag{5.5}$$

Based on the Ito's lemma, the exact discrete time model of (5.5) is given by,

$$X_{ih} = \exp(-\kappa h)X_{(i-1)h} + \sigma\sqrt{\frac{1-\exp(-2\kappa h)}{2\kappa}}\varepsilon_i, \tag{5.6}$$

where the distribution of ε_i depends on the specification of the Lévy measure $L(t)$. This is a discrete time AR(1) model with a possibly non-Gaussian error term. When $L(t)$ is the Brownion motion, $\varepsilon_i \sim N(0,1)$. If $L(t)$ is the variance gamma process of Madan and Seneta (1990) (i.e. $L(t) = B(\gamma(t;1,\nu))$ where $\gamma(t;1,\nu)$ is a gamma distribution with mean 1 and variance ν), then ε_i follows the variance gamma distribition whose density and the moment generate function are given by,

$$f(x) = \int_0^\infty \frac{1}{\sqrt{2\pi g}}e^{-x^2/(2g)}\frac{g^{1/\nu-1}e^{-g/\nu}}{\Gamma(1/\nu)\nu^{1/\nu}}dg, \tag{5.7}$$

and

$$mgf(u) = \left(1-\nu u^2/2\right)^{-1/\nu}, \tag{5.8}$$

where Γ is the gamma function. The variance gamma distribution is conditional Gaussian given that the conditional variance is distributed as a gamma variate whose mean is 1 and variance is ν. It is known that, for the variance gamma distribution, the moments of all orders exist with the mean 0, the variance 1, and the kurtosis $3+3\nu$. Since the excess kurtosis is determined by the parameter ν, ν measures the degree of tail thickness.

If $L_t = B_t$ and $X_0 = x_0$. the exact discrete time model of (5.4) is

$$X_{ih} = \phi X_{(i-1)h} + \mu(1-e^{-\kappa h}) + \sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_i, \varepsilon_i \sim N(0,1), X_0 = x_0, \tag{5.9}$$

where $\phi = e^{-\kappa h}$. When $\kappa \to 0$ or $h \to 0$, $\phi \to 1$ and Equation (5.9) has a unit root in the limit. To simplify notation, we write X_{ih} as X_i. The transition density in Equation

(5.9) is

$$X_i \mid X_{i-1} \sim N\left\{X_{i-1}e^{-\kappa h} + \mu(1-e^{-\kappa h}), \frac{1}{2\kappa}\sigma^2(1-e^{-2\kappa h})\right\}, \tag{5.10}$$

facilitating the maximum likelihood (ML) estimation, or equivalently ordinary least squares (OLS) estimation of κ,

$$\widehat{\kappa} = -\frac{\ln\hat{\phi}}{h} \text{ with } \hat{\phi} = \frac{n^{-1}\sum_{i=1}^{n} X_i X_{i-1} - n^{-2}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_{i-1}}{n^{-1}\sum_{i=1}^{n} X_{i-1}^2 - n^{-2}\left(\sum_{i=1}^{n} X_{i-1}\right)^2}. \tag{5.11}$$

Taking the Taylor expansion to the second order, we obtain

$$\begin{aligned}
\widehat{\kappa} &= -\frac{\ln\phi}{h} - \frac{1}{h}\left(\frac{\hat{\phi}-\phi}{\phi}\right) + \frac{1}{2h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^2 + o_p(T^{-1}),\\
\widehat{\kappa} - \kappa &= -\frac{1}{h}\left(\frac{\hat{\phi}-\phi}{\phi}\right) + \frac{1}{2h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^2 + o_p(T^{-1}),\\
E(\widehat{\kappa}) - \kappa &= -\frac{1}{h\phi}E\left(\hat{\phi}-\phi\right) + \frac{1}{2h\phi^2}E\left(\hat{\phi}-\phi\right)^2 + o(T^{-1})\\
&= -\frac{Bias(\hat{\phi})}{h\phi} + \frac{MSE(\hat{\phi})}{2h\phi^2} + o(T^{-1}),
\end{aligned} \tag{5.12}$$

where $MSE(\hat{\phi}) = E\left(\hat{\phi}-\phi\right)^2$ represents the mean square errors (MSE) of $\hat{\phi}$.

For general Lévy processes, the transition density is not Gaussian any more. As a result, $\hat{\phi}$ and, hence, $\widehat{\kappa}$ is not the MLE. However, $\hat{\phi}$ and $\widehat{\kappa}$ are the QMLEs and can be obtained by OLS. Although the QML/OLS is not as efficient as the ML, it is analytically more tractable. To approximate the bias of $\widehat{\kappa}$, we follow Bao and Ullah (2010) and make the same assumptions about ε_i. In particular, we assume ε_i is *i.i.d* and follows a distribution with eight moments:

$$\begin{aligned}
m_1 &= 0, m_2 = 1, m_3 = \gamma_1, m_4 = \gamma_2 + 3, \qquad (5.13)\\
m_5 &= \gamma_3 + 10\gamma_1, m_6 = \gamma_4 + 15\gamma_2 + 10\gamma_1^2 + 15,\\
m_7 &= \gamma_5 + 21\gamma_3 + 35\gamma_2\gamma_1 + 105\gamma_1,\\
m_8 &= \gamma_6 + 28\gamma_4 + 56\gamma_3\gamma_1 + 35\gamma_2^2 + 210\gamma_2 + 280\gamma_1^2 + 105,
\end{aligned}$$

where γ_1 and γ_2 are the Pearson's measures of skewness and kurtosis of the distribution and $\gamma_1, \ldots, \gamma_6$ can be regarded as measures for deviation from normality. For a normal distribution, $\gamma_1, \ldots, \gamma_6$ all equal 0.

5.2.2 Bias approximation when the long run mean is known

When $\mu = 0$ and it is known, the exact discrete time model of the Lévy process is,

$$X_{ih} = \phi X_{(i-1)h} + \sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_i. \tag{5.14}$$

Bao and Ullah (2007 , 2010) and Bao (2007) give the approximate bias and the MSE of the OLS estimator for the AR(1) model without intercept but with a non-Gaussian error term:

$$\begin{aligned}
Bias(\hat{\phi}) &= -\frac{2\phi}{n} + o(n^{-1}) \\
MSE(\hat{\phi}) &= \frac{1-\phi^2}{n} + \frac{1}{n^2}\left[14\phi^2 - 1 - \frac{(1-\phi^2)x_0^2}{\sigma_0^2} - \frac{4\gamma_1\phi(1-\phi^2)^2}{1-\phi^3} - \gamma_2(1-\phi^2)\right] \\
&\quad + o(n^{-2}),
\end{aligned}$$

where x_0 is fixed. In the Guassian case, $\varepsilon_t \sim$iid$N(0,1)$, we have, for fixed x_0,

$$\begin{aligned}
E(\widehat{\kappa} - \kappa | x_0) &= -\frac{1}{h\phi}\left(-\frac{2\phi}{n}\right) + \frac{1}{2h\phi^2}\left[\frac{1-\phi^2}{n} + \frac{1}{n^2}\left(14\phi^2 - 1 - \frac{2\kappa x_0^2}{\sigma^2}\right)\right] + o(T^{-1}) \\
&= \frac{e^{2\kappa h}+3}{2T} + \frac{1}{2nT}\left[14 - e^{2\kappa h} - \frac{2\kappa e^{2\kappa h}x_0^2}{\sigma^2}\right] + o(T^{-1}).
\end{aligned}$$

In the non-Guassian case, i.e., $\varepsilon_t \sim$iid$(0,1)$, the skewness and the excess kurtosis coefficients matter for approximating the MSE up to $O(T^{-2})$. Consequently, the bias formula, for fixed x_0, can be obtained as

$$
\begin{aligned}
E[\widehat{\kappa}-\kappa \quad | \quad x_0] &= -\frac{1}{h\phi}(-\frac{2\phi}{n})+\frac{1}{2h\phi^2}\{\frac{1-\phi^2}{n}+\frac{1}{n^2}[14\phi^2-1-\frac{2\kappa x_0^2}{\sigma^2} \\
&\quad -\frac{4\gamma_1\phi(1-\phi^2)^2}{1-\phi^3}-\gamma_2(1-\phi^2)]\}+o(T^{-1}) \\
&= \frac{e^{2\kappa h}+3}{2T}+\frac{1}{2nT}\{14-e^{2\kappa h}-\frac{2\kappa e^{2\kappa h}x_0^2}{\sigma^2}-\frac{4\gamma_1 e^{\kappa h}(1-e^{-\kappa h})(1+e^{-\kappa h})^2}{e^{-2\kappa h}+e^{-\kappa h}+1} \\
&\quad -\gamma_2(e^{2\kappa h}-1)\}+o(T^{-1}).
\end{aligned}
$$

We summarize the above results in Theorem 2.1.

Theorem 9 *Under Model (5.14) with a known mean, a non-Gaussian error term with moments given in (5.13), and a fixed initial condition x_0, the approximation to the bias of $\widehat{\kappa}$ is given by,*

$$
\begin{aligned}
E[\widehat{\kappa}-\kappa \quad | \quad x_0] = \frac{e^{2\kappa h}+3}{2T}+\frac{1}{2nT}\{14-e^{2\kappa h}-\frac{2\kappa e^{2\kappa h}x_0^2}{\sigma^2} \\
-\frac{4\gamma_1 e^{\kappa h}(1-e^{-\kappa h})(1+e^{-\kappa h})^2}{e^{-2\kappa h}+e^{-\kappa h}+1}-\gamma_2(e^{2\kappa h}-1)\}+o(T^{-1}). \quad (5.15)
\end{aligned}
$$

Corollary 10 *Under the Lévy process (5.14) with a known mean, a non-Gaussian error term with moments given in (5.13), and a random non-Gaussian initial condition x_0 with mean 0 and varinace $\sigma^2/(2\kappa)$, the approximation to the bias of $\widehat{\kappa}$ is given by,*

$$
\begin{aligned}
E(\widehat{\kappa}-\kappa) &= \frac{e^{2\kappa h}+3}{2T}+\frac{1}{2nT}\{14-2e^{2\kappa h}-\frac{4\gamma_1 e^{\kappa h}(1-e^{-\kappa h})(1+e^{-\kappa h})^2}{e^{-2\kappa h}+e^{-\kappa h}+1} \\
&\quad -\gamma_2(e^{2\kappa h}-1)\}+o(T^{-1}). \quad (5.16)
\end{aligned}
$$

In Theorem 9 the results on $Bias(\widehat{\kappa})$ are obtained conditional on x_0. When x_0 is assumed to be random with mean 0 and varinace $\sigma^2/(2\kappa)$, the uncoditional bias is obtained by the iterated expectation, namely, $E(\widehat{\kappa}-\kappa)=E_{x_0}[E(\widehat{\kappa}-\kappa)\mid x_0]$.

Corollary 11 *Under the Lévy process (5.14) with a known mean, a Gaussian error term ($\gamma_1=0$ and $\gamma_2=0$), and a fixed x_0, the approximation to the bias of $\widehat{\kappa}$ is given by,*

$$
E[\widehat{\kappa}-\kappa \mid x_0]=\frac{e^{2\kappa h}+3}{2T}+\frac{1}{2nT}\left[14-e^{2\kappa h}-\frac{2\kappa e^{2\kappa h}x_0^2}{\sigma^2}\right]+o(T^{-1}). \quad (5.17)
$$

Corollary 12 *Under the Lévy process (5.14) with a known mean, a Gaussian error term, and a random Gaussian initial condition* x_0 *with mean* 0 *and varinace* $\sigma^2/(2\kappa)$*, the approximation to the bias of* $\widehat{\kappa}$ *is given by*

$$E(\widehat{\kappa}-\kappa)=\frac{e^{2\kappa h}+3}{2T}+\frac{1}{nT}[7-e^{2\kappa h}]+o(T^{-1}). \tag{5.18}$$

Remark 2.1.1 Here we have considered the bias of the AR(1) coefficient up to $O(n^{-1})$ and the MSE of the AR(1) coefficient up to $O(n^{-2})$ to obtain the results in Theorem 9 and Corollary 10 for the non-Gaussian case, and Corollary 11 and Corollary 12 for the Gaussian case. In Theorem 9 and Corollary 11 the results are obtained conditional on x_0. In Corollary 2.2 and Corollary 2.4 the results are obtained for a random x_0. Yu (2009) derived the result for the bias of $\widehat{\kappa}$ for the case of normality and $x_0 \sim N(0,\sigma^2/(2\kappa))$ as,

$$E(\widehat{\kappa}-\kappa)=\frac{e^{2\kappa h}+3}{2T}-\frac{2(1-e^{-2n\kappa h})}{Tn(1-e^{-2\kappa h})}. \tag{5.19}$$

The first term on the right hand side of (5.19) is the same as the first term in (5.18), but the second term is different.

Remark 2.1.2 The second term in (5.17) incorporates the initial condition x_0, suggesting that the initial condition affects the bias. Notice that if $x_0 > 0 (< 0)$, $\partial Bias(\widehat{\kappa})/\partial x_0 < 0 (> 0)$, implying that the bias is a decreasing (increasing) function of the initial condition.

Remark 2.1.3 Results obtained in Theorem 5.15 show that σ^2, the initial condition x_0, the skewness and the excess kurtosis all affect the bias of $\widehat{\kappa}$. Note that $\partial Bias(\widehat{\kappa})/\partial\gamma_2 < 0$, which implies that the bias is a monotonically decreasing function of the excess kurtosis. If $x_0 > 0 (< 0)$, $\partial Bias(\widehat{\kappa})/\partial x_0 < 0 (> 0)$, implying that the bias is a decreasing (increasing) function of the initial condition when $x_0 > 0(< 0)$. If $\gamma_1 > 0(< 0)$, $\partial Bias(\widehat{\kappa})/\partial\gamma_1 < 0(> 0)$. Moreover, $\partial Bias(\widehat{\kappa})/\partial\sigma^2 > 0$ implies that the bias is a monotonically increasing function of the variance of error terms, σ^2.

5.2.3 Bias approximation when the long run mean is unknown

For the discrete time AR(1) model with an unknown intercept, the second-order bias of the OLS estimator $\hat{\phi}$, up to $O(n^{-1})$, is $Bias(\hat{\phi}) = -\frac{1+3\phi}{n}$, as obtained in Bao and Ullah (2007). The MSE, up to $O(n^{-2})$, is given by Bao and Ullah (2010) as

$$\begin{aligned} M(\hat{\phi}) &= \frac{1-\phi^2}{n} + \frac{1}{n^2}\{23\phi^2 + 10\phi - \frac{1+\phi}{1-\phi}\left(\frac{\beta-(1-\phi)x_0}{\sigma_0}\right)^2 \\ &\quad -\frac{4\gamma_1\phi(1-\phi^2)^2}{1-\phi^3} - \gamma_2(1-\phi^2)\} + o(n^{-2}) \end{aligned}$$

where x_0 is the initial condition, γ_1 the skewness, and γ_2 the excess kurtosis. In a special case of the Gaussian error term, we have,

$$M(\hat{\phi}) = \frac{1-\phi^2}{n} + \frac{1}{n^2}\left[23\phi^2 + 10\phi - \frac{1+\phi}{1-\phi}\left(\frac{\beta-(1-\phi)x_0}{\sigma_0}\right)^2\right] + o(n^{-2}).$$

Substituting above results into (5.12), the bias of $\widehat{\kappa}$ in the Gaussian case is,

$$\begin{aligned} E(\widehat{\kappa}-\kappa|x_0) &= \frac{1+3\phi}{T\phi} + \frac{1-\phi^2}{2T\phi^2} + \frac{1}{2Tn\phi^2}\{23\phi^2 + 10\phi \\ &\quad -\frac{1+\phi}{1-\phi}\left(\frac{\beta-(1-\phi)x_0}{\sigma_0}\right)^2\} + o(T^{-1}) \end{aligned}$$

where $\beta = \mu(1-e^{-\kappa h})$, $\sigma_0 = \sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}$. We can rewrite the bias in terms of κ, h, and μ as,

$$\begin{aligned} E(\widehat{\kappa}-\kappa|x_0) &= \frac{e^{\kappa h}+3}{T} + \frac{e^{2\kappa h}-1}{2T} + \frac{1}{2nT}\left(23 + 10e^{\kappa h} - \frac{2\kappa e^{2\kappa h}(\mu-x_0)^2}{\sigma^2}\right) + o(T^{-1}) \\ &= \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5) + \frac{1}{2Tn}\left(23 + 10e^{\kappa h} - \frac{2\kappa e^{2\kappa h}(\mu-x_0)^2}{\sigma^2}\right) \\ &\quad +o(T^{-1}). \end{aligned}$$

If we consider the case of non-normality, i.e., $\varepsilon_t \sim$iid$(0,1)$, the skewness and the excess kurtosis coefficients show up in the approximate MSE, up to $O(n^{-2})$. Therefore, we can obtain the formula of the bias of $\widehat{\kappa}$, for a fixed x_0, as

$$\begin{aligned} E(\widehat{\kappa}) - \kappa &= \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5) + \frac{1}{2Tn}\{23 + 10e^{\kappa h} - \frac{2\kappa e^{2\kappa h}(\mu - x_0)^2}{\sigma^2} \\ &\quad -\frac{4\gamma_1 e^{\kappa h}(1 - e^{-\kappa h})(1 + e^{-\kappa h})^2}{1 + e^{-\kappa h} + e^{-2\kappa h}} - \gamma_2(e^{2\kappa h} - 1)\} + o(T^{-1}). \end{aligned}$$

The results of the estimation bias of $\widehat{\kappa}$ for Lévy process with an unknow mean are formally stated in the following Theorem.

Theorem 13 *Under the Lévy process (5.9) with an unknown mean, a non-Gaussian error term with moments given in (5.13), and a fixed initial condition x_0, the approximation to the bias of $\widehat{\kappa}$ is,*

$$\begin{aligned} E[\widehat{\kappa} - \kappa \quad | \quad x_0] &= \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5) + \frac{1}{2Tn}\{23 + 10e^{\kappa h} - \frac{2\kappa e^{2\kappa h}(\mu - x_0)^2}{\sigma^2} \\ &\quad -\frac{4\gamma_1 e^{\kappa h}(1 - e^{-\kappa h})(1 + e^{-\kappa h})^2}{1 + e^{-\kappa h} + e^{-2\kappa h}} - \gamma_2(e^{2\kappa h} - 1)\} + o(T^{-1}). \end{aligned} \tag{5.20}$$

Corollary 14 *Under the Lévy process (5.9) with an unknown mean, a non-Gaussian error term with moments given in (5.13), and a random initial condition x_0 whose mean is μ and varinace $\sigma^2/(2\kappa)$, the approximation to the bias of $\widehat{\kappa}$ is*

$$\begin{aligned} E(\widehat{\kappa} - \kappa) &= \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5) + \frac{1}{2Tn}\{23 + 10e^{\kappa h} - e^{2\kappa h} \\ &\quad -\frac{4\gamma_1 e^{\kappa h}(1 - e^{-\kappa h})(1 + e^{-\kappa h})^2}{1 + e^{-\kappa h} + e^{-2\kappa h}} - \gamma_2(e^{2\kappa h} - 1)\} + o(T^{-1}) \end{aligned} \tag{5.21}$$

In Theorem 13 the results on $Bias(\widehat{\kappa})$ are obtained conditional on x_0. When x_0 is assumed to be random with mean 0 and varinace $\sigma^2/(2\kappa)$, the uncoditional bias is obtained by the iterated expectation, namely, $E(\widehat{\kappa} - \kappa) = E_{x_0}[E(\widehat{\kappa} - \kappa) \mid x_0]$.

Corollary 15 *Under Lévy process (5.9) with an unknown mean, a Gaussian error term ($\gamma_1 = 0$ and $\gamma_2 = 0$), and a fixed x_0, the approximation to the bias of $\widehat{\kappa}$ is,*

$$E[\widehat{\kappa} - \kappa \mid x_0] = \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5) + \frac{1}{2Tn}\left(23 + 10e^{\kappa h} - \frac{2\kappa e^{2\kappa h}(\mu - x_0)^2}{\sigma^2}\right) + o(T^{-1}) \tag{5.22}$$

Corollary 16 *Under the Lévy process (5.9) with an unknown mean, a Gaussian error term ($\gamma_1 = 0$ and $\gamma_2 = 0$), and a random Gaussian x_0 with mean μ and varinace $\sigma^2/(2\kappa)$, the approximation to the Bias of $\widehat{\kappa}$ is,*

$$E(\widehat{\kappa} - \kappa) = \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5) + \frac{1}{2Tn}(23 + 10e^{\kappa h} - e^{2\kappa h}) + o(T^{-1}) \tag{5.23}$$

Remark 2.2.1 Here we consider the bias of the AR(1) coefficient up to $O(n^{-1})$, and the MSE of the AR(1) coefficient up to $O(n^{-2})$ to obtain the results in Theorem 13 and Corollary 14 for the non-Gaussian case, and Corollary 15 and 16 for the Gaussian case. In Theorem 13 and Corollary 15 the results on $Bias(\widehat{\kappa})$ are obtained conditional on x_0. In Corollary 14 and Corollary 16 the results on $Bias(\widehat{\kappa})$ are obtained for a random x_0. Tang and Chen (2009) derived the result for the bias of $\widehat{\kappa}$ for the case of normality and $x_0 \sim N(\mu, \sigma^2/(2\kappa))$ as,

$$E(\widehat{\kappa}) - \kappa = \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5) \tag{5.24}$$

which is the first term of (5.23). Therefore, our results in Theorem 13 under Lévy-based OU process with an unknown mean, provides an improvement over that of Tang and Chen (2009). In addition, we derive the results for ther Lévy-based OU process. Yu (2009) also gave the bias of $\widehat{\kappa}$ for fixed x_0 case as

$$E(\widehat{\kappa} - \kappa) = \frac{e^{2\kappa h} + 3}{2T}. \tag{5.25}$$

Remark 2.2.2 We note that the second term $\frac{1}{2Tn}\left(23 + 10e^{\kappa h} - \frac{2\kappa e^{2\kappa h}(\mu - x_0)^2}{\sigma^2}\right)$ in (5.22) incorporates both μ and x_0. If x_0 is fixed and $\mu > x_0 (< x_0)$, then $\partial Bias(\widehat{\kappa})/\partial\mu < 0 (> 0)$, which implies that the higher μ lowers the bias. Also note that the bias is not a monotonic function of x_0; if $\mu > x_0$, $\partial Bias(\widehat{\kappa})/\partial x_0 > 0$; otherwise, $\partial Bias(\widehat{\kappa})/\partial x_0 < 0$. Furthermore, when Tn is very large, the effects of μ and x_0 on the bias are negligible. However, when $x_0 = \mu$, the bias term is free from μ and σ.

Remark 2.2.3 Result (5.20) shows that not only μ and x_0 but also the skewness and the excess kurtosis affect the bias. We note that $\partial Bias(\widehat{\kappa})/\partial\gamma_2 < 0$, which imply the bias is the monotonically decreasing function of the excess kurtosis. If $\gamma_1 > 0$, $\partial Bias(\widehat{\kappa})/\partial\gamma_1 < 0$; if $\gamma_1 < 0$, $\partial Bias(\widehat{\kappa})/\partial\gamma_1 > 0$.

5.3 Bias Approximations with Higher Order Bias and MSE

This section shows the bias approximation by considering both the bias and the MSE of the AR(1) coefficient up to $O(n^{-2})$. Bao (2007) gave the approximate bias and the MSE of the OLS estimator for the AR(1) model without intercept and with a general error term as,

$$\begin{aligned} Bias(\hat{\phi}) &= -\frac{2\phi}{n} + \frac{1}{n^2}\Bigg[4\phi + \frac{2\phi x_0^2}{\sigma_0^2} - \frac{(1+\phi)x_0\gamma_1}{\sigma_0} + \frac{2\gamma_1(1+\phi+3\phi^2)(1-\phi^2)}{1-\phi^3} \\ &\quad + 2\gamma_2\phi\Bigg] + o(n^{-2}) \\ MSE(\hat{\phi}) &= \frac{1-\phi^2}{n} + \frac{1}{n^2}\Bigg[14\phi^2 - 1 - \frac{(1-\phi^2)x_0^2}{\sigma_0^2} - \frac{4\gamma_1\phi(1-\phi^2)^2}{1-\phi^3} \\ &\quad - \gamma_2(1-\phi^2)\Bigg] + o(n^{-2}), \end{aligned}$$

and those for the AR(1) model with intercept and with a general error term as,

$$
\begin{aligned}
Bias(\hat{\phi}) &= -\frac{1+3\phi}{n} + \frac{1}{n^2}\left[\frac{3\phi - 9\phi^2 - 1}{1-\phi} + \frac{1+3\phi}{(1-\phi)^2}\left(\frac{\beta - (1-\phi)x_0}{\sigma_0}\right)^2\right. \\
&\quad \left. + \frac{4\gamma_1\phi^2(1-\phi^2)}{1-\phi^3} + 2\gamma_2\phi\right] \\
&\quad + o(n^{-2}), \\
MSE(\hat{\phi}) &= \frac{1-\phi^2}{n} + \frac{1}{n^2}\left[23\phi^2 + 10\phi - \frac{1+\phi}{1-\phi}\left(\frac{\beta - (1-\phi)x_0}{\sigma_0}\right)^2\right. \\
&\quad \left. - \frac{4\gamma_1\phi(1-\phi^2)^2}{1-\phi^3} - \gamma_2(1-\phi^2)\right] \\
&\quad + o(n^{-2}).
\end{aligned}
$$

Based on these results, we obtain the bias approximations of the OLS estimator of $\widehat{\kappa}$ in the context of the Lévy OU process with a known mean and with an unknown mean, which are presented below.

Theorem 17 *Under Model (5.14) with a known mean, a non-Gaussian error term with moments given in (5.13), and a fixed x_0, the approximation to the bias of $\widehat{\kappa}$ is given by*

$$
\begin{aligned}
E[\widehat{\kappa} - \kappa \quad | \quad x_0] &= \frac{e^{2\kappa h} + 3}{2T} + \frac{1}{2nT}\{6 - e^{2\kappa h} - \frac{2\kappa e^{2\kappa h} x_0^2 (e^{2\kappa h} + 3)}{\sigma^2 (e^{2\kappa h} - 1)} + \frac{2(e^{\kappa h} + 1)x_0\gamma_1}{\sigma\sqrt{\frac{1-e^{-2\kappa h}}{2k}}} \\
&\quad - \frac{4\gamma_1(3 + 2e^{\kappa h} + 3e^{-\kappa h} + 2e^{-2\kappa h})}{e^{-2\kappa h} + e^{-\kappa h} + 1} - \gamma_2(e^{2\kappa h} + 3)\} + o(T^{-1}).
\end{aligned} \tag{5.26}
$$

Corollary 18 *Under the Lévy process (5.14) with a known mean, a non-Gaussian error term with moments given in (5.13), and a random nonnormal x_0 with mean 0 and varinace $\sigma^2/(2\kappa)$, the approximation to the bias of $\widehat{\kappa}$ is,*

$$
\begin{aligned}
E(\widehat{\kappa} - \kappa) &= \frac{e^{2\kappa h} + 3}{2T} + \frac{1}{2nT}\{6 - e^{2\kappa h} - \frac{e^{2\kappa h}(e^{2\kappa h} + 3)}{(e^{2\kappa h} - 1)} \\
&\quad - \frac{4\gamma_1(3 + 2e^{\kappa h} + 3e^{-\kappa h} + 2e^{-2\kappa h})}{e^{-2\kappa h} + e^{-\kappa h} + 1} - \gamma_2(e^{2\kappa h} + 3)\} + o(T^{-1}).
\end{aligned} \tag{5.27}
$$

Theorem 19 *Under the Lévy process (5.9) with an unknown mean, a non-Gaussian error term with moments given in (5.13), and the initial condition x_0, the approximation* *to the bias of $\widehat{\kappa}$ is*

$$
\begin{aligned}
E[\widehat{\kappa}-\kappa \quad | \quad x_0] = & \frac{1}{2T}(e^{2\kappa h}+2e^{\kappa h}+5)+\frac{1}{2Tn}\{23+\frac{12e^{2\kappa h}-16e^{\kappa h}+18}{e^{\kappa h}(e^{\kappa h}-1)} \\
& -\frac{2\kappa(\mu-x_0)^2(1+2e^{-\kappa h}+5e^{-2\kappa h})}{\sigma^2 e^{-2\kappa h}(1-e^{-2\kappa h})} \\
& -\frac{4\gamma_1\left[(1-e^{-\kappa h})(1+e^{-\kappa h})^2+2e^{-2\kappa h}(1+e^{-\kappa h})\right]}{e^{-\kappa h}(1+e^{-\kappa h}+e^{-2\kappa h})} \\
& -\gamma_2(e^{2\kappa h}+3)\}+o(T^{-1}) \qquad (5.28)
\end{aligned}
$$

Corollary 20 *Under the Lévy process (5.9) with an unknown mean, a non-Gaussian error term with moments given in (5.13), and a random initial condition whose mean is μ and varinace $\sigma^2/(2\kappa)$, the approximation to the bias of $\widehat{\kappa}$ is*

$$
\begin{aligned}
E(\widehat{\kappa}-\kappa) \quad = \quad & \frac{1}{2T}(e^{2\kappa h}+2e^{\kappa h}+5)+\frac{1}{2Tn}\{23+\frac{12e^{2\kappa h}-16e^{\kappa h}+18}{e^{\kappa h}(e^{\kappa h}-1)} \\
& -\frac{(1+2e^{-\kappa h}+5e^{-2\kappa h})}{e^{-2\kappa h}(1-e^{-2\kappa h})} \\
& -\frac{4\gamma_1\left[(1-e^{-\kappa h})(1+e^{-\kappa h})^2+2e^{-2\kappa h}(1+e^{-\kappa h})\right]}{e^{-\kappa h}(1+e^{-\kappa h}+e^{-2\kappa h})} \\
& -\gamma_2(e^{2\kappa h}+3)\}+o((T)^{-1}) \qquad (5.29)
\end{aligned}
$$

Remark 3.1 Here we consider the bias and the MSE of the AR(1) coefficient up to $O(n^{-2})$ to obtain our new results in Theorem 17 and Corollary 18 for the Lévy process with a known mean. The bias approximations for the Gaussian OU process may be straightforward developed by substituting $\gamma_1 = 0$ and $\gamma_2 = 0$ into above results. As before, the initial condition, the variance, the skewness and the excess kurtosis of of the error term all enter the higher order bias approximations. Compared with Theorem 9, the second term is different. With the use of a higher order term for the AR(1) coefficient, the esitmaiton bias approximation of $\hat{\kappa}$ has a cross product term of x_0 and γ_1. In addition, the approximated estimation bias is non-monotonical function of the initial value and the kewness. The excess kurtosis continues to have a negative effect on the estimation bias, and its negative impact is larger than the results in Theorem 9.

Remark 3.2 The second term in Collary 18 is different from that in Corollary 10. Corollary 18 shows that the higher order bias approximation continues to be a non-monotonical function of the skewness and that the kurtosis of the error term distribution has a larger negative effect on the estimation bias than that obtained in Corollary 10.

Remark 3.3 With the aid of the higher order approximation, the second terms obtained in Theorem 19 differ from those obtained in Theorem 13. The marginal effects of the long run mean, the initial condition, the skewness and the kurtosis of the error term are all different. The squared skewness and the kurtosis in Theorem 19 have a larger negative impact on the bias.

Remark 3.4 The results obtained in Corollary 20 differ from those in Corollary 14 in terms of the second term.Corollary 20 shows that the marginal impacts of the squared skewness and the kurtosis are higher than waht is implied in Corollary 14.

5.4 Bias Approximation with Higher Order Talyor Expansion

The theoretical results in the previous sections are obtained by considering talyor expansion up to $o(T^{-1})$. This section presents the bias approximation by incorporating higher order taylor expansion up to $o(T^{-2})$.By higher order taylor expansion

$$\widehat{\kappa} = \kappa - \frac{\hat{\phi}-\phi}{h\phi} + \frac{1}{2h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^2 - \frac{1}{3h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^3 + \frac{1}{4h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^4 + o_p(n^{-2}h^{-1}).$$

Therefore, to approximate the bias of $\widehat{\kappa}$, not only the bias and MSE of $\widehat{\kappa}$ are needed be considered, but also the third and fourth moment of $\widehat{\kappa}$, i.e., $E\left(\hat{\phi}-\phi\right)^3$ and $E\left(\hat{\phi}-\phi\right)^4$, should be incorporated.

For pure model, we obtain

$$\begin{aligned} E\left(\hat{\phi}-\phi\right)^3 &= n^{-2}[\beta_2^{-3}(\gamma_1^2\beta_3 - 12\phi\beta_2)] + o(n^{-2}), \\ E\left(\hat{\phi}-\phi\right)^4 &= = 3n^{-2}\left(1-\phi^2\right)^2 + o(n^{-2}), \end{aligned}$$

where $\beta_i = (1-\phi^i)^{-1}$. The following gives the conditional bias approximation and uncondition bias approximation for pure model.

Theorem 21 *Under Model (5.14) with a known mean, a non-Gaussian error term with moments given in (5.13), and a fixed x_0, the approximation to the bias of $\widehat{\kappa}$ is given by*

$$\begin{aligned} Bias(\hat{\kappa}|x_0) &= E(\hat{\kappa}-\kappa|x_0) \\ &= \frac{3+e^{2\kappa h}}{2T} + \frac{1}{nT}\{3 + \frac{1}{4}\left[3e^{4\kappa h} + 8e^{2\kappa h} - 29 + 16e^{-2\kappa h}\right] \\ &\quad - \frac{\kappa e^{2\kappa h}\left(e^{2\kappa h}+3\right)x_0^2}{\sigma^2\left(e^{2\kappa h}-1\right)} \\ &\quad + \gamma_1\frac{\left(e^{\kappa h}+1\right)x_0}{\sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}} - \gamma_2\frac{3+e^{2\kappa h}}{2e^{2\kappa h}} \\ &\quad - \gamma_1^2\frac{\left(e^{\kappa h}+1\right)\left(e^{4\kappa h}+10e^{2\kappa h}+6e^{\kappa h}+13\right)}{3\left(e^{2\kappa h}+e^{\kappa h}+1\right)}\} + o(\frac{1}{nT}) \end{aligned} \tag{5.30}$$

Corollary 22 *Under the Lévy process (5.14) with a known mean, a non-Gaussian error term with moments given in (5.13), and a random nonnormal x_0 with mean 0 and varinace $\sigma^2/(2\kappa)$, the approximation to the bias of $\widehat{\kappa}$ is,*

$$\begin{aligned} Bias(\hat{\kappa}) &= E(\hat{\kappa}-\kappa) \\ &= \frac{3+e^{2\kappa h}}{2T} + \frac{1}{nT}\{3 + \frac{1}{4}\left[3e^{4\kappa h} + 8e^{2\kappa h} - 29 + 16e^{-2\kappa h}\right] - \frac{e^{2\kappa h}\left(e^{2\kappa h}+3\right)}{2\left(e^{2\kappa h}-1\right)} \\ &\quad - \gamma_2\frac{3+e^{2\kappa h}}{2e^{2\kappa h}} \\ &\quad - \gamma_1^2\frac{\left(e^{\kappa h}+1\right)\left(e^{4\kappa h}+10e^{2\kappa h}+6e^{\kappa h}+13\right)}{3\left(e^{2\kappa h}+e^{\kappa h}+1\right)}\} + o(\frac{1}{nT}) \end{aligned}$$

Remark 4.1 Comparing Theorem 21 with the previous results obtained for the conditional bias of $\hat{\kappa}$ in the pure model, we find that the bias expression is a nonlinear function of the skewness when considering higher order Taylor expansion. As shown in Corollary 22, the unconditional bias of $\hat{\kappa}$ in the pure model is also a nonlinear function of the skewness, which is different from the previous results. Also, note that by higher order Taylor expansion, both conditional and unconditional bias is up to $o(\frac{1}{nT})$.

For intercept model, we derive that

$$\begin{aligned} E\left(\hat{\phi}-\phi\right)^3 &= n^{-2}[\beta_2^{-3}(\beta_3\gamma_1^2 - 3\beta_1\beta_2 - 12\phi\beta_2^2)] + o(n^{-2}), \\ E\left(\hat{\phi}-\phi\right)^4 &= 12n^{-2}\beta_2^{-2} + o(n^{-2}). \end{aligned}$$

The following gives the conditional bias approximation and uncondition bias approximation for intercept model.

Theorem 23 *Under the Lévy process (5.9) with an unknown mean, a non-Gaussian error term with moments given in (5.13), and the initial condition x_0, the approximation to the bias of $\widehat{\kappa}$ is*

$$\begin{aligned} Bias(\hat{\kappa}|x_0) &= E(\hat{\kappa}-\kappa|x_0) \\ &= \frac{5+2e^{\kappa h}+e^{2\kappa h}}{2T} + \frac{1}{nT}\{\left(\frac{15}{2}+7e^{\kappa h}+3e^{2\kappa h}+e^{3\kappa h}+3e^{4\kappa h}+\frac{7}{e^{\kappa h}-1}\right) \\ &\quad + \frac{e^{4\kappa h}+4e^{3\kappa h}+7e^{2\kappa h}}{(e^{\kappa h}+1)(e^{\kappa h}-1)}\frac{\kappa\left(\mu-x_0\right)^2}{\sigma^2} \\ &\quad -\gamma_1^2\frac{6\left(e^{\kappa h}+1\right)\left(e^{2\kappa h}+1\right)+\left(e^{\kappa h}-1\right)^2\left(e^{\kappa h}+1\right)^3}{3\left(e^{2\kappa h}+e^{\kappa h}+1\right)} \\ &\quad -\gamma_2\frac{3+e^{2\kappa h}}{2}\} + o(\frac{1}{nT}) \end{aligned}$$

Corollary 24 *Under the Lévy process (5.9) with an unknown mean, a non-Gaussian error term with moments given in (5.13), and a random initial condition whose mean* *is μ and varinace $\sigma^2/(2\kappa)$, the approximation to the bias of $\widehat{\kappa}$ is*

$$\begin{aligned} Bias(\hat{\kappa}) &= E(\hat{\kappa}-\kappa) \\ &= \frac{5+2e^{\kappa h}+e^{2\kappa h}}{2T} + \frac{1}{nT}\{\left(\frac{15}{2}+7e^{\kappa h}+3e^{2\kappa h}+e^{3\kappa h}+3e^{4\kappa h}+\frac{7}{e^{\kappa h}-1}\right) \\ &+\frac{e^{4\kappa h}+4e^{3\kappa h}+7e^{2\kappa h}}{2(e^{\kappa h}+1)(e^{\kappa h}-1)} \\ &-\gamma_1^2\frac{6\left(e^{\kappa h}+1\right)\left(e^{2\kappa h}+1\right)+\left(e^{\kappa h}-1\right)^2\left(e^{\kappa h}+1\right)^3}{3\left(e^{2\kappa h}+e^{\kappa h}+1\right)} \\ &-\gamma_2\frac{3+e^{2\kappa h}}{2}\}+o(\frac{1}{nT}) \end{aligned}$$

The derivations for the above theorems are outlined in the section III of Appendix A.

Remark 4.2 Comparing Theorem 23 with the previous results obtained for the conditional bias of $\hat{\kappa}$ in the intercept model, we find that the bias expression is a nonlinear function of the skewness when considering higher order Taylor expansion. As shown in Corollary 24, the unconditional bias of $\hat{\kappa}$ in the intercept model is also a nonlinear function of the skewness, which is different from the previous results. Also, note that by higher order Taylor expansion, both conditional and unconditional bias for the intercept model is up to $o(\frac{1}{nT})$.

5.5 Simulation Results

In this section, we perform Monte Carlo simulations to check the finite sample performance of our bias formulae. We also propose ways for bias correcting and check the performance of the bias corrected $\widehat{\kappa}$ in terms of mean, relative bias, and root mean squared error. We further compare the performance of the bias corrected $\widehat{\kappa}$ with those based on the bias formulae derived in Yu (2009) and Tang and Chen (2009). The Lévy processes both with a known mean and with an unknown mean are considered.

All simulation results are calculated from 10,000 replications. It is not entirely fair to compare our bias formulae with those derived in Yu (2009) and Tang and Chen (2009) because both Yu and Tang and Chen assumed the true model is the Gaussian model.

5.5.1 Bias correction for Lévy process with a known mean under non-Gaussianity

First we consider four estimators for Lévy process with a known mean under nonnormality: OLS, Yu (2009) estimator corrected by the bias (5.19) for random x_0 case and (5.25) for fixed x_0 case which are given in Remark 2.1.1, the estimator (UWY) corrected by the bias corresponding to (5.15) for fixed initial condition case and (10) for random initial condition case, the estimator (UWYH) corrected by the bias corresponding to (5.26) for fixed case and (5.27) for random case, and the estimator (UWYHT) corrected by the bias corresponding to (21) for fixed case and (5.31) for random case. In order to obtain non-Gaussian error terms we generate the random numbers from the gamma distribution with mean 1 and variance ν ($v = 0.25, 1$), then make the transformation on the generated errors to satisfy the assumption in (5.13), and then generate the discrete time observations under the model (5.14). We set κ to be small so that it is empirically realistic for the U.S. data. In particular, we consider four values for κ, $0.1, 0.5, 1.0, 3.0$. We set $T = 5$ and $h = 1/12, 1/52, 1/252$. For the fixed x_0 case, we set $x_0 = 0$. For the random x_0 case, we generate x_0 from the variance gamma distribution.

Figure 1 and 2 in figure 5.1 on page 109 plot the true bias for $h = 1/12$ and $v = 1$, the bias according to Yu (2009), UWY, UWYH, and UWYHT for the Levy processes with a known mean. The red solid line represents the true bias, the black dashed line is the Yu's bias, the blue dashed line is the UWY corrected estimator, the green line with star shows the UWYH corrected estimator, and the light blue with circle is the

UWYHT corrected estimator. We notice that UWYH and UWYHT behave almost the same. Those two lines are almost overlaped. For the random x_0 case, Figure 1 shows that when κ is smaller than 0.5 the Yu, UWYH and UWYHT estimators drop below the true bias. When κ is greater than 0.5 both Yu and UWYH bias approximations can match the true bias very well. However, the blue dashed line shows that UWY bias approximation is a little above the true bias. As κ gets larger, all three bias approximations, Yu, UWYH, and UWYHT, are approaching the true bias more closely. For the fixed x_0 case, Figure 2 shows that all bias approximations have small discrepency from the true bias, especially when κ is less than 0.5. UWYH and UWYH bias approximations are closest to the true bias compared to Yu and UWY.

Simulations results are also reported in Tables 5.1- 5.4 on pages 111- 114 and the results can be summarized as follows. First, Yu's method has the smallest bias among all estimators when $\kappa = 0.1, 0.5, 1.0$ when $v = 0.25$ and x_0 is fixed. Second, when κ is moderately larger ($\kappa = 3.0$), UWYH and UWYHT have smaller relative bias than Yu's estimator. Third, RMSEs are very close among Yu, UWY, UWYH, and UWYHT estimators. Fourth, when $v = 1$ and x_0 is fixed, UWYH and UWYHT performs slightly better than Yu estimator in terms of relative bias for $\kappa = 0.1, 0.5, 1.0, 3.0$. For random x_0 case shown in Table 5.3 and Table 5.4, when $\kappa = 0.1$, Yu's method performs slightly better than UWYH in the sense of having a lower bias and RMSE. When κ is moderately larger ($\kappa = 0.5, 1.0, 3.0$) and $v = 0.25$, UWYH and UWYHT have lower relative bias than Yu. The findings in Tables 5.1- 5.4 are consistent with the plots in Figure 1 and Figure 2.

5.5.2 Bias correction for Lévy process with an unknown mean under non-Gaussianity

In this case, we also consider five estimators under the Lévy process with an unknown mean and a non-Gaussian error term: OLS, Tang and Chen (2009) estimator (TC) corrected by the bias (5.24) given in Remark 2.2.1, the estimator UWY corrected by the bias corresponding to (5.20) for fixed initial condition case and (5.21) for random initial condition case, UWYH corrected by the bias expression in (5.28) for fixed case and (5.29) for random case, and UWYH corrected by the bias expression in (5.31) for fixed case and (5.31) for random case. As before, the error term is first generated from the gamma distribution with mean 1 and variance ν ($\nu = 0.25, 1$). We set $\mu = 0.1$ and $\sigma^2 = 0.1$. For the case of the fixed initial condition, x_0 is fixed at μ. For the case of the random initial condition, x_0 is generated from the gamma distribution with mean 1 and variance ν ($\nu = 0.25, 1$).

Figure 3 and 4 in figure 5.2 plot the true bias, the biases according to Tang and Chen (2009), UWY, UWYH, and UWYHT for the Levy processes with an unknown mean. The red solid line represents the true bias, the black dashed line with dots is the TC bias according to Tang and Chen (2009), the blue dashed line is UWY, the green line with stars shows the UWYH bias expression, and the blue line with circle is UWYHT. For random initial condition case, among all bias approximations shown in Figure 3, UWYHT performs the best and shows the curvature as κ is getting smaller. Figure 4 shows the performance of all bias approximations for the Lévy process with an unknown mean and the fixed x_0 case. When κ is close to zero, UWYH and UWYHT bias approximation goes up dramatically. When κ is greater than 1.0, UWYH and UWYHT are very close to the true bias.

Tables 5.5- 5.8 report the simulation results. As for the case with a known mean, the simulation results under the Lévy process with an unknown mean provides the evidence that UWYH is useful in finite samples. The simulations in Table 5.5- 5.8 show that UWYH always has the smallest bias and the lowest RMSE for most of cases, especially, for larger $\kappa = 1.0, 3.0$. For $\kappa = 0.5$, UWY has the best performance in terms of relative bias and RMSE. These results are consistent with those in Figures 3 and 4, namely, our estimators (UWY, UWYH, and UWYHT) offer improvement over OLS and TC, especially when $\kappa = 0.5, 1.0, 3.0$. UWYH is the most efficient estimator in the sense of having the smallest bias and the lowest RMSE. UWYHT performs closely to UWYH.

The main findings the two experiements are as follow. First, for the Lévy process with a unknown mean, no matter if x_0 is fixed or random, under non-Gaussanity the RMSE of UWY is always smaller than that of the TC estimator and UWYH has the smallest bias and the smallest RMSE when $\kappa = 1.0, 3.0$. Second, for the Lévy process with a known mean, Yu, UWYH and UWYHT perform similarly. When $\kappa = 0.1$, Yu has a slightly smaller bias and RMSE than UWYH. However, when $\kappa = 0.5, 1.0, 3.0$, UWYH performs slightly better than Yu in the sense of having a slightly lower relative bias and RMSE. Second, Figures 1-4 show that the UWYH and UWYHT bias approximations have large distance from the true bias as κ is very close to 0. However, as κ gets larger, the UWYH and UWYHT bias approximations get closer to the true bias. Finally, all the simulation results in this seciton point outs that if the true model us non-Gaussian, it is important to take into account of the feature for the sake of bias correction and the higher order bias approximation is useful to improve the efficiency and the accuracy of $\widehat{\kappa}$ in finite samples.

5.6 Conclusions

This chapter considers the effect of the non-Gaussianity of error terms under the Lévy processes with a known mean and with an unknown mean. We obtain the bias approximations of the mean reversion parameter estimator under a general error distribution and find that the skewness (γ_1), the kurtorsis (γ_2), the initial condition, the long term mean (μ), and the diffusion parameter (σ^2) all affect the bias of κ. Monte Carlo simulations provide supports that our proposed bias corrected estimator of the mean reversion parameter is effective in finite samples.

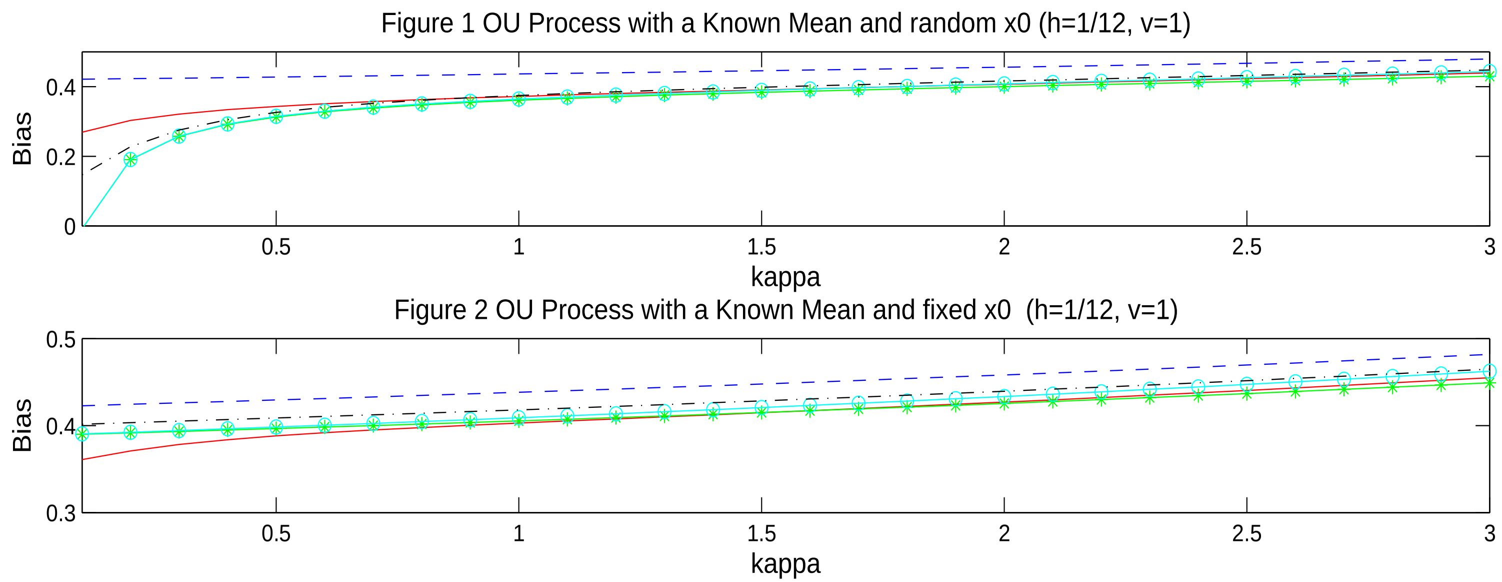

Figure 5.1: Lévy process with a known mean under non-Gaussianity

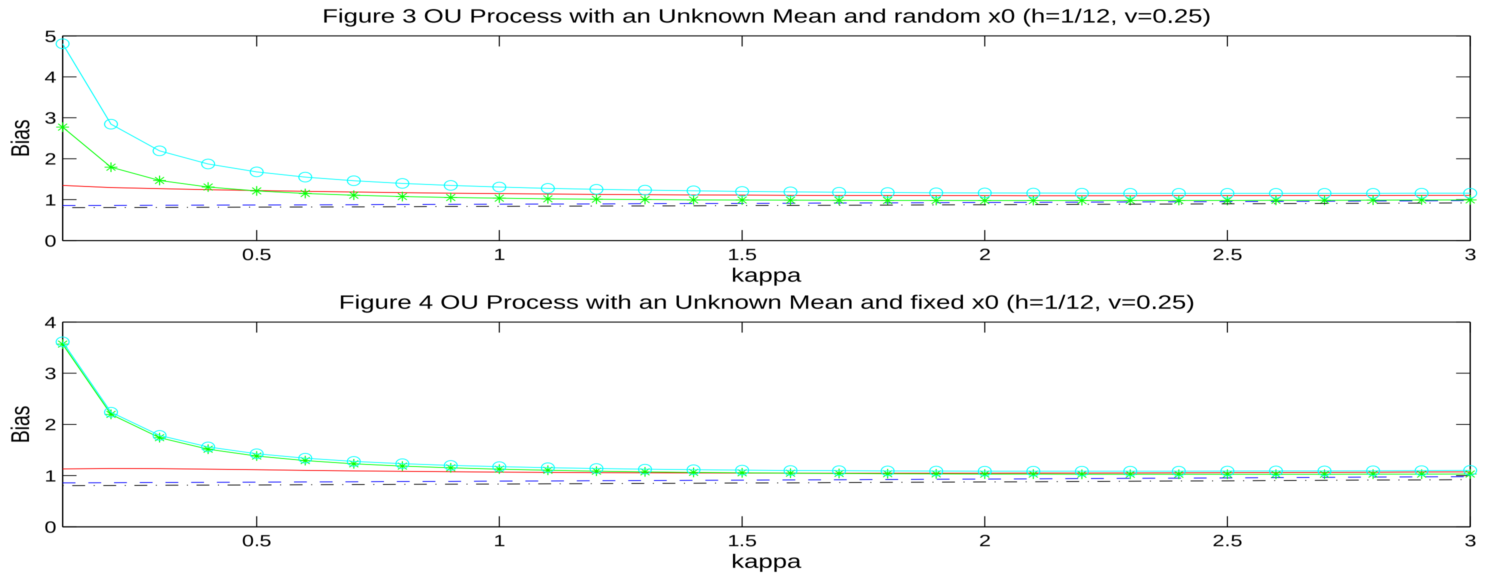

Figure 5.2: Lévy process with a known mean under non-Gaussianity

Table 5.1: Bias correction for the Lévy process with a known mean and a fixed x_0 ($v = 0.25$)

$v = 0.25$	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.1					T=5, h=1/52, κ=0.1					T=5, h=1/252, κ=0.1				
Mean	0.4624	0.0607	0.0391	0.0574	0.0573	0.4687	0.0683	0.0633	0.0676	0.0676	0.4683	0.0682	0.0672	0.0681	0.0681
r. bias(%)	362.4	-39.28	-60.90	-42.56	-42.65	368.71	-31.68	-36.68	-32.44	-32.45	368.29	-31.79	-32.82	-31.94	-31.94
RMSE	0.7681	0.6784	0.6800	0.6786	0.6786	0.7663	0.6725	0.6728	0.6726	0.6726	0.7666	0.6730	0.6731	0.6730	0.6730
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	0.8911	0.4824	0.4610	0.4794	0.4788	0.8905	0.4886	0.4836	0.4878	0.4878	0.8901	0.4897	0.4887	0.4895	0.4895
r. bias(%)	78.23	-3.51	-7.80	-4.13	-4.23	78.11	-2.28	-3.28	-2.43	-2.44	78.02	-2.06	-2.27	-2.09	-2.09
RMSE	0.8876	0.7970	0.7978	0.7971	0.7971	0.8721	0.7799	0.7800	0.7799	0.7799	0.8639	0.7708	0.7709	0.7708	0.7708
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	1.4089	0.9907	0.9696	0.9879	0.9867	1.3986	0.9947	0.9897	0.9940	0.9939	1.3975	0.9967	0.9957	0.9966	0.9966
r. bias(%)	40.89	-0.93	-3.04	-1.64	-1.3316	39.86	-0.53	-1.03	-0.60	-0.61	39.75	-0.33	-0.43	-0.34	-0.34
RMSE	1.0334	0.9491	0.9496	0.9492	0.9492	0.9896	0.9058	0.9058	0.9058	0.9058	0.9749	0.8901	0.8901	0.8901	0.8901
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	3.4687	3.0039	2.9841	3.0024	2.9957	3.4141	3.0019	2.9969	3.0012	3.0010	3.4020	2.9996	2.9985	2.9995	2.9995
r. bias(%)	15.62	0.13	-0.53	-0.08	-0.14	13.80	0.06	-0.10	0.04	0.03	13.40	-0.01	-0.05	-0.02	-0.02
RMSE	1.5831	1.5122	1.5122	1.5122	1.5122	1.3808	1.3172	1.3172	1.3172	1.3172	1.3377	1.2759	1.2759	1.2759	1.2759

Table 5.2: Bias correction for the Lévy process with a known mean and a fixed x_0 ($v = 1.0$)

$v = 1.0$	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.1					T=5, h=1/52, κ=0.1					T=5, h=1/252, κ=0.1				
Mean	0.4609	0.0592	0.0376	0.0710	0.0706	0.4614	0.0610	0.0560	0.0637	0.0637	0.4712	0.0711	0.0701	0.0717	0.0717
r. bias(%)	360.88	-40.80	-62.35	-29.02	-29.36	361.38	-39.00	-44.00	-36.30	-36.32	371.20	-28.88	-29.91	-28.33	-28.33
RMSE	0.7732	0.6850	0.6866	0.6844	0.6844	0.7486	0.6567	0.6570	0.6566	0.6566	0.7686	0.6736	0.6737	0.6736	0.6736
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	0.8882	0.4795	0.4585	0.4918	0.4901	0.8822	0.4803	0.4753	0.4830	0.4829	0.8879	0.4875	0.4864	0.4880	0.4880
r. bias(%)	77.65	-4.09	-8.31	-1.64	-1.99	76.44	-3.95	-4.94	-3.40	-3.42	77.58	-2.50	-2.71	-2.39	-2.39
RMSE	0.8950	0.8066	0.8075	0.8064	0.8065	0.8542	0.7641	0.7643	0.7641	0.7641	0.8594	0.7670	0.7670	0.7669	0.7669
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	1.4029	0.9848	0.9643	0.9976	0.9940	1.3886	0.9848	0.9798	0.9875	0.9873	1.3899	0.9891	0.9881	0.9896	0.9896
r. bias(%)	40.29	-1.52	-3.57	-0.24	-0.60	38.87	-1.52	-2.02	-1.2490	-1.2671	38.99	-1.09	-1.19	-1.04	-1.04
RMSE	1.0369	0.9555	0.9560	0.9554	0.9554	0.9732	0.8924	0.8925	0.8923	0.8923	0.9642	0.8819	0.8819	0.8819	0.8819
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	3.4551	2.9902	2.9729	3.0062	2.9926	3.3969	2.9847	2.9799	2.9876	2.9870	3.3829	2.9804	2.9794	2.9810	2.9810
r. bias(%)	15.17	-0.33	-0.90	0.21	-0.25	13.23	-0.51	-0.67	-0.41	-0.43	12.76	-0.65	-0.69	-0.63	-0.63
RMSE	1.5771	1.5100	1.5102	1.5100	1.5100	1.3638	1.3048	1.3049	1.3048	1.3048	1.3192	1.2626	1.2626	1.2626	1.2626

Table 5.3: Bias correction for the Lévy process with a known mean and a random x_0 ($v = 0.25$)

$v = 0.25$	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.1					T=5, h=1/52, κ=0.1					T=5, h=1/252, κ=0.1				
Mean	0.3521	0.2053	-0.0695	0.3521	0.3521	0.3509	0.2039	-0.0541	0.3509	0.3509	0.3557	0.2086	-0.0453	0.3557	0.3557
r. bias(%)	252.07	105.35	-169.54	252.14	252.05	250.91	103.86	-154.08	250.92	250.91	255.71	108.58	-145.32	255.71	255.71
RMSE	0.5891	0.5427	0.5587	0.5891	0.5891	0.5701	0.5223	0.5346	0.5701	0.5701	0.5837	0.5359	0.5445	0.5837	0.5837
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	0.8448	0.5190	0.4165	0.5183	0.5177	0.8439	0.5221	0.4373	0.5223	0.5223	0.8444	0.5237	0.4431	0.5242	0.5242
r. bias(%)	68.97	3.79	-16.69	3.65	3.55	68.77	4.43	-12.54	4.46	4.46	68.89	4.74	-11.37	4.83	4.83
RMSE	0.8018	0.7241	0.7287	0.7241	0.7241	0.7882	0.7096	0.7120	0.7096	0.7096	0.7919	0.7134	0.7153	0.7134	0.7134
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	1.3778	1.0031	0.9405	1.0023	1.0010	1.3708	1.0077	0.9623	1.0073	1.0073	1.3684	1.0078	0.9667	1.0077	1.0077
r. bias(%)	37.78	0.31	-5.95	0.23	0.10	37.08	0.77	-3.77	0.73	0.73	36.84	0.78	-3.33	0.77	0.77
RMSE	0.9684	0.8916	0.8936	0.8916	0.8916	0.9386	0.8623	0.8631	0.8623	0.8623	0.9289	0.8527	0.8533	0.8527	0.8527
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	3.4525	3.0046	2.9706	3.0059	2.9991	3.4023	3.0042	2.9856	3.0039	3.0038	3.3897	3.0007	2.9863	3.0007	3.0007
r. bias(%)	15.08	0.15	-0.98	0.20	-0.03	13.41	0.14	-0.48	0.13	0.13	12.99	0.03	-0.46	0.02	0.02
RMSE	1.5355	1.4673	1.4676	1.4673	1.4673	1.3563	1.2952	1.2953	1.2952	1.0952	1.3175	1.2586	1.2587	1.2586	1.2586

Table 5.4: Bias correction for the Lévy process with a known mean and a random x_0 ($v = 1.0$)

$v = 1.0$	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT	OLS	Yu	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.1					T=5, h=1/52, κ=0.1					T=5, h=1/252, κ=0.1				
Mean	0.3691	0.2224	-0.0524	0.3843	0.3839	0.3682	0.2212	-0.0368	0.3717	0.3717	0.3683	0.2212	-0.0327	0.3690	0.3690
r. bias(%)	269.13	122.41	-152.42	282.26	283.92	268.24	121.19	-136.75	271.71	271.69	268.31	121.18	-132.72	269.02	269.02
RMSE	0.6364	0.5895	0.5965	0.6429	0.6428	0.5968	0.5467	0.5503	0.5983	0.5983	0.5990	0.5491	0.5517	0.5993	0.5993
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	0.8432	0.5173	0.4153	0.5319	0.5302	0.8430	0.5213	0.4365	0.5250	0.5249	0.8432	0.5224	0.4418	0.5236	0.5236
r. bias(%)	68.63	3.46	-16.95	6.38	6.03	68.60	4.26	-12.70	4.99	4.97	68.64	4.48	-11.61	4.72	4.72
RMSE	0.8192	0.7440	0.7486	0.7445	0.7444	0.7887	0.7106	0.7131	0.7107	0.7107	0.7798	0.7006	0.7026	0.7006	0.7006
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	1.3720	0.9973	0.9354	1.0122	1.0085	1.3635	1.0004	0.9550	1.0035	1.0033	1.3613	1.0006	0.9595	1.0013	1.0013
r. bias(%)	37.20	-0.27	-6.46	1.22	0.8494	36.35	0.04	-4.50	0.35	0.33	36.13	0.06	-4.05	0.13	0.13
RMSE	0.9818	0.9086	0.9109	0.9087	0.9086	0.9307	0.8568	0.8580	0.8568	0.8568	0.9121	0.8375	0.8385	0.8375	0.8375
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	3.4393	2.9914	2.9598	3.0101	2.9965	3.3859	2.9878	2.9693	2.9911	2.9905	3.3702	2.9813	2.9668	2.9819	2.9819
r. bias(%)	14.64	-0.29	-1.33	0.34	-0.12	12.86	-0.41	-1.02	-0.30	-0.32	12.34	-0.62	-1.11	-0.60	-0.60
RMSE	1.5374	1.4733	1.4738	1.4733	1.4733	1.3406	1.2839	1.2842	1.2838	1.2838	1.2959	1.2420	1.2423	1.2420	1.2420

Table 5.5: Bias correction for the Lévy process with an unknown mean and a fixed x_0 ($v = 0.25$)

$v = 0.25$	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	1.6142	0.7970	0.7414	0.22413	0.17793	1.5787	0.7748	0.7621	0.2120	0.2019	1.5634	0.7626	0.7599	0.2020	0.1999
r. bias(%)	222.84	59.40	48.28	-55.17	-64.42	215.74	54.96	52.42	-57.59	-59.62	212.67	52.51	51.99	-59.60	-60.01
RMSE	1.5723	1.1484	1.1353	1.1431	1.1551	1.5096	1.0912	1.0881	1.0946	1.0973	1.4737	1.0535	1.0529	1.0629	1.0635
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	2.0680	1.2324	1.1763	0.9385	0.8888	2.0101	1.2023	1.1895	0.9195	0.9092	1.9922	1.1905	1.1879	0.9100	0.9079
r. bias(%)	106.80	23.25	17.62	-6.16	-11.12	101.01	20.23	18.95	-8.0524	-9.0817	99.21	19.05	18.79	-9.00	-9.21
RMSE	1.6297	1.2527	1.2435	1.2325	1.2360	1.5291	1.1657	1.1636	1.1508	1.1516	1.4922	1.1305	1.1301	1.1180	1.1181
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	4.0732	3.1515	3.0926	3.0402	2.9695	3.9231	3.0990	3.0861	3.0459	2.9917	3.8874	3.0825	3.0799	2.9886	2.9865
r. bias(%)	35.77	5.05	3.09	1.34	-1.02	30.77	3.30	2.87	0.09	-0.28	29.58	2.75	2.66	-0.38	-0.45
RMSE	2.0620	1.7672	1.7631	1.7612	1.7610	1.7401	1.4784	1.4775	1.4750	1.4751	1.6729	1.4205	1.4204	1.4182	1.4182

Table 5.6: Bias correction for the Lévy process with an unknown mean and a fixed x_0 ($v = 1.0$)

$v = 1.0$	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	1.6122	0.7950	0.7397	0.2375	0.1913	1.5645	0.7606	0.7479	0.2014	0.1912	1.5525	0.7517	0.7491	0.1918	0.1898
r. bias	222.44	59.00	47.95	-52.50	-61.75	212.90	52.13	49.59	-59.73	-61.76	210.49	50.33	49.81	-61.63	-62.05
RMSE	1.5841	1.1659	1.1532	1.1581	1.1695	1.4849	1.0675	1.0645	1.0774	1.0803	1.4535	1.0336	1.0329	1.0487	1.0494
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	2.0643	1.2288	1.1732	0.9505	0.9008	1.9935	1.1857	1.1731	0.9065	0.8962	1.9782	1.1764	1.1738	0.8966	0.8945
r. bias	106.43	22.88	17.32	-4.95	-9.92	99.36	18.58	17.31	-9.35	-10.38	97.80	17.64	17.38	-10.34	-10.55
RMSE	1.6502	1.2817	1.2729	1.2621	1.2650	1.5089	1.1506	1.1486	1.1394	1.1403	1.4652	1.1047	1.1043	1.0954	1.0956
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	4.0596	3.1379	3.0814	3.0440	2.9733	3.9011	3.0769	3.0642	2.9841	2.9731	3.8666	3.0617	3.0590	2.9685	2.9664
r. bias	35.32	4.60	2.71	1.47	-0.8890	30.04	2.56	2.14	-0.530	-0.90	28.88	2.06	1.97	-1.05	-1.12
RMSE	2.0689	1.7822	1.7787	1.7774	1.7771	1.7246	1.4725	1.4718	1.4705	1.4707	1.6465	1.4011	1.40101	1.4001	1.4002

Table 5.7: Bias correction for the Lévy process with an unknown mean and a random x_0 ($v = 0.25$)

$v = 0.25$	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	1.5248	0.7076	0.6538	0.3051	-0.1670	1.4976	0.6937	0.6814	0.2933	-0.1227	1.4895	0.6887	0.6861	0.2886	-0.1147
r. bias(%)	204.96	41.52	30.76	-38.98	-133.4	199.52	38.75	36.28	-41.34	-124.53	197.89	37.74	37.23	-42.28	-122.93
RMSE	1.4714	1.0761	1.0670	1.0737	1.2489	1.4251	1.0359	1.0337	1.0384	1.1930	1.4070	1.0179	1.0174	1.0223	1.1740
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	1.9962	1.1607	1.1065	0.9574	0.6810	1.9486	1.1408	1.1284	0.9403	0.7242	1.9353	1.1337	1.1312	0.9337	0.7304
r. bias(%)	99.62	16.07	10.65	-4.25	-31.90	94.86	14.08	12.84	-5.97	-27.58	93.53	13.37	13.012	-6.63	-26.96
RMSE	1.5381	1.1829	1.1768	1.1727	1.2146	1.4573	1.1152	1.1137	1.1079	1.1402	1.4301	1.0914	1.0911	1.0852	1.1162
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	4.0342	3.1125	3.0564	3.0403	2.8721	3.8958	3.0717	3.0592	3.0044	2.9207	3.8630	3.0582	3.0557	2.9915	2.9215
r. bias(%)	34.47	3.75	1.88	1.34	-4.26	29.86	2.39	1.97	0.15	-2.64	28.77	1.94	1.86	-0.2847	-2.62
RMSE	1.9927	1.7070	1.7042	1.7038	1.7081	1.7023	1.4493	1.4488	1.4476	1.4497	1.6436	1.4002	1.4000	1.3990	1.4011

Table 5.8: Bias correction for the Lévy process with an unknown mean and a random x_0 ($v = 1.0$)

$v = 1.0$	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT	OLS	TC	UWY	UWYH	UWYHT
	T=5, h=1/12, κ=0.5					T=5, h=1/52, κ=0.5					T=5, h=1/252, κ=0.5				
Mean	1.5282	0.7110	0.6575	0.3238	-0.1483	1.4863	0.6825	0.6701	0.2855	-0.1305	1.4846	0.6837	0.6812	0.2844	-0.1189
r. bias(%)	205.64	42.20	31.51	-35.24	-129.65	197.27	36.49	34.03	-42.90	-126.09	196.91	36.75	36.24	-43.12	-123.77
RMSE	1.5010	1.1137	1.1049	1.1077	1.2713	1.4040	1.0157	1.0135	1.0219	1.1814	1.3873	0.9945	0.9940	1.0009	1.1568
	T=5, h=1/12, κ=1.0					T=5, h=1/52, κ=1.0					T=5, h=1/252, κ=1.0				
Mean	1.9953	1.1597	1.1062	0.9721	0.6957	1.9340	1.1262	1.1139	0.9293	0.7131	1.9249	1.1233	1.1208	0.9240	0.7207
r. bias(%)	99.53	15.98	10.62	-2.79	-30.43	93.41	12.62	11.39	-7.07	-28.69	92.49	12.33	12.08	-7.60	-27.93
RMSE	1.5668	1.2206	1.2147	1.2104	1.2477	1.4394	1.1025	1.1011	1.0975	1.1322	1.4068	1.0671	1.0668	1.0627	1.0961
	T=5, h=1/12, κ=3.0					T=5, h=1/52, κ=3.0					T=5, h=1/252, κ=3.0				
Mean	4.0220	3.1003	3.0469	3.0455	2.8773	3.8740	3.0499	3.0375	2.9861	2.9025	3.8424	3.0376	3.0351	2.9716	2.9016
r. bias(%)	34.07	3.34	1.55	1.52	-4.09	29.13	1.66	1.25	-0.46	-3.25	28.08	1.25	1.17	-0.94	-3.28
RMSE	2.0056	1.7286	1.7263	1.7263	1.7301	1.6859	1.4425	1.4421	1.4417	1.4449	1.6155	1.3790	1.3789	1.3788	1.3820

Chapter 6

Exact Distribution and Density of Mean Reversion Parameter Estimator in Continuous Time Models *

6.1 Introduction

Since the seminal works of Merton (1971) and Black and Scholes (1973), continuous-time models have been used extensively in financial economics, see the excellent survey by Sundaresan (2000). Econometricians have also paid close attention to this line of literature. Maximum likelihood, generalized method of moments, simulated method of moments, and nonparametric approaches have been developed for model estimation, see, for instance, Singleton (2001), Aït-Sahalia (2002), Bandi and Phillips

*This chapter is a joint work with Dr. Yong Bao and Dr. Aman Ullah

(2003), Hong and Li (2005), and Phillips and Yu (2009a). As shown in the literature, there exists serious estimation bias in the mean reversion parameter (κ) by almost all the methods, especially when the diffusion process has a linear drift function and the speed of mean reversion is slow (i.e., small values of κ).[1] For example, Phillips and Yu (2005) revealed that the bias of the maximum likelihood estimator (MLE) for κ in the CIR model (Cox, Ingersoll, and Ross, 1985) can be extremely large for data sets with very long time spans, regardless of data frequency. Recently, Tang and Chen (2009) showed that the bias of $\widehat{\kappa}$ is up to $O(T^{-1})$ in the stationary Vasicek model, where T is the time span. They also derived the approximate biases of the diffusion and drift estimators, and their simulations demonstrated that the estimation biases of diffusion and drift parameters are virtually zero, but $\widehat{\kappa}$ could be substantially biased. Since the mean reversion parameter κ is of most importance for asset pricing, risk management, and forecasting, considerable attention in the literature has arisen to improve its estimation accuracy. Recent contributions include indirect inference (Phillips and Yu, 2009b), bootstrapping (Tang and Chen, 2009), and analytical bias approximation (Yu, 2011).

In addition to the classical asymptotic analysis under expanding domain ($T \to \infty$), asymptotic results under infill ($n \to \infty$, where n is the number of sample observations within a data span T) and mixed ($n \to \infty$ and $T \to \infty$) domains are also analyzed in the literature. In the context of Vasicek (1977) and CIR processes with unknown drift, Tang and Chen (2009) showed that asymptotic distributions of the MLE are quite different under expanding and mixed domains. Aït-Sahalia (2002) derived the asymptotic distribution of his approximate MLE under the expanding domain in diffusions models. A striking observation from his simulations is that under the stationary case, the asymptotic distribution of the estimated mean reversion parameter deviates more

[1] Here we use the word "bias" in a very loose term. In Section 2, we discuss this issue more formally.

seriously from its corresponding finite-sample distribution as the true parameter value deceases from 10 to 1, i.e., as the process is getting closer to a unit root process, even with a very large sample size ($n = 1000$). Under the mixed domain, Brown and Hewitt (1975) obtained the limit normal distribution for the MLE of κ in the Vasicek model with a known drift term, see also Bandi and Phillips (2003 , 2007), and Phillips and Yu (2009c) for asymptotic analysis under mixed domain. In a recent paper, Zhou and Yu (2010) derived the asymptotic distributions of the least squares (LS) estimator of κ in a general class of diffusion models under the three different domains. They provided Monte Carlo evidence that the infill asymptotic distribution is much more accurate in approximating the true finite-sample distribution than the asymptotic distributions under the other two domains.

The problems of approximate estimation bias and inaccurate and different distribution approximations floating in the literature are largely due to the absence of exact analytical distribution results. Moreover, in reality, given the discretized data (with a given finite data span T and finite sample size n), we do not really know under which asymptotic domain our inference about $\hat{\kappa}$ shall be, but the asymptotic distribution results can behave quite differently under expanding, infill, and mixed domains. To address these problems, in this chapter we investigate the exact distribution of the estimated mean reversion parameter. To the best of our knowledge, this chapter is the first to examine the exact finite-sample distribution of the estimated κ in continuous-time models. Since the MLE of κ is a simple transformation of the LS estimator of the autoregressive coefficient ϕ in a first-order autoregressive (AR(1)) model with discrete data, our study is intrinsically related to the vast literature studying the finite-sample distribution of the AR(1) coefficient estimator $\hat{\phi}$. The Imhof (1961) technique, in conjunction with Davies (1973, 1980), was typically used to develop the exact distribution of $\hat{\phi}$, see Ullah

(2004) for a comprehensive review. Nevertheless, the Imhof (1961) technique is applicably only when the process is strictly stationary with an initial random observation included in formulating $\hat{\phi}$, or when the first observation is discarded. Computational burden of the Imhof (1961) technique also increases tremendously as the sample size of the AR process increases, since it involves computation of eigenvalues of a matrix whose dimension is the same as the sample size. In this chapter, we take a different approach by first analytically evaluating the joint characteristic function of the random numerator and denominator in defining $\hat{\phi}$, and then inverting it via Gurland (1948) and Gil-Pelaez (1951) to calculate the exact finite-sample distribution. This approach is in line with Tsui and Ali (1992 , 1994) and Ali (2002) . However, note that in Tsui and Ali (1992 , 1994) and Ali (2002), no intercept term was included in the AR(1) model. This is equivalent to a known drift term in our continuous-time model. In this paper, we consider explicitly the case when the drift term is unknown. Moreover, Tsui and Ali (1992 , 1994) did not include the initial observation in formulating the LS estimator $\hat{\phi}$. However, the initial observation does matter in studying the finite-sample distributions; in fact, it also matters even for the asymptotic distributions under several scenarios. The initial observation was included in Ali (2002) , but he studied the approximate distributions.

The reminder of this chapter is as follows. In Section 2, we derive the exact distribution of the MLE of the mean reversion parameter κ. Section 3 offers some insights to the issues of moment and asymptotic distribution. Section 4 presents the simulation results and compares the exact distribution results with the asymptotic results under the three different domains. Section 5 concludes. Technical details are collected in the section IV of Appendix A.

6.2 Finite-Sample Properties

We consider the Ornstein-Uhlenbeck (OU) process with initial value $x(0)$,

$$\mathrm{d}x(t) = \kappa(\mu - x(t))\mathrm{d}t + \sigma \mathrm{d}B(t), \tag{6.1}$$

where $\kappa \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$, and $B(t)$ is a standard Brownian motion. We are interested in estimating the parameter κ. When $\kappa \neq 0$, the solution to the above process is

$$x(t) = \mu + (x(0) - \mu)\exp(-\kappa t) + \sigma \int_0^t \exp(\kappa(s-t))\,\mathrm{d}B(s),\ \ t \geq 0.$$

Usually $\kappa > 0$ is assumed, and then as $t \to \infty$, the deterministic part of x tends to the mean level μ, so we have a mean-reverting process. When $\kappa = 0$, the process is no longer mean reverting:

$$x(t) = x(0) + \sigma B(t),$$

where the parameter μ vanishes.

The exact discrete model corresponding to (6.1) is given by

$$x_{ih} = \alpha + \phi x_{(i-1)h} + \varepsilon_{ih}, \tag{6.2}$$

where $0 < \phi = \exp(-\kappa h) \leq 1$, $\alpha = \mu(1 - \exp(-\kappa h))$, $\varepsilon_{ih} = \sigma\epsilon_i\sqrt{(1 - \exp(-2\kappa h))/(2\kappa)}$ when $\kappa > 0$ and $\varepsilon_{ih} = \sigma\sqrt{h}\epsilon_i$ when $\kappa = 0$, $\epsilon_i \sim i.i.d.\mathrm{N}(0,1)$, h is the sampling interval, $i = 0, 1, \cdots, n$ such that the observed data are discretely recorded at $(0, h, 2h, \cdots, nh)$ in the time interval $[0, T]$ and $nh = T$. Thus $n + 1$ is the total number of discrete observations and T is the data span. When $\kappa > 0$, $\phi < 1$; when $\kappa = 0$, $\phi = 1$, $\alpha = 0$, so (6.2) becomes a random walk (with no drift). In the following, we suppress h in x_{ih} and ε_{ih} for notational convenience.

It is well known that the LS/ML estimator of κ is

$$\hat{\kappa} = -\frac{\ln(\hat{\phi})}{h}, \tag{6.3}$$

where $\hat{\phi}$ is the LS estimator of the autoregression coefficient ϕ from the AR(1) model (6.2), defined as

$$\hat{\phi} = \begin{cases} \frac{\sum_{i=1}^{n} x_{i-1} x_i}{\sum_{i=1}^{n} x_{i-1}^2} & \kappa = 0, \text{ or } \kappa > 0 \text{ and } \mu \text{ is known} \\ & \text{(without loss of generality, } \mu = 0) \\ \frac{\sum_{i=1}^{n} (x_{i-1} - \bar{x}) x_i}{\sum_{i=1}^{n} (x_{i-1} - \bar{x})^2} & \kappa > 0 \text{ and } \mu \text{ is unknown} \end{cases}, \tag{6.4}$$

where $\bar{x} = n^{-1} \sum_{i=1}^{n} x_{i-1}$.[2]

We are interested in studying the properties of $\hat{\kappa}$ estimated from the discrete sample via $\hat{\phi}$. As can be expected, the exact properties of $\hat{\kappa}$ depend on how to spell out the initial observation $x(0) = x_0$. We distinguish between three cases: (A) x_0 is fixed at 0; (B) x_0 is fixed at a constant c; (C) x_0 is a random draw from $\text{N}((1-\phi)^{-1}\alpha, (1-\phi^2)^{-1}\sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_i)$, or, equivalently, $\text{N}\left(\mu, (2\kappa)^{-1}\sigma^2\right)$, $\kappa > 0$, and x_0 is independent of $(\varepsilon_1, \cdots, \varepsilon_n)$. Under case C, the time series $(x_0, x_1, \cdots, x_n)$ is stationary. Since case A is a special case of B by setting $c = 0$, in the sequel, we focus on cases B and C and discuss briefly case A.

[2]There seems to be some confusion in the literature regarding the sample size n in formulating the LS estimator $\hat{\phi}$. In Tsui and Ali (1992 ,1994), the initial observation x_0 is discarded, following the convention of Hurwicz (1950) . Yet in Ali (2002) the initial observation is included, which possibly leads the author to state (wrongly) that there might be an error in Tsui and Ali (1994). Since we are interested in studying the finite sample properties of $\hat{\kappa}$, the initial condition x_0 matters and we include it in the estimation procedure.

6.2.1 Distribution and Density

We note that $\hat{\phi}$ can be negative with a non-zero probability; in fact, (6.3) is defined only if $\hat{\phi} > 0$. Thus, we define exact distribution of $\hat{\kappa} - \kappa$ as

$$\begin{aligned}
\Pr(\hat{\kappa} - \kappa \le w) &\equiv \Pr\left(\hat{\kappa} - \kappa \le w | \hat{\phi} > 0\right) \\
&= \Pr\left(\hat{\phi} \ge \phi \exp(-hw) | \hat{\phi} > 0\right) \\
&= \frac{\Pr\left(\hat{\phi} \ge \phi \exp(-hw)\right)}{\Pr\left(\hat{\phi} > 0\right)} \\
&= \frac{1 - \Pr\left(\hat{\phi} \le \phi \exp(-hw)\right)}{1 - \Pr\left(\hat{\phi} < 0\right)} \\
&= \frac{1 - \Pr\left(\hat{\phi} - \phi \le (\exp(-hw) - 1)\phi\right)}{1 - \Pr\left(\hat{\phi} - \phi < -\phi\right)} \\
&= \frac{1 - F_{\hat{\phi}}(d)}{1 - F_{\hat{\phi}}(-\phi)}, \qquad (6.5)
\end{aligned}$$

where $d = (\exp(-hw) - 1)\phi$ and $F_{\hat{\phi}}(d)$ denotes the cumulative distribution function (CDF) of $\hat{\phi} - \phi$ at d. Thus the distribution of $\hat{\kappa} - \kappa$ at w follows from the distribution of $\hat{\phi} - \phi$, given the sampling frequency h. From (6.5), we have the probability distribution function (PDF) of $\hat{\kappa} - \kappa$, conditional on $\hat{\phi} > 0$,

$$f_{\hat{\kappa}}(w) = \frac{h\phi \exp(-hw) f_{\hat{\phi}}(d)}{1 - F_{\hat{\phi}}(-\phi)}, \qquad (6.6)$$

where $f_{\hat{\phi}}(d)$ denotes the PDF of $\hat{\phi} - \phi$ at d. [3]

We note from (6.5) that evaluation of the cumulative distribution of $\hat{\kappa}$ depends on evaluation of the distribution of $\hat{\phi}$. When $\kappa > 0$ and x_0 is random, we can write $\hat{\phi} - \phi$ as a ratio of quadratic forms in the normal random vector $(x_0, x_1, \cdots, x_n)'$, and the technique of Imhof (1961) can be used to evaluate $F_{\hat{\phi}}$, and thus $F_{\hat{\kappa}}$. For fixed x_0, it is not obvious how to directly apply Imhof (1961).[4] More fundamentally, as pointed

[3] Note that (6.5) and (6.6) hold regardless of the distribution assumption.

[4] If we discard x_0, then the Imhof (1961) technique is still applicable, as we can define $\hat{\phi}$ in terms of quadratic forms in $\boldsymbol{x}_n$.

out by Tsui and Ali (1994) and Ali (2002), Imhof's procedure requires computation of eigenvalues of an $(n+1)\times(n+1)$ matrix, which becomes very cumbersome as the sampling interval h decreases.[5] Therefore, we proceed to derive the distribution of $\hat{\phi}$, and hence that of $\hat{\kappa}$, by an alternative method.

Following Tsui and Ali (1994), we use the results from Gurland (1948) and Gil-Pelaez (1951) on a ratio of two random variables. Let Y_1 and Y_2 have the joint characteristic function (CF) $\varphi(u,v)=\mathrm{E}(\exp(\mathrm{i}uY_1+\mathrm{i}vY_2))$. If $\Pr(Y_2\le 0)=0$, then the distribution of $Y=Y_1/Y_2$ is given by

$$F_Y(y)=\Pr\left(\frac{Y_1}{Y_2}\le y\right)=\frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\Im\left(\frac{\varphi(u,-uy)}{u}\right)\mathrm{d}u, \tag{6.7}$$

and the density function is

$$f_Y(y)=F_Y'(y)=\frac{1}{\pi}\int_0^{\infty}\Im\left(\frac{\partial\varphi(u,v)}{\partial v}|_{v=-uy}\right)\mathrm{d}u, \tag{6.8}$$

which can be used to derive $F_{\hat{\phi}}(d)$, $F_{\hat{\phi}}(-\phi)$, and $f_{\hat{\phi}}(d)$, and thus $F_{\hat{\kappa}}(w)$ and $f_{\hat{\kappa}}(w)$, via (6.5) and (6.6), respectively.

6.2.2 Characteristic Function

To be able to use (6.7) and (6.8), an essential task is to derive the characteristic function of $\hat{\phi}-\phi$. Let $\mathbf{0}_n$ be an $n\times 1$ vector of zeros, $\boldsymbol{I}_n$ be the identity matrix of size n, $\boldsymbol{\iota}_n$ be an $n\times 1$ vector of ones, $\boldsymbol{d}_n=(\boldsymbol{\iota}_{n-1},0)'$, $\boldsymbol{M}_n=\boldsymbol{I}_n-n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'$, and $\boldsymbol{e}_{i,n}$ be unit/elementary vector in n-dimensional Euclidean space with its ith element being 1. Denote $\boldsymbol{\chi}_{n+1}=(x_0,\boldsymbol{x}_n)'$, $\boldsymbol{x}_n=(x_1,\cdots,x_n)'$, $\boldsymbol{z}_n=\boldsymbol{x}_n/\sigma_\varepsilon$, $z_0=x_0/\sigma_\varepsilon$, and

$$\boldsymbol{A}_n^{\boldsymbol{C}}=\boldsymbol{A}_n(\boldsymbol{C}_{n-1})=\begin{pmatrix}\mathbf{0}_{n-1}' & 0\\ \boldsymbol{C}_{n-1} & \mathbf{0}_{n-1}\end{pmatrix},\quad \boldsymbol{B}_n^{\boldsymbol{C}}=\boldsymbol{B}_n(\boldsymbol{C}_{n-1})=\begin{pmatrix}\boldsymbol{C}_{n-1} & \mathbf{0}_{n-1}\\ \mathbf{0}_{n-1}' & 0\end{pmatrix}, \tag{6.9}$$

[5]Another issue is that Imhof's (1961) procedure is not directly applicable to work out $f_{\hat{\kappa}}(w)$, even though Lu (2006) discussed numerical evaluation of the probability distribution function of a normal quadratic form.

where $\boldsymbol{C}_{n-1}$ is an $(n-1)\times(n-1)$ matrix; when $\boldsymbol{C}_{n-1}=\boldsymbol{I}_{n-1}$, we suppress it and simply put $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$. For an $n\times n$ matrix $\boldsymbol{C}_n$, we use $c_{n,ij}$ to denote its ij-th element, and $c_n^{(ij)}$ to denote the ij-th element of $\boldsymbol{C}_n^{-1}$, whenever it exists.

6.2.2.1 Known Intercept ($\mu = 0$)

When the mean level μ is known (0), regardless of the value of mean-reverting parameter κ, the corresponding intercept in the discrete AR(1) model is zero and $\hat{\phi}$ is the ratio of $\sum_{i=1}^n x_{i-1}x_i$ to $\sum_{i=1}^n x_{i-1}^2$. Note that $(x_2,\cdots,x_3)'=(\boldsymbol{0}_{n-1},\boldsymbol{I}_{n-1})\,\boldsymbol{x}_n$ and $(x_1,\cdots,x_{n-1})'=(\boldsymbol{I}_{n-1},\boldsymbol{0}_{n-1})\,\boldsymbol{x}_n$, so

$$\begin{aligned}
\sum_{i=2}^n x_{i-1}x_i &= \boldsymbol{x}_n'\begin{pmatrix}\boldsymbol{0}_{n-1}'\\ \boldsymbol{I}_{n-1}\end{pmatrix}(\boldsymbol{I}_{n-1},\boldsymbol{0}_{n-1})\,\boldsymbol{x}_n\\
&= \boldsymbol{x}_n'\boldsymbol{A}_n\boldsymbol{x}_n,\\
\sum_{i=2}^n x_{i-1}^2 &= \boldsymbol{x}_n'\begin{pmatrix}\boldsymbol{I}_{n-1}\\ \boldsymbol{0}_{n-1}'\end{pmatrix}(\boldsymbol{I}_{n-1},\boldsymbol{0}_{n-1})\,\boldsymbol{x}_n\\
&= \boldsymbol{x}_n'\boldsymbol{B}_n\boldsymbol{x}_n.
\end{aligned}$$

Therefore, we can write $\sum_{i=1}^n x_{i-1}x_i = \boldsymbol{\chi}_{n+1}'\boldsymbol{A}_{n+1}\boldsymbol{\chi}_{n+1} = x_0\boldsymbol{x}_n'\boldsymbol{e}_{1,n}+\boldsymbol{x}_n'\boldsymbol{A}_n\boldsymbol{x}_n$ and $\sum_{i=1}^n x_{i-1}^2=\boldsymbol{\chi}_{n+1}'\boldsymbol{B}_{n+1}\boldsymbol{\chi}_{n+1}=x_0^2+\boldsymbol{x}_n'\boldsymbol{B}_n\boldsymbol{x}_n$.

If x_0 is fixed, then

$$\begin{aligned}
\hat{\phi}-\phi &= \frac{x_0x_1+\boldsymbol{x}_n'\boldsymbol{A}_n\boldsymbol{x}_n}{x_0^2+\boldsymbol{x}_n'\boldsymbol{B}_n\boldsymbol{x}_n}-\phi\\
&= \frac{\left[x_0\boldsymbol{x}_n'\boldsymbol{e}_{1,n}+\boldsymbol{x}_n'(\boldsymbol{A}_n-\phi\boldsymbol{B}_n)\boldsymbol{x}_n-\phi x_0^2\right]/\sigma_\varepsilon^2}{\left(x_0^2+\boldsymbol{x}_n'\boldsymbol{B}_n\boldsymbol{x}_n\right)/\sigma_\varepsilon^2}\\
&= \frac{z_0\boldsymbol{z}_n'\boldsymbol{e}_{1,n}+\boldsymbol{z}_n'(\boldsymbol{A}_n-\phi\boldsymbol{B}_n)\boldsymbol{z}_n-\phi z_0^2}{\left(z_0^2+\boldsymbol{z}_n'\boldsymbol{B}_n\boldsymbol{z}_n\right)}.
\end{aligned}$$

The density function of $\boldsymbol{z}_n$ (conditional on z_0) is

$$\begin{aligned}
f(\boldsymbol{z}_n) &= (2\pi)^{-\frac{n}{2}}\exp\left[-\frac{\sum_{i=1}^n(z_i-\phi z_{i-1})^2}{2}\right]\\
&= (2\pi)^{-\frac{n}{2}}\exp\left[-\frac{\phi^2z_0^2}{2}+\phi z_0\boldsymbol{z}_n'\boldsymbol{e}_{1,n}-\frac{1}{2}\boldsymbol{z}_n'\left(\boldsymbol{I}_n+\phi^2\boldsymbol{B}_n-2\phi\boldsymbol{A}_n\right)\boldsymbol{z}_n\right]
\end{aligned}$$

and the joint CF of the numerator and denominator in defining $\hat{\phi}-\phi$, conditional on z_0, is

$$\begin{aligned}
\varphi(u,v) &= (2\pi)^{-\frac{n}{2}} \int_{-\infty}^{+\infty} \exp\left\{ \mathrm{i}u \left[z_0 \boldsymbol{z}_n' \boldsymbol{e}_{1,n} + \boldsymbol{z}_n'(\boldsymbol{A}_n - \phi \boldsymbol{B}_n)\boldsymbol{z}_n - \phi z_0^2 \right] \right. \\
&\quad \left. + \mathrm{i}v \left(z_0^2 + \boldsymbol{z}_n' \boldsymbol{B}_n \boldsymbol{z}_n \right) \right\} f(\boldsymbol{z}_n)\, \mathrm{d}\boldsymbol{z}_n \\
&= (2\pi)^{-\frac{n}{2}} \exp\left(-\mathrm{i}u\phi z_0^2 + \mathrm{i}v z_0^2 - \frac{\phi^2 z_0^2}{2} \right) \\
&\quad \cdot \int_{-\infty}^{+\infty} \exp\{ -\frac{1}{2} \boldsymbol{z}_n' \left[\boldsymbol{I}_n + \left(\phi^2 + 2\mathrm{i}u\phi - 2\mathrm{i}v \right) \boldsymbol{B}_n \right. \\
&\quad \left. - (\phi + \mathrm{i}u) \left(\boldsymbol{A}_n + \boldsymbol{A}_n' \right) \right] \boldsymbol{z}_n + (\phi + \mathrm{i}u) z_0 \boldsymbol{z}_n' \boldsymbol{e}_{1,n} \} \mathrm{d}\boldsymbol{z}_n.
\end{aligned}$$

$$\boldsymbol{R}_n = \boldsymbol{R}_n(u,v) = \boldsymbol{I}_n + \left(\phi^2 + 2\mathrm{i}u\phi - 2\mathrm{i}v \right) \boldsymbol{B}_n - (\phi + \mathrm{i}u)(\boldsymbol{A}_n + \boldsymbol{A}_n'), \tag{6.10}$$

which is a tridiagonal matrix with its main diagonal elements $r_{n,ii} = 1 + \phi^2 + 2\mathrm{i}(u\phi - v)$, $i = 1, \cdots\ n-1$, $r_{n,ii} = 1$, $i = n$, and sub- and super-diagonal elements $-\phi - \mathrm{i}u$. Note that

$$\begin{aligned}
-\frac{1}{2} \boldsymbol{z}_n' \boldsymbol{R}_n \boldsymbol{z}_n + (\phi + \mathrm{i}u) z_0 \boldsymbol{z}_n' \boldsymbol{e}_{1,n} &= -\frac{1}{2} \left[\boldsymbol{z}_n - (\phi + \mathrm{i}u) z_0 \boldsymbol{R}_n^{-1} \boldsymbol{e}_{1,n} \right]' \boldsymbol{R}_n \\
&\quad \cdot \left[\boldsymbol{z}_n - (\phi + \mathrm{i}u) z_0 \boldsymbol{R}_n^{-1} \boldsymbol{e}_{1,n} \right] \\
&\quad + \frac{1}{2} \left[(\phi + \mathrm{i}u) z_0 \right]^2 \boldsymbol{e}_{1,n}' \boldsymbol{R}_n^{-1} \boldsymbol{e}_{1,n},
\end{aligned}$$

so

$$\begin{aligned}
\varphi(u,v) &= \exp\left\{ \frac{z_0^2}{2} \left(2\mathrm{i}v - 2\mathrm{i}u\phi - \phi^2 \right) + \frac{1}{2} \left[(\phi + \mathrm{i}u) z_0 \right]^2 \boldsymbol{e}_{1,n}' \boldsymbol{R}_n^{-1} \boldsymbol{e}_{1,n} \right\} \\
&\quad \cdot (2\pi)^{-\frac{n}{2}} \int_{-\infty}^{+\infty} \exp\left\{ -\frac{1}{2} \left[\boldsymbol{z}_n - (\phi + \mathrm{i}u) z_0 \boldsymbol{R}_n^{-1} \boldsymbol{e}_{1,n} \right]' \boldsymbol{R}_n \right. \\
&\quad \left. \cdot \left[\boldsymbol{z}_n - (\phi + \mathrm{i}u) z_0 \boldsymbol{R}_n^{-1} \boldsymbol{e}_{1,n} \right] \right\} \mathrm{d}\boldsymbol{z}_n \\
&= |\boldsymbol{R}_n|^{-1/2} \exp\left\{ \frac{z_0^2}{2} \left[-2\mathrm{i}(u\phi - v) - \phi^2 + (\phi + \mathrm{i}u)^2 \boldsymbol{e}_{1,n}' \boldsymbol{R}_n^{-1} \boldsymbol{e}_{1,n} \right] \right\}.
\end{aligned}$$

Also, $\boldsymbol{e}_{1,n}'\boldsymbol{R}_n^{-1}\boldsymbol{e}_{1,n} = r_n^{(11)}$, where $r_n^{(11)}$ is the element at the first row and column of $\boldsymbol{R}_n^{-1}$. It is obvious that $r_n^{(11)} = |\boldsymbol{R}_{n-1}| / |\boldsymbol{R}_n|$. By expanding along the first row of $\boldsymbol{R}_n$, we can verify $|\boldsymbol{R}_n| = \left[1 + \phi^2 + 2\mathrm{i}(u\phi - v)\right] |\boldsymbol{R}_{n-1}| - (\phi + \mathrm{i}u)^2 |\boldsymbol{R}_{n-2}|$, and thus $1 - |\boldsymbol{R}_{n+1}| / |\boldsymbol{R}_n| = -\phi^2 - 2\mathrm{i}(u\phi - v) + (\phi + \mathrm{i}u)^2 r_n^{(11)}$, which lead to

$$\varphi(u,v) = |\boldsymbol{R}_n|^{-1/2} \exp\left[\frac{z_0^2}{2}\left(1 - \frac{|\boldsymbol{R}_{n+1}|}{|\boldsymbol{R}_n|}\right)\right], \tag{6.11}$$

and thus

$$\frac{\partial\varphi(u,v)}{\partial v} = -\varphi(u,v)\left[\frac{\partial |\boldsymbol{R}_n|}{\partial v}\left(\frac{1}{2|\boldsymbol{R}_n|} - \frac{z_0^2 |\boldsymbol{R}_{n+1}|}{2|\boldsymbol{R}_n|^2}\right) + \frac{\partial |\boldsymbol{R}_{n+1}|}{\partial v}\frac{z_0^2}{2|\boldsymbol{R}_n|}\right]. \tag{6.12}$$

If $x_0 = 0$, then the characteristic function and its derivative degenerate to $\varphi(u,v) = |\boldsymbol{R}_n|^{-1/2}$ and $\partial\varphi(u,v)/\partial v = -(|\boldsymbol{R}_n|^{-3/2}/2)\partial|\boldsymbol{R}_n|/\partial v$, respectively.

If x_0 is random (and $\kappa > 0$), we write

$$\hat{\phi} - \phi = \frac{\boldsymbol{\chi}_{n+1}'(\boldsymbol{A}_{n+1} - \phi\boldsymbol{B}_{n+1})\boldsymbol{\chi}_{n+1}}{\boldsymbol{\chi}_{n+1}'\boldsymbol{B}_{n+1}\boldsymbol{\chi}_{n+1}},$$

which is invariant to σ_ε^2. Without loss of generality, normalize $\sigma_\varepsilon^2 = 1$, and the density function of $\boldsymbol{\chi}_{n+1}$ is

$$f(\boldsymbol{\chi}_{n+1}) = (2\pi)^{-\frac{n+1}{2}}\left|\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right|^{-1/2}\exp\left[-\frac{1}{2}\boldsymbol{\chi}_{n+1}'\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1}\boldsymbol{\chi}_{n+1}\right],$$

where $\boldsymbol{V}_{n+1}$ is $(n+1)\times(n+1)$ with its elements $v_{n,ij} = \phi^{|i-j|}$. Given its special structure, we can verify that $(1-\phi^2)\boldsymbol{V}_{n+1}^{-1}$ is tridiagonal with main diagonal elements 1 at positions 1 and $n+1$, $1+\phi^2$ at positions 2 to n, and sub- and super-diagonal elements

$-\phi$. Also, $\left|\boldsymbol{V}_{n+1}^{-1}\right| = (1-\phi^2)^{-n}$. Then immediately,

$$
\begin{aligned}
\varphi(u,v) &= (2\pi)^{-\frac{n+1}{2}} \left|\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right|^{-1/2} \\
&\quad \cdot \int_{-\infty}^{+\infty} \exp\Bigg\{ -\frac{1}{2}\boldsymbol{\chi}_{n+1}' \Bigg[\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1} - 2\mathrm{i}u\left(\boldsymbol{A}_{n+1} - \phi \boldsymbol{B}_{n+1}\right) \\
&\quad - 2\mathrm{i}v\boldsymbol{B}_{n+1} \Bigg] \boldsymbol{\chi}_{n+1} \Bigg\} \mathrm{d}\boldsymbol{\chi}_{n+1} \\
&= \sqrt{1-\phi^2}\left|\left(1-\phi^2\right)\boldsymbol{V}_{n+1}^{-1} - \mathrm{i}u(\boldsymbol{A}_{n+1} + \boldsymbol{A}_{n+1}') + 2\mathrm{i}\left(u\phi - v\right)\boldsymbol{B}_{n+1}\right|^{-1/2} \\
&= \sqrt{1-\phi^2}\left|\boldsymbol{R}_{n+1} - \phi^2 \boldsymbol{e}_{1,n+1}\boldsymbol{e}_{1,n+1}'\right|^{-1/2},
\end{aligned}
$$

where $\boldsymbol{R}_{n+1} - \phi^2 \boldsymbol{e}_{1,n+1}\boldsymbol{e}_{1,n+1}'$ is tridiagonal with its main diagonal elements $1+2\mathrm{i}(u\phi - v)$ at position 1, $1+\phi^2+2\mathrm{i}(u\phi-v)$ at positions 2 to n, and 1 at position $n+1$, and sub- and super-diagonal elements $-\phi - \mathrm{i}u$. Expanding $\boldsymbol{R}_{n+1} - \phi^2 \boldsymbol{e}_{1,n+1}\boldsymbol{e}_{1,n+1}'$ by its first row leads to $\left|\boldsymbol{R}_{n+1} - \phi^2 \boldsymbol{e}_{1,n+1}\boldsymbol{e}_{1,n+1}'\right| = [1+2\mathrm{i}(u\phi - v)]\left|\boldsymbol{R}_n\right| - (\phi + \mathrm{i}u)^2\left|\boldsymbol{R}_{n-1}\right|$. Recall $\left|\boldsymbol{R}_n\right| = \left[1+\phi^2+2\mathrm{i}(u\phi-v)\right]\left|\boldsymbol{R}_{n-1}\right| - (\phi+\mathrm{i}u)^2\left|\boldsymbol{R}_{n-2}\right|$. So $\left|\boldsymbol{R}_{n+1} - \phi^2 \boldsymbol{e}_{1,n+1}\boldsymbol{e}_{1,n+1}'\right| = \left|\boldsymbol{R}_{n+1}\right| - \phi^2\left|\boldsymbol{R}_n\right|$,

$$
\varphi(u,v) = \sqrt{1-\phi^2}\left(\left|\boldsymbol{R}_{n+1}\right| - \phi^2\left|\boldsymbol{R}_n\right|\right)^{-1/2}, \tag{6.13}
$$

and

$$
\frac{\partial \varphi(u,v)}{\partial v} = -\frac{\varphi(u,v)}{2\left(\left|\boldsymbol{R}_{n+1}\right| - \phi^2\left|\boldsymbol{R}_n\right|\right)}\left(\frac{\partial\left|\boldsymbol{R}_{n+1}\right|}{\partial v} - \phi^2\frac{\partial\left|\boldsymbol{R}_n\right|}{\partial v}\right). \tag{6.14}
$$

6.2.2.2 Unknown Intercept ($\mu \neq 0$ and $\kappa > 0$)

When the mean level μ is unknown and the mean-reverting parameter $\kappa > 0$, the corresponding intercept in the discrete AR(1) model is nonzero and $\hat{\phi}$ is the ratio of

$\sum_{i=1}^{n}(x_{i-1}-\bar{x})x_i$ to $\sum_{i=1}^{n}(x_{i-1}-\bar{x})^2$. Note that

$$\begin{aligned}
(x_0-\bar{x},\cdots,x_{n-1}-\bar{x})' &= \boldsymbol{M}_n(x_0,\cdots,x_{n-1})' \\
&= \boldsymbol{M}_n(0,x_1,\cdots,x_{n-1})' + \boldsymbol{M}_n\,(x_0,0,\cdots 0)' \\
&= \boldsymbol{M}_n \begin{pmatrix} \boldsymbol{0}'_{n-1} & 0 \\ \boldsymbol{I}_{n-1} & \boldsymbol{0}_{n-1} \end{pmatrix} \boldsymbol{x}_n + x_0\boldsymbol{M}_n\boldsymbol{e}_{1,n} \\
&\equiv \boldsymbol{M}_n\boldsymbol{A}_n\boldsymbol{x}_n + x_0\boldsymbol{M}_n\boldsymbol{e}_{1,n},
\end{aligned}$$

so we can write

$$\begin{aligned}
\sum_{i=1}^{n}(x_{i-1}-\bar{x})x_i &= \boldsymbol{x}'_n\boldsymbol{M}_n\boldsymbol{A}_n\boldsymbol{x}_n + x_0\boldsymbol{x}'_n\boldsymbol{M}_n\boldsymbol{e}_{1,n} = \boldsymbol{\chi}'_{n+1}\boldsymbol{A}^{\boldsymbol{M}}_{n+1}\boldsymbol{\chi}_{n+1}, \\
\sum_{i=1}^{n}(x_{i-1}-\bar{x})^2 &= \boldsymbol{x}'_n\boldsymbol{A}'_n\boldsymbol{M}_n\boldsymbol{A}_n\boldsymbol{x}_n + x_0^2\boldsymbol{e}'_{1,n}\boldsymbol{M}_n\boldsymbol{e}_{1,n} + 2x_0\boldsymbol{x}'_n\boldsymbol{A}'_n\boldsymbol{M}_n\boldsymbol{e}_{1,n} \\
&= \boldsymbol{\chi}'_{n+1}\boldsymbol{B}^{\boldsymbol{M}}_{n+1}\boldsymbol{\chi}_{n+1},
\end{aligned}$$

where $\boldsymbol{x}_n = (\boldsymbol{0}_n, \boldsymbol{I}_n)\boldsymbol{\chi}_{n+1}$, and $\boldsymbol{A}^{\boldsymbol{M}}_{n+1}$ and $\boldsymbol{B}^{\boldsymbol{M}}_{n+1}$ are defined in (6.9).

If x_0 is fixed, then

$$\begin{aligned}
\hat{\phi}-\phi &= \frac{\boldsymbol{x}'_n\boldsymbol{M}_n\boldsymbol{A}_n\boldsymbol{x}_n + x_0\boldsymbol{x}'_n\boldsymbol{M}_n\boldsymbol{e}_{1,n}}{\boldsymbol{x}'_n\boldsymbol{A}'_n\boldsymbol{M}_n\boldsymbol{A}_n\boldsymbol{x}_n + x_0^2\boldsymbol{e}'_{1,n}\boldsymbol{M}_n\boldsymbol{e}_{1,n} + 2x_0\boldsymbol{x}'_n\boldsymbol{A}'_n\boldsymbol{M}_n\boldsymbol{e}_{1,n}} - \phi \\
&= \frac{\boldsymbol{z}'_n(\boldsymbol{M}_n\boldsymbol{A}_n - \phi\boldsymbol{A}'_n\boldsymbol{M}_n\boldsymbol{A}_n)\boldsymbol{z}_n + z_0\boldsymbol{z}'_n(\boldsymbol{I}_n - 2\phi\boldsymbol{A}'_n)\boldsymbol{M}_n\boldsymbol{e}_{1,n} - \phi z_0^2\boldsymbol{e}'_{1,n}\boldsymbol{M}_n\boldsymbol{e}_{1,n}}{\boldsymbol{z}'_n\boldsymbol{A}'_n\boldsymbol{M}_n\boldsymbol{A}_n\boldsymbol{z}_n + 2z_0\boldsymbol{z}'_n\boldsymbol{A}'_n\boldsymbol{M}_n\boldsymbol{e}_{1,n} + z_0^2\boldsymbol{e}'_{1,n}\boldsymbol{M}_n\boldsymbol{e}_{1,n}}.
\end{aligned}$$

The density function of $\boldsymbol{z}_n$ (conditional on z_0) is

$$\begin{aligned}
f(\boldsymbol{z}_n) &= (2\pi)^{-\frac{n}{2}}\exp\left[-\frac{\sum_{i=1}^{n}(z_i-\alpha-\phi z_{i-1})^2}{2}\right] \\
&= (2\pi)^{-\frac{n}{2}}\exp\Bigg\{-\frac{1}{2}\bigg[\boldsymbol{z}'_n\left(\boldsymbol{I}_n + \phi^2\boldsymbol{B}_n - 2\phi\boldsymbol{A}_n\right)\boldsymbol{z}_n \\
&\quad + 2\boldsymbol{z}'_n\left(\alpha\phi\boldsymbol{d}_n - \alpha\boldsymbol{\iota}_n - \phi z_0\boldsymbol{e}_{1,n}\right) + n\alpha^2 + \phi^2 z_0^2 + 2\alpha\phi z_0\bigg]\Bigg\},
\end{aligned}$$

and the joint CF of the numerator and denominator in defining $\hat{\phi}-\phi$ is

$$\begin{aligned}\varphi(u,v) &= (2\pi)^{-\frac{n}{2}}\exp\left[-\frac{n\alpha^2+\phi^2z_0^2+2\alpha\phi z_0}{2}+(\mathrm{i}v-\mathrm{i}u\phi)\,z_0^2\boldsymbol{e}_{1,n}'\boldsymbol{M}_n\boldsymbol{e}_{1,n}\right]\\ &\quad\cdot\int_{-\infty}^{+\infty}\exp\Bigg\{-\frac{1}{2}\boldsymbol{z}_n'\Bigg[\boldsymbol{I}_n+\phi^2\boldsymbol{B}_n-2\phi\boldsymbol{A}_n-2\mathrm{i}u\left(\boldsymbol{M}_n\boldsymbol{A}_n-\phi\boldsymbol{A}_n'\boldsymbol{M}_n\boldsymbol{A}_n\right)\\ &\quad-2\mathrm{i}v\boldsymbol{A}_n'\boldsymbol{M}_n\boldsymbol{A}_n\Bigg]\boldsymbol{z}_n+\boldsymbol{z}_n'\boldsymbol{\delta}_n\Bigg\}\mathrm{d}\boldsymbol{z}_n,\end{aligned}$$

where

$$\boldsymbol{\delta}_n=\boldsymbol{\delta}_n(u,v)=\mathrm{i}uz_0(\boldsymbol{I}_n-2\phi\boldsymbol{A}_n')\boldsymbol{M}_n\boldsymbol{e}_{1,n}+2\mathrm{i}vz_0\boldsymbol{A}_n'\boldsymbol{M}_n\boldsymbol{e}_{1,n}-(\alpha\phi\boldsymbol{d}_n-\alpha\boldsymbol{\iota}_n-\phi z_0\boldsymbol{e}_{1,n})\,, \tag{6.15}$$

and the (symmetrized) matrix in the quadratic form of $\boldsymbol{z}_n$ in the exponent of the integral, denoted by $\boldsymbol{S}_n=\boldsymbol{S}_n(u,v)$, can be written as

$$\begin{aligned}\boldsymbol{S}_n &= \boldsymbol{I}_n+\phi^2\boldsymbol{B}_n-\phi\left(\boldsymbol{A}_n+\boldsymbol{A}_n'\right)-\mathrm{i}u(\boldsymbol{M}_n\boldsymbol{A}_n+\boldsymbol{A}_n'\boldsymbol{M}_n)+2\mathrm{i}u\phi\boldsymbol{A}_n'\boldsymbol{M}_n\boldsymbol{A}_n \qquad (6.16)\\ &\quad -2\mathrm{i}v\boldsymbol{A}_n'\boldsymbol{M}_n\boldsymbol{A}_n\\ &= \boldsymbol{I}_n+\phi^2\boldsymbol{B}_n-\phi\left(\boldsymbol{A}_n+\boldsymbol{A}_n'\right)-\mathrm{i}u(\boldsymbol{A}_n+\boldsymbol{A}_n')+2\mathrm{i}u\phi\boldsymbol{A}_n'\boldsymbol{A}_n-2\mathrm{i}v\boldsymbol{A}_n'\boldsymbol{A}_n\\ &\quad +n^{-1}\mathrm{i}u\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n+n^{-1}\mathrm{i}u\boldsymbol{A}_n'\boldsymbol{\iota}_n\boldsymbol{\iota}_n'-2n^{-1}\mathrm{i}u\phi\boldsymbol{A}_n'\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n+2n^{-1}\mathrm{i}v\boldsymbol{A}_n'\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n\\ &= \boldsymbol{R}_n+\frac{\mathrm{i}}{n}(u\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n+u\boldsymbol{A}_n'\boldsymbol{\iota}_n\boldsymbol{\iota}_n'-2u\phi\boldsymbol{A}_n'\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n+2v\boldsymbol{A}_n'\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n)\\ &= \boldsymbol{R}_n+\frac{\mathrm{i}}{n}\begin{pmatrix}2(u+v-u\phi)\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}_{n-1}' & u\boldsymbol{\iota}_{n-1}\\ u\boldsymbol{\iota}_{n-1}' & 0\end{pmatrix}, \qquad (6.17)\end{aligned}$$

by noticing

$$\boldsymbol{A}_n'\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n=\begin{pmatrix}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}_{n-1}' & \boldsymbol{0}_{n-1}\\ \boldsymbol{0}_{n-1}' & 0\end{pmatrix},\quad \boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{A}_n=\begin{pmatrix}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}_{n-1}' & \boldsymbol{0}_{n-1}\\ \boldsymbol{\iota}_{n-1}' & 0\end{pmatrix}.$$

Thus we have

$$\begin{aligned}\varphi(u,v) &= |\boldsymbol{S}_n|^{-1/2}\exp\Bigg[-\frac{n\alpha^2+\phi^2z_0^2+2\alpha\phi z_0}{2}+\mathrm{i}\,(v-u\phi)\,z_0^2\boldsymbol{e}_{1,n}'\boldsymbol{M}_n\boldsymbol{e}_{1,n}\\ &\quad+\frac{1}{2}\boldsymbol{\delta}_n'\boldsymbol{S}_n^{-1}\boldsymbol{\delta}_n\Bigg],\end{aligned} \tag{6.18}$$

and

$$\frac{\partial\varphi\left(u,v\right)}{\partial v}=\varphi\left(u,v\right)\big(\mathrm{i}z_0^2\boldsymbol{e}_{1,n}'\boldsymbol{M}_n\boldsymbol{e}_{1,n}+2\mathrm{i}z_0\boldsymbol{e}_{1,n}'\boldsymbol{M}_n\boldsymbol{A}_n\boldsymbol{S}_n^{-1}\boldsymbol{\delta}_n$$
$$+\frac{1}{2}\boldsymbol{\delta}_n'\frac{\partial\boldsymbol{S}_n^{-1}}{\partial v}\boldsymbol{\delta}_n-\frac{1}{2\left|\boldsymbol{S}_n\right|}\frac{\partial\left|\boldsymbol{S}_n\right|}{\partial v}\bigg). \tag{6.19}$$

When x_0 is random,

$$\hat{\phi}-\phi=\frac{\boldsymbol{\chi}_{n+1}'\left(\boldsymbol{A}_{n+1}^{\boldsymbol{M}}-\phi\boldsymbol{B}_{n+1}^{\boldsymbol{M}}\right)\boldsymbol{\chi}_{n+1}}{\boldsymbol{\chi}_{n+1}'\boldsymbol{B}_{n+1}^{\boldsymbol{M}}\boldsymbol{\chi}_{n+1}},$$

which is invariant to σ_ε^2. With σ_ε^2 normalized to be 1, the density of $\boldsymbol{\chi}_{n+1}$ is

$$f\left(\boldsymbol{\chi}_{n+1}\right)=(2\pi)^{-\frac{n+1}{2}}\left|\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right|^{-1/2}$$
$$\cdot\exp\left[-\frac{1}{2}\left(\boldsymbol{\chi}_{n+1}-\frac{\alpha\boldsymbol{\iota}_{n+1}}{1-\phi}\right)'\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1}\left(\boldsymbol{\chi}_{n+1}-\frac{\alpha\boldsymbol{\iota}_{n+1}}{1-\phi}\right)\right],$$

and the joint CF of $\boldsymbol{\chi}_{n+1}'\left(\boldsymbol{A}_{n+1}^{\boldsymbol{M}}-\phi\boldsymbol{B}_{n+1}^{\boldsymbol{M}}\right)\boldsymbol{\chi}_{n+1}$ and $\boldsymbol{\chi}_{n+1}'\boldsymbol{B}_{n+1}^{\boldsymbol{M}}\boldsymbol{\chi}_{n+1}$ is

$$\begin{aligned}\varphi\left(u,v\right) &= (2\pi)^{-\frac{n+1}{2}}\left|\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right|^{-1/2}\exp\left[-\frac{1}{2}\frac{\alpha^2}{(1-\phi)^2}\boldsymbol{\iota}_{n+1}'\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1}\boldsymbol{\iota}_{n+1}\right]\\ &\quad\cdot\int_{-\infty}^{+\infty}\exp\Bigg\{-\frac{1}{2}\boldsymbol{\chi}_{n+1}'\Bigg[\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1}-2\mathrm{i}u\left(\boldsymbol{A}_{n+1}^{\boldsymbol{M}}-\phi\boldsymbol{B}_{n+1}^{\boldsymbol{M}}\right)\\ &\quad-2\mathrm{i}v\boldsymbol{B}_{n+1}^{\boldsymbol{M}}\Bigg]\boldsymbol{\chi}_{n+1}+\frac{\alpha}{1-\phi}\boldsymbol{\chi}_{n+1}'\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1}\boldsymbol{\iota}_{n+1}\Bigg\}\mathrm{d}\boldsymbol{\chi}_{n+1}\\ &= (2\pi)^{-\frac{n+1}{2}}\left|\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right|^{-1/2}\exp\left[-\frac{1}{2}\boldsymbol{\gamma}_{n+1}'\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)\boldsymbol{\gamma}_{n+1}\right]\\ &\quad\cdot\int_{-\infty}^{+\infty}\exp\Bigg\{-\frac{1}{2}\boldsymbol{\chi}_{n+1}'\left[\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1}-2\mathrm{i}u\boldsymbol{A}_{n+1}^{\boldsymbol{M}}+2\mathrm{i}(u\phi-v)\boldsymbol{B}_{n+1}^{\boldsymbol{M}}\right]\boldsymbol{\chi}_{n+1}\\ &\quad+\boldsymbol{\chi}_{n+1}'\boldsymbol{\gamma}_{n+1}\Bigg\}\mathrm{d}\boldsymbol{\chi}_{n+1},\end{aligned}$$

where

$$\boldsymbol{\gamma}_{n+1}=\frac{\alpha}{1-\phi}\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1}\boldsymbol{\iota}_{n+1}. \tag{6.20}$$

Note that

$$\boldsymbol{\gamma}_{n+1}'\left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)\boldsymbol{\gamma}_{n+1}=n\alpha^2+\left(1+\phi\right)\alpha\mu,$$

and that the (symmetrized) matrix in the quadratic form of $\boldsymbol{\chi}_{n+1}$ in the exponent of the above integral, denoted by $\boldsymbol{T}_{n+1} = \boldsymbol{T}_{n+1}(u,v)$, is

$$\begin{aligned}
\boldsymbol{T}_{n+1} &= \left(\frac{\boldsymbol{V}_{n+1}}{1-\phi^2}\right)^{-1} - \mathrm{i}u(\boldsymbol{A}^{\boldsymbol{M}}_{n+1} + \boldsymbol{A}^{\boldsymbol{M}\prime}_{n+1}) + 2\mathrm{i}(u\phi - v)\boldsymbol{B}^{\boldsymbol{M}}_{n+1} \\
&= (1-\phi^2)\boldsymbol{V}^{-1}_{n+1} - \mathrm{i}u(\boldsymbol{A}_{n+1} + \boldsymbol{A}'_{n+1}) + 2\mathrm{i}(u\phi - v)\boldsymbol{B}_{n+1} \\
&\quad + \frac{\mathrm{i}u}{n}\left[\begin{pmatrix} \boldsymbol{0}'_n & 0 \\ \boldsymbol{\iota}_n\boldsymbol{\iota}'_n & \boldsymbol{0}_n \end{pmatrix} + \begin{pmatrix} \boldsymbol{0}_n & \boldsymbol{\iota}_n\boldsymbol{\iota}'_n \\ 0 & \boldsymbol{0}'_n \end{pmatrix}\right] - \frac{2\mathrm{i}(u\phi - v)}{n}\begin{pmatrix} \boldsymbol{\iota}_n\boldsymbol{\iota}'_n & \boldsymbol{0}_n \\ \boldsymbol{0}'_n & 0 \end{pmatrix} \\
&= \boldsymbol{R}_{n+1} - \phi^2 \boldsymbol{e}_{1,n+1}\boldsymbol{e}'_{1,n+1} \\
&\quad + \frac{\mathrm{i}}{n}\begin{pmatrix} 2(v-u\phi) & (u+2v-2u\phi)\boldsymbol{\iota}'_{n-1} & u \\ (u+2v-2u\phi)\boldsymbol{\iota}_{n-1} & 2(v+u-u\phi)\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}'_{n-1} & u\boldsymbol{\iota}_{n-1} \\ u & u\boldsymbol{\iota}'_{n-1} & 0 \end{pmatrix}. \qquad (6.21)
\end{aligned}$$

Therefore,

$$\varphi(u,v) = \sqrt{1-\phi^2}\,|\boldsymbol{T}_{n+1}|^{-1/2}\exp\left(-\frac{n\alpha^2 + (1+\phi)\,\alpha\mu}{2} + \frac{1}{2}\boldsymbol{\gamma}'_{n+1}\boldsymbol{T}^{-1}_{n+1}\boldsymbol{\gamma}_{n+1}\right), \qquad (6.22)$$

and

$$\frac{\partial\varphi(u,v)}{\partial v} = \frac{\varphi(u,v)}{2}\left(\boldsymbol{\gamma}'_{n+1}\frac{\partial \boldsymbol{T}^{-1}_{n+1}}{\partial v}\boldsymbol{\gamma}_{n+1} - \frac{1}{|\boldsymbol{T}_{n+1}|}\frac{\partial|\boldsymbol{T}_{n+1}|}{\partial v}\right). \qquad (6.23)$$

6.2.2.3 Unknown Intercept ($\mu \neq 0$ but $\kappa = 0$)

When the mean level μ is unknown but the mean-reverting parameter $\kappa = 0$, the corresponding intercept in the discrete AR(1) model is zero. Given that the intercept term is unknown, we still estimate $\hat{\phi}$ as the ratio of $\sum_{i=1}^{n}(x_{i-1}-\bar{x})x_i$ to $\sum_{i=1}^{n}(x_{i-1}-\bar{x})^2$, though the true intercept is zero. The joint CF $\varphi(u,v)$ and its partial derivative with respective to v of the numerator and denominator in defining $\hat{\phi}-\phi$ are (6.18) and (6.19), respectively, with $\phi = 1$ and $\alpha = 0$. Note that when $\kappa = 0$, we consider only the case when x_0 is fixed.

6.2.3 Evaluation of Characteristic Functions

To evaluate the characteristic functions (6.11), (6.13), (6.18), and (6.22) with $\boldsymbol{R}_n$, $\boldsymbol{\delta}_n$, $\boldsymbol{S}_n$, $\boldsymbol{\gamma}_{n+1}$, and $\boldsymbol{T}_{n+1}$ defined in (6.10), (6.15), (6.17), (6.20), and (6.21), respectively, we need to find the determinants and inverses of $(n \times n)$ and $(n+1) \times (n+1)$ matrices.

As emphasized in Tsui and Ali (1994), the typical way of evaluating matrix determinant/inverse by using eigenvalues can be very expensive as n increases. When there is no intercept term in the AR(1) model, we have already noticed that $|\boldsymbol{R}_n| = \left[1+\phi^2+2\mathrm{i}(u\phi - v)\right]|\boldsymbol{R}_{n-1}| - (\phi + \mathrm{i}u)^2 |\boldsymbol{R}_{n-2}|$, starting with $|\boldsymbol{R}_1| = 1$ and $|\boldsymbol{R}_2| = 1 + \phi^2 + 2\mathrm{i}(u\phi - v) - (\phi+\mathrm{i}u)^2$. Tsui and Ali (1992, 1994) proposed expanding along the last row of $\boldsymbol{R}_n$ so that

$$|\boldsymbol{R}_n| = |\boldsymbol{D}_{n-1}| - (\phi + \mathrm{i}u)^2 |\boldsymbol{D}_{n-2}|, \tag{6.24}$$

where $\boldsymbol{D}_n = \boldsymbol{D}_n(u,v)$ is the determinant of an $n \times n$ tridiagonal matrix with $1+\phi^2+2\mathrm{i}(u\phi - v)$ on its main diagonal and $-(\phi+\mathrm{i}u)$ on its super- and sub- diagonals, which can be evaluated by using the result of Muir (1884):

$$|\boldsymbol{D}_n| = \prod_{i=1}^{n}\left[1+\phi^2+2\mathrm{i}(u\phi - v) - 2(\phi+\mathrm{i}u)\cos\left(\pi i/(n+1)\right)\right].$$

A probably more direct and efficient way (see Berstein (2009, page 235)) is perhaps to use

$$|\boldsymbol{D}_n| = \begin{cases} (n+1)\left[\frac{1+\phi^2+2\mathrm{i}(u\phi-v)}{2}\right]^n & \left[1+\phi^2+2\mathrm{i}(u\phi - v)\right]^2 = 4(\phi+\mathrm{i}u)^2 \\ \frac{\beta_1^{n+1}-\beta_2^{n+1}}{\beta_1-\beta_2} & \left[1+\phi^2+2\mathrm{i}(u\phi - v)\right]^2 \neq 4(\phi+\mathrm{i}u)^2 \end{cases}, \tag{6.25}$$

where

$$\begin{aligned}\beta_1 &= \frac{1+\phi^2+2\mathrm{i}(u\phi-v)+\sqrt{\left[1+\phi^2+2\mathrm{i}(u\phi-v)\right]^2 - 4(\phi+\mathrm{i}u)^2}}{2}, \\ \beta_2 &= \frac{1+\phi^2+2\mathrm{i}(u\phi-v)-\sqrt{\left[1+\phi^2+2\mathrm{i}(u\phi-v)\right]^2 - 4(\phi+\mathrm{i}u)^2}}{2}.\end{aligned}$$

The more challenging task is to deal with the case when $\alpha \neq 0$ and $\kappa > 0$, namely, to evaluate the determinants and inverses of $\boldsymbol{S}_n$ and $\boldsymbol{T}_n$. First, we note that $\boldsymbol{S}_n$ has some special structure:

$$\boldsymbol{S}_n = \boldsymbol{R}_n + \frac{\mathrm{i}}{n}\begin{bmatrix} 2(u+v-u\phi)\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}'_{n-1} & u\boldsymbol{\iota}_{n-1} \\ u\boldsymbol{\iota}'_{n-1} & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Delta}_{n-1} & \boldsymbol{a}_{n-1} \\ \boldsymbol{a}_{n-1} & 1 \end{bmatrix},$$

where

$$\boldsymbol{\Delta}_{n-1} = \boldsymbol{D}_{n-1} + \frac{2\mathrm{i}(u+v-u\phi)}{n}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}'_{n-1}, \quad \boldsymbol{a}_{n-1} = \frac{\mathrm{i}u}{n}\boldsymbol{\iota}_{n-1} - (\phi + \mathrm{i}u)\,\boldsymbol{e}_{n-1,n-1}.$$

Note that

$$|\boldsymbol{S}_n| = |\boldsymbol{\Delta}_{n-1}|\left(1 - \boldsymbol{a}'_{n-1}\boldsymbol{\Delta}^{-1}_{n-1}\boldsymbol{a}_{n-1}\right), \tag{6.26}$$

where

$$|\boldsymbol{\Delta}_{n-1}| = |\boldsymbol{D}_{n-1}|\left(1 + \frac{2\mathrm{i}(u+v-u\phi)}{n}\boldsymbol{\iota}'_{n-1}\boldsymbol{D}^{-1}_{n-1}\boldsymbol{\iota}_{n-1}\right), \tag{6.27}$$

and

$$\boldsymbol{\Delta}^{-1}_{n-1} = \boldsymbol{D}^{-1}_{n-1} - \frac{2\mathrm{i}(u+v-u\phi)}{n + 2\mathrm{i}(u+v-u\phi)\boldsymbol{\iota}'_{n-1}\boldsymbol{D}^{-1}_{n-1}\boldsymbol{\iota}_{n-1}}\boldsymbol{D}^{-1}_{n-1}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}'_{n-1}\boldsymbol{D}^{-1}_{n-1}. \tag{6.28}$$

Keep in mind that (6.26) is valid only if $\boldsymbol{\Delta}_{n-1}$ is nonsingular; (6.27) is valid only if $\boldsymbol{D}_{n-1}$ is nonsingular; (6.28) is valid only if $\boldsymbol{D}_{n-1}$ is nonsingular and $n + 2\mathrm{i}(u+v-u\phi)\boldsymbol{\iota}'_{n-1}\boldsymbol{D}^{-1}_{n-1}\boldsymbol{\iota}_{n-1} \neq 0$. From (6.25), we see that $|\boldsymbol{D}_n| \neq 0$; Appendix A section IV part (i) also shows that $n + 2\mathrm{i}(u+v-u\phi)\boldsymbol{\iota}'_{n-1}\boldsymbol{D}^{-1}_{n-1}\boldsymbol{\iota}_{n-1} \neq 0$. Further, these two conditions ensure that $|\boldsymbol{\Delta}_{n-1}| \neq 0$.

Given that we already know how to evaluate analytically the determinant of $\boldsymbol{D}_n$ via (6.25), we need to work out $\boldsymbol{D}^{-1}_n$ to be able to evaluate (6.26) via (6.27) and (6.28) . From Hu and O'Connell (1996), with slight modification[6]:

[6] Hu and O'Connell (1996) presents the result for the case when the main diagonal is real. Yes, it is still valid when the condition regarding the real diagonal is changed to its real part if the diagonal is complex. In our case, the matrix $\boldsymbol{D}_n$ is divided by $-(\phi+\mathrm{i}u)$, which is always non-zero as $\phi > 0$, so the new matrix has main diagonal $[1+\phi^2+2\mathrm{i}(u\phi-v)]/[-(\phi+\mathrm{i}u)]$ and super- and sub- diagonals 1. Now the real part of $[1+\phi^2+2\mathrm{i}(u\phi-v)]/[-(\phi+\mathrm{i}u)]$ is $\left[-\phi\left(1+\phi^2\right) - 2u(u\phi-v)\right]/\left(\phi^2+u^2\right)$. Their determinant result is also valid with this modification, comparable to (6.25).

$$d_n^{(ij)} = \begin{cases} -\frac{\cosh[(n+1-|j-i|)\lambda]-\cosh[(n+1-i-j)\lambda]}{2\sinh(\lambda)\sinh[(n+1)\lambda]}\left(\frac{\mathrm{i}u-\phi}{\phi^2+u^2}\right), & (i) \\ (-1)^{i+j}\frac{\cosh[(n+1-|j-i|)\lambda]-\cosh[(n+1-i-j)\lambda]}{2\sinh(\lambda)\sinh[(n+1)\lambda]}\left(\frac{\mathrm{i}u-\phi}{\phi^2+u^2}\right), & (ii) \\ (-1)^{i+j+1}\frac{\cos[(n+1-|j-i|)\lambda]-\cos[(n+1-i-j)\lambda]}{2\sin(\lambda)\sin[(n+1)\lambda]}\left(\frac{\mathrm{i}u-\phi}{\phi^2+u^2}\right), & (iii) \end{cases}, \tag{6.29}$$

where

$$(i)\frac{-\phi\left(1+\phi^2\right)-2u(u\phi-v)}{\phi^2+u^2} \leq -2$$
$$(ii)\frac{-\phi\left(1+\phi^2\right)-2u(u\phi-v)}{\phi^2+u^2} \geq 2$$
$$(iii) - 2 < \frac{-\phi\left(1+\phi^2\right)-2u(u\phi-v)}{\phi^2+u^2} < 2$$

with corresponding λ equal to $\operatorname{arccosh}([1+\phi^2+2\mathrm{i}(u\phi-v)]/[2(\phi+\mathrm{i}u)])$, $\operatorname{arccosh}(-[1+\phi^2+2\mathrm{i}(u\phi-v)]/[2(\phi+\mathrm{i}u)])$, and $\arccos(-[1+\phi^2+2\mathrm{i}(u\phi-v)]/[2(\phi+\mathrm{i}u)])$, respectively. The above analytical inverse is "piece-wise". For fast programming and in the need of the derivatives of $\boldsymbol{D}_n^{-1}$, an equivalent formula is given by da Fonseca and Petronilho (2001) (note that $\boldsymbol{D}_n^{-1}$ is symmetric),

$$d_n^{(ij)} = (-1)^{i+j}\frac{1}{-(\phi+\mathrm{i}u)}\frac{U_{i-1}\left(\frac{1+\phi^2+2\mathrm{i}(u\phi-v)}{-2(\phi+\mathrm{i}u)}\right)U_{n-j}\left(\frac{1+\phi^2+2\mathrm{i}(u\phi-v)}{-2(\phi+\mathrm{i}u)}\right)}{U_n\left(\frac{1+\phi^2+2\mathrm{i}(u\phi-v)}{-2(\phi+\mathrm{i}u)}\right)}, \ i \leq j \tag{6.30}$$

where $U_n(x)$ is Chebyshev polynomial of the second kind, defined by a second-order recursion, $U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x)$, $n \geq 1$, with two initial conditions $U_0(x) = 1$ and $U_1(x) = 2x$. It also has an analytic expression: $U_n(x) = 2^n \prod_{i=1}^n [x - \cos(\pi i/(n+1))]$.

With $\boldsymbol{\Delta}_{n-1}^{-1}$ given by (6.28), evaluation of the inverse of $\boldsymbol{S}_n$ is straightforward:

$$\boldsymbol{S}_n^{-1} = \begin{pmatrix} \boldsymbol{\Delta}_{n-1}^{-1} & \boldsymbol{0}_{n-1} \\ \boldsymbol{0}'_{n-1} & 0 \end{pmatrix} + \frac{1}{1-\boldsymbol{a}'_{n-1}\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{a}_{n-1}}\begin{pmatrix} \boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{a}_{n-1} \\ -1 \end{pmatrix}\begin{pmatrix} \boldsymbol{a}'_{n-1}\boldsymbol{\Delta}_{n-1}^{-1} & -1 \end{pmatrix}, \tag{6.31}$$

if $1-\boldsymbol{a}'_{n-1}\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{a}_{n-1} \neq 0$, which is verified to hold true in Appendix A section IV part (i).

Next, we write

$$\boldsymbol{T}_{n+1} = \begin{pmatrix} 1 + 2\mathrm{i}(u\phi - v) + \frac{2\mathrm{i}(v-u\phi)}{n} & \boldsymbol{b}'_{n-1} & \frac{\mathrm{i}u}{n} \\ \boldsymbol{b}_{n-1} & \boldsymbol{\Delta}_{n-1} & \boldsymbol{a}_{n-1} \\ \frac{\mathrm{i}u}{n} & \boldsymbol{a}'_{n-1} & 1 \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{\Delta}^*_n & \boldsymbol{a}^*_n \\ \boldsymbol{a}^{*\prime}_n & 1 \end{pmatrix},$$

where

$$\boldsymbol{b}_{n-1} = \frac{\mathrm{i}(u + 2v - 2u\phi)}{n}\boldsymbol{\iota}_{n-1} - (\phi + \mathrm{i}u)\,\boldsymbol{e}_{1,n-1}.$$

Following the same strategy as before,

$$|\boldsymbol{T}_{n+1}| = |\boldsymbol{\Delta}^*_n|\left(1 - \boldsymbol{a}^{*\prime}_n \boldsymbol{\Delta}^{*-1}_n \boldsymbol{a}^*_n\right), \tag{6.32}$$

where

$$|\boldsymbol{\Delta}^*_n| = |\boldsymbol{\Delta}_{n-1}|\left(1 + 2\mathrm{i}(u\phi - v) + \frac{2\mathrm{i}(v - u\phi)}{n} - \boldsymbol{b}'_{n-1}\boldsymbol{\Delta}^{-1}_{n-1}\boldsymbol{b}_{n-1}\right), \tag{6.33}$$

and

$$\begin{aligned} \boldsymbol{\Delta}^{*-1}_n &= \begin{pmatrix} 0 & \boldsymbol{0}'_{n-1} \\ \boldsymbol{0}_{n-1} & \boldsymbol{\Delta}^{-1}_{n-1} \end{pmatrix} \\ &+ \frac{1}{1 + 2\mathrm{i}(u\phi - v) + \frac{2\mathrm{i}(v-u\phi)}{n} - \boldsymbol{b}'_{n-1}\boldsymbol{\Delta}^{-1}_{n-1}\boldsymbol{b}_{n-1}} \\ &\cdot \begin{pmatrix} -1 \\ \boldsymbol{\Delta}^{-1}_{n-1}\boldsymbol{b}_{n-1} \end{pmatrix} \begin{pmatrix} -1 & \boldsymbol{b}'_{n-1}\boldsymbol{\Delta}^{-1}_{n-1} \end{pmatrix}. \end{aligned} \tag{6.34}$$

(Note that $|\boldsymbol{\Delta}_{n-1}|$ and $\boldsymbol{\Delta}^{-1}_{n-1}$ are given by (6.27) and (6.28), respectively.) With $\boldsymbol{\Delta}^{*-1}_n$ given above, evaluation of the inverse of $\boldsymbol{T}_{n+1}$ easily follows:

$$\boldsymbol{T}^{-1}_{n+1} = \begin{pmatrix} \boldsymbol{\Delta}^{*-1}_n & \boldsymbol{0}_n \\ \boldsymbol{0}'_n & 0 \end{pmatrix} + \frac{1}{1 - \boldsymbol{a}^{*\prime}_n \boldsymbol{\Delta}^{*-1}_n \boldsymbol{a}^*_n} \begin{pmatrix} \boldsymbol{\Delta}^{*-1}_n \boldsymbol{a}^*_n \\ -1 \end{pmatrix} \begin{pmatrix} \boldsymbol{a}^{*\prime}_n \boldsymbol{\Delta}^{*-1}_n & -1 \end{pmatrix}. \tag{6.35}$$

Again, for (6.32) to be valid, $\boldsymbol{\Delta}^*_n$ needs to be nonsingular; for (6.33) to be valid, $\boldsymbol{\Delta}_{n-1}$ needs to be nonsingular; for (6.34) to be valid, $\boldsymbol{\Delta}_{n-1}$ needs to be nonsingular and

$1 + 2\mathrm{i}(u\phi - v) + 2\mathrm{i}(v - u\phi)/n - \boldsymbol{b}'_{n-1}\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{b}_{n-1} \neq 0$; for (6.35) to be valid, $\boldsymbol{\Delta}_n^*$ needs to be nonsingular and $1 - \boldsymbol{a}_n^{*\prime}\boldsymbol{\Delta}_n^{*-1}\boldsymbol{a}_n^* \neq 0$. Appendix A section IV part (i) verifies that these conditions in fact hold.

For us to be able to use (6.8) to evaluate the density function (6.6), we need also work out the derivatives of (6.11), (6.13), (6.18), and (6.22), given by (6.12), (6.14), (6.19), and (6.23), respectively. Essentially, we seek analytical expressions for $\partial\left|\boldsymbol{R}_n\right|/\partial v$, $\partial\left|\boldsymbol{S}_n\right|/\partial v$, $\partial\boldsymbol{S}_n^{-1}/\partial v$, $\partial\left|\boldsymbol{T}_{n+1}\right|/\partial v$, and $\partial\boldsymbol{T}_{n+1}^{-1}/\partial v$. Appendix A section IV part (ii) gives detailed expressions for these derivatives.

6.3 On the Moment and Asymptotic Distribution

Given the density function of $\hat{\kappa} - \kappa$, conditional on $\hat{\phi} > 0$, we can write the moment of $\hat{\kappa}$, if existing, as

$$\mathrm{E}(\hat{\kappa}|\hat{\phi} > 0) = \kappa + \int_{-\infty}^{+\infty} f_{\hat{\kappa}}(w)\mathrm{d}w = -\frac{1}{h}\mathrm{E}(\ln(\hat{\phi})|\hat{\phi} > 0),$$

where

$$\mathrm{E}(\ln(\hat{\phi})|\hat{\phi} > 0) = \frac{1}{1 - F_{\hat{\phi}}(-\phi)}\int_0^{+\infty} \ln(w)\, f_{\hat{\phi}}(w - \phi)\,\mathrm{d}w.$$

Note that $\int_0^{+\infty} \ln(w)\, f_{\hat{\phi}}(w-\phi)\,\mathrm{d}w$ exists if and only if $\int_0^{+\infty} \left|\ln(w)\right| f_{\hat{\phi}}(w-\phi)\,\mathrm{d}w$ exists. For $w \in [1, +\infty)$, $\left|\ln(w)\right| \leq w - 1$, so

$$\begin{aligned}
\int_0^{+\infty} \left|\ln(w)\right| f_{\hat{\phi}}(w-\phi)\,\mathrm{d}w &= \int_0^{1} \left|\ln(w)\right| f_{\hat{\phi}}(w-\phi)\,\mathrm{d}w \\
&\quad + \int_1^{+\infty} \left|\ln(w)\right| f_{\hat{\phi}}(w-\phi)\,\mathrm{d}w \\
&\leq \int_0^{1} \left|\ln(w)\right| f_{\hat{\phi}}(w-\phi)\,\mathrm{d}w \\
&\quad + \int_1^{+\infty} (w-1)\, f_{\hat{\phi}}(w-\phi)\,\mathrm{d}w.
\end{aligned}$$

Since $f_{\hat{\phi}}(w-\phi)$ is a PDF function, it is bounded, say, by a positive constant c, then $\int_0^1 |\ln(w)| f_{\hat{\phi}}(w-\phi)\, \mathrm{d}w \leq -c\int_0^1 \ln(w)\, \mathrm{d}w = c$. Next observe that

$$\begin{aligned}\int_1^{+\infty} (w-1) f_{\hat{\phi}}(w-\phi)\, \mathrm{d}w &= \int_{1-\phi}^{+\infty} (z+\phi-1) f_{\hat{\phi}}(z)\, \mathrm{d}z \\ &= \int_{1-\phi}^{+\infty} z f_{\hat{\phi}}(z)\, \mathrm{d}z + (\phi-1)\int_{1-\phi}^{+\infty} f_{\hat{\phi}}(z)\, \mathrm{d}z,\end{aligned}$$

where the second part $(\phi-1)\int_{1-\phi}^{+\infty} f_{\hat{\phi}}(z)\, \mathrm{d}z$ is bounded, and the first part

$$\int_{1-\phi}^{+\infty} z f_{\hat{\phi}}(z)\, \mathrm{d}z = \int_{-\infty}^{+\infty} z f_{\hat{\phi}}(z)\, \mathrm{d}z - \int_0^{1-\phi} z f_{\hat{\phi}}(z)\, \mathrm{d}z.$$

Again, $\int_0^{1-\phi} z f_{\hat{\phi}}(z)\, \mathrm{d}z \leq c\int_0^{1-\phi} z\mathrm{d}z$, which is bounded, and $\int_{-\infty}^{+\infty} z f_{\hat{\phi}}(z)\, \mathrm{d}z = \mathrm{E}(\hat{\phi}) - \phi$, assuming $\mathrm{E}(\hat{\phi})$ exists. (Keep in mind that $F_{\hat{\phi}}(\cdot)$ and $f_{\hat{\phi}}(\cdot)$ denote the CDF and PDF of $\hat{\phi}-\phi$, respectively.) Thus, existence or not of $\mathrm{E}(\hat{\kappa}|\hat{\phi}>0)$ depends on existence or not of $\mathrm{E}(\hat{\phi})$. When x_0 is random, $\hat{\phi}$ can be written as a ratio of quadratic forms in a normal random vector (see the next subsection), and from Roberts (1995), we can easily verify that $\mathrm{E}(\hat{\phi})$ always exists if $n>1$ or 2 for the AR(1) model without or with intercept. When x_0 is fixed and is not used in formulating the LS estimator $\hat{\phi}$, a very similar argument can show that $\mathrm{E}(\hat{\phi})$ always exists if $n>2$ or 3 for the AR(1) model without or with intercept. Including an extra fixed data point in formulating $\hat{\phi}$ should not affect existence or not of $\mathrm{E}(\hat{\phi})$. Thus, in any interesting case, say, with at least 4 data points, $\mathrm{E}(\hat{\phi})$ always exists, and hence $\mathrm{E}(\hat{\kappa}|\hat{\phi}>0)$ always exists.

Unconditionally, however, $\mathrm{E}(\hat{\kappa})$ is not well defined in the real domain. This is because $\Pr(\hat{\phi}<0) = \Pr(\hat{\phi}-\phi<-\phi) = F_{\hat{\phi}}(-\phi) \neq 0$, and $\hat{\kappa} = -\ln(\hat{\phi})/h$ takes on complex values with a positive probability. Given this observation, one has to be cautions to interpret the approximate bias results developed in the literature. Note that for $0<\phi\leq 1$, $\Pr(\hat{\phi}\leq 0)\to 0$ asymptotically, since $\hat{\phi}$ is consistent. In other words, $\hat{\kappa}$ is always well defined asymptotically, and so is its asymptotic distribution.

In this perspective, we might interpret the approximate moment as the moment of the asymptotic distribution. We summarize the asymptotic distribution of $\hat{\kappa}$ from Zhou and Yu (2010) as follows.[7]

	$\mu = 0$	
$\kappa > 0$	$T \to \infty$ and h fixed	$\sqrt{T}\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{e^{2k\delta}-1}{\delta}\right)$
	$T \to \infty$ and $h \to 0$	$\sqrt{T}\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} \mathrm{N}\left(0, 2\kappa\right)$
	$h \to 0$ and T fixed	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{A_1(\gamma_0,c)}{B_1(\gamma_0,c)}$
$\kappa = 0$	$T \to \infty$ and h fixed	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{\int_0^1 B(r)\mathrm{d}B(r)}{\int_0^1 B^2(r)\mathrm{d}r}$
	$T \to \infty$ and $h \to 0$	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{\int_0^1 B(r)\mathrm{d}B(r)}{\int_0^1 B^2(r)\mathrm{d}r}$
	$h \to 0$ and T fixed	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{A_1(\gamma_0,c)}{B_1(\gamma_0,c)}$

	$\mu \neq 0$	
$\kappa > 0$	$T \to \infty$ and h fixed	$\sqrt{T}\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{e^{2k\delta}-1}{\delta}\right)$
	$T \to \infty$ and $h \to 0$	$\sqrt{T}\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} \mathrm{N}\left(0, 2\kappa\right)$
	$h \to 0$ and T fixed	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{A_2(\gamma_0,c)}{B_2(\gamma_0,c)}$
$\kappa = 0$	$T \to \infty$ and h fixed	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{\int_0^1 B(r)\mathrm{d}B(r) - B(1)\int_0^1 B(r)\mathrm{d}r}{\int_0^1 B^2(r)\mathrm{d}r - \left(\int_0^1 B(r)\mathrm{d}r\right)^2}$
	$T \to \infty$ and $h \to 0$	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{\int_0^1 B(r)\mathrm{d}B(r) - B(1)\int_0^1 B(r)\mathrm{d}r}{\int_0^1 B^2(r)\mathrm{d}r - \left(\int_0^1 B(r)\mathrm{d}r\right)^2}$
	$h \to 0$ and T fixed	$T\left(\hat{\kappa} - \kappa\right) \xrightarrow{d} -\frac{A_2(\gamma_0,c)}{B_2(\gamma_0,c)}$

[7] Note that Zhou and Yu (2010) did not give the expanding and infill asymptotic distribution results when $\kappa = 0$ and $\mu \neq 0$. This corresponds to the scenario, in a discrete framework, when no intercept is present in the true model, but a constant term is included in the regression. The expanding and infill asymptotic distribution results easily follow via the generalized delta method.

where

$$
\begin{aligned}
A_1(\gamma_0, c) &= \gamma_0 \int_0^1 e^{cr} \mathrm{d}B(r) + \int_0^1 J_c(r) \mathrm{d}B(r), \\
B_1(\gamma_0, c) &= \frac{\gamma_0^2 (e^{2c} - 1)}{2c} + 2\gamma_0 \int_0^1 e^{cr} J_c(r) \mathrm{d}B(r) + \int_0^1 J_c^2(r) \mathrm{d}r, \\
A_2(\gamma_0, c) &= \frac{b}{c} \int_0^1 c_1 \mathrm{d}B(r) + \int_0^1 J_c(r) \mathrm{d}B(r) + \gamma_0 \int_0^1 e^{cr} \mathrm{d}B(r) \\
&\quad - \int_0^1 \mathrm{d}B(r) \left(c_2 b + \int_0^1 J_c(r) \mathrm{d}r + c_4 \gamma_0 \right), \\
B_2(\gamma_0, c) &= c_3 b^2 + \frac{2b}{c} \int_0^1 c_1 J_c(r) \mathrm{d}r + \int_0^1 J_c^2(r) \mathrm{d}r + c_4^2 b \gamma_0 + 2\gamma_0 \int_0^1 e^{cr} J_c(r) \mathrm{d}r \\
&\quad + \frac{\gamma_0^2 (e^{2c} - 1)}{2c} - \left(c_2 b + \int_0^1 J_c(r) \mathrm{d}r + c_4 \gamma_0 \right)^2,
\end{aligned}
$$

with $c = -\kappa T$, $c_1 = e^{rc} - 1$, $c_2 = (e^c - c - 1)/c^2$, $c_3 = (e^{2c} - 4e^c + 2c + 3)/(2c^3)$, $c_4 = (e^c - 1)/c$, $b = \mu\sqrt{-c\kappa}/\sigma$, $\gamma_0 = x_0/(\sigma\sqrt{T})$, and $J_c(r) = \int_0^r e^{c(r-s)} \mathrm{d}B(s)$. Note that under the infill asymptotics, the results are conditional on the initial x_0.

6.4 Numerical Results

In this section, we conduct Monte Carlo simulations to illustrate the finite sample performance of our exact distribution in comparison with the "true" distribution and the asymptotic distribution. The data generating process follows the OU model in (6.2), and the error term is generated from normal distribution. Then we adopt the algorithm mentioned in section 2 to compute the exact distribution of LS/ML of κ.

We set $T = 1, 2, 5, 10$, $h = 1/12, 1/52, 1/252$, $\kappa = 0.01, 0.1, 1$, $\mu = 0$, 0.1, $\sigma = 0.1$, $x_0 = \mu$ or $x_0 \sim \mathrm{N}(\mu, \sigma^2/(2\kappa))$. Compared with Zhou and Yu (2010), we have a more comprehensive experiment design, so as to have a better understanding of the finite-sample distributions. For the fixed start-up case ($x_0 = 0$), we also consider $\kappa = 0$. As pointed out in Zhou and Yu (2010), the values of 0.01 and 0.1 for κ are empirically realistic for interest rate data while the value of 1 is empirically realistic for volatility.

In Tables 6.1- 6.8, the "true" distribution results come from 1,000,000 replications, and we make comparison of the exact (p), true (p_{edf}), and asymptotic results under the three asymptotics (p_{exp}, p_{mix}, p_{inf}). In simulating the asymptotic non-normal results, 10,000 replications are used and a sample size of 5,000 is used to approximate the integrals involving the Brownian motion by the discrete Riemann sums.[8] To save space, we report only four tables (each with two panels corresponding to $T = 1, 10$, respectively). Tables 6.1 and 6.4 report the cumulative distributions of $T(\hat{\kappa} - \kappa)$ under a fixed start-up when $\kappa = 0.01$, with $x_0 = \mu = 0, 0.1$, respectively, and Tables 6.5 and 6.8 report the results when x_0 is random.

Several striking features are present in these tables. First, the exact distribution results match to at least the third decimal place with those obtained by 1 million simulations, in all the cases considered. This indicates high accuracy of the exact results calculated by our numerical integration algorithm. In consistent with the asymptotic results in Zhou and Yu (2010), there is no much difference between the results under the expanding and mixed domains, and the infill asymptotics provide relatively better performance. Yet, the asymptotic distribution under the infill domain may still provide poor approximation to the true distribution when the data span is short, especially so in the left tails. While increasing data frequency does not affect much the asymptotic distributions, it does affect the true distribution, and the remarkable performance of the exact distribution is robust to data frequency, as well as to data span and other aspects of model specification.

Second, the true distribution of $\hat{\kappa}$ is highly skewed to the right. Normality is a terrible approximation of the finite-sample distribution of $\hat{\kappa}$. As data frequency or data

[8]Given that the infill asymptotic results are conditional on x_0, in comparison with the exact results with a random start-up, they are calculated as averaging over 2,000 replications, where in each replication, $x_0 \sim \mathrm{N}(\mu, \sigma^2/(2\kappa))$. Also, the discrete AR(1) process is simulated with a sample size of 2,000 when x_0 is random.

span increases, the true distribution tends to exhibit a longer left tail and a shorter right tail. Moreover, we can infer from these tables the exact/true median of $T(\hat{\kappa}-\kappa)$ in all cases are substantially positive. (A direct calculation of the median is also possible, see the paragraph to follow.) This suggests that $\hat{\kappa}$ can significantly over estimate κ in finite samples. This degree of overestimation does not decrease with a higher data frequency (given a fixed data span). This is in line with the observations made by Phillips and Yu (2005) and Tang and Chen (2009). On the other hand, increasing data span might help alleviate this problem, though somewhat marginally.

Third, how the initial observation is spelled out affects significantly the exact distribution of $\hat{\kappa}$. For example, for the fixed start-up case, the distortions of the asymptotic distributions are less severe when $x_0 = 0$ compared with when $x_0 \neq 0$, and the exact distribution is less skewed to the right. This feature is related to the role of initial observation in the unit test literature. It also suggests that the conclusions in Tsui and Ali (1992 , 1994) with x_0 discarded should be examined with more scrutiny. Given the CDF function (6.7) and PDF function (6.8), one might be tempted to calculate the quantile function $F_{\hat{\kappa}}^{-1}(t)$, $t \in [0,1]$ by Newton's method of interpolation. Yet, calculation of the PDF function also involves numerical integration. To reduce computational time, we instead employ a very simple bisection search algorithm. Since it is relatively cheap to simulate the asymptotic results and we have observed that the in-fill asymptotic results are more reliable compared with the expanding and mixed asymptotic results, we start with the t-th empirical quantile of the simulated sample for approximating the in-fill asymptotic results, say c_0. If $F_{\hat{\kappa}}(c_0) < t$, we set c_1 as the $\min\{2t, 1\}$-th empirical quantile of the simulated sample. (Typically, $F_{\hat{\kappa}}(c_1) > t$. If not, one can set c_1 as the $\min\{ct, 1\}$-th empirical quantile of the simulated sample, $c = 3, 4, \cdots$, until one finds $F_{\hat{\kappa}}(c_1) > t$.) If $F_{\hat{\kappa}}(c_0) > t$, we set c_1 as the $t/2$-th empirical quantile of the simulated sample. (Typically, $F_{\hat{\kappa}}(c_1) < t$. If not, one can set c_1 as the ct-th empirical quantile of the simulated sample, $c = 1/3, 1/4, \cdots$, until one finds $F_{\hat{\kappa}}(c_1) < t$.) Given the two initial points c_0 and c_1, a bisection search can then be straightforwardly applied to search numerically for $F_{\hat{\kappa}}^{-1}(t)$. This algorithm is in a similar spirit of the algorithm in Lu and King (2002) . We have calculated some typical percentiles of $T(\hat{\kappa} - \kappa)$ under different scenarios. To save space, they are not reported here but are available upon request.

6.5 Conclusions

We have investigated the exact finite-sample distribution of the estimated mean-reversion parameter in the Ornstein-Uhlenbeck diffusion process. We have considered several different set-ups: known or unknown drift term, fixed or random start-up value, and zero or positive mean-reversion parameter. In particular, we employ numerical integration via analytical evaluation of a joint characteristic function. Our numerical calculations demonstrate the remarkably reliable performance of the exact approach. It is found that the true distribution of the maximum likelihood estimator of the mean-reversion parameter can be severely skewed in finite samples. The asymptotic results under expanding and mixed domains in general perform worse than those under the in-fill domain, though the latter may still perform poorly in the left tails when data spans are short. Our exact approach always provides distribution results of high accuracy, and thus should be used for conducting hypothesis testing and constructing confidence intervals.

Table 6.1: Panel 1.A ($T = 1$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Fixed x_0, $\kappa = 0.01$, $x_0 = \mu = 0$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0002	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
-3	0.0037	0.0036	0.0000	0.0000	0.0013	0.0016	0.0016	0.0000	0.0000	0.0015	0.0013	0.0013	0.0000	0.0000	0.0014
-2	0.0201	0.0202	0.0000	0.0000	0.0109	0.0126	0.0125	0.0000	0.0000	0.0115	0.0110	0.0112	0.0000	0.0000	0.0105
-1.5	0.0487	0.0485	0.0000	0.0000	0.0317	0.0354	0.0352	0.0000	0.0000	0.0323	0.0323	0.0326	0.0000	0.0000	0.0310
-1	0.1126	0.1129	0.0000	0.0000	0.0884	0.0933	0.0926	0.0000	0.0000	0.0898	0.0892	0.0888	0.0000	0.0000	0.0876
-0.8	0.1511	0.1515	0.0000	0.0000	0.1249	0.1305	0.1297	0.0000	0.0000	0.1261	0.1261	0.1253	0.0000	0.0000	0.1245
-0.6	0.1958	0.1963	0.0000	0.0000	0.1679	0.1748	0.1741	0.0000	0.0000	0.1699	0.1691	0.1697	0.0000	0.0000	0.1688
-0.4	0.2439	0.2444	0.0023	0.0023	0.2173	0.2233	0.2226	0.0023	0.0023	0.2191	0.2190	0.2186	0.0023	0.0023	0.2171
-0.2	0.2928	0.2931	0.0787	0.0786	0.2682	0.2732	0.2724	0.0787	0.0786	0.2692	0.2688	0.2686	0.0787	0.0786	0.2682
-0.1	0.3169	0.3174	0.2398	0.2398	0.2933	0.2978	0.2974	0.2398	0.2398	0.2940	0.2946	0.2932	0.2398	0.2398	0.2934
-0.01	0.3380	0.3384	0.4718	0.4718	0.3153	0.3196	0.3193	0.4718	0.4718	0.3156	0.3154	0.3153	0.4718	0.4718	0.3157
-0.001	0.3401	0.3405	0.4972	0.4972	0.3175	0.3218	0.3214	0.4972	0.4972	0.3180	0.3180	0.3175	0.4972	0.4972	0.3179
0	0.3403	0.3408	0.5000	0.5000	0.3177	0.3220	0.3217	0.5000	0.5000	0.3182	0.3183	0.3177	0.5000	0.5000	0.3181
0.001	0.3406	0.3410	0.5028	0.5028	0.3180	0.3223	0.3219	0.5028	0.5028	0.3184	0.3185	0.3180	0.5028	0.5028	0.3184
0.01	0.3427	0.3432	0.5282	0.5282	0.3202	0.3244	0.3240	0.5282	0.5282	0.3206	0.3207	0.3202	0.5282	0.5282	0.3203
0.1	0.3632	0.3638	0.7602	0.7602	0.3411	0.3456	0.3452	0.7602	0.7602	0.3425	0.3421	0.3414	0.7602	0.7602	0.3414
0.2	0.3853	0.3856	0.9213	0.9214	0.3644	0.3685	0.3685	0.9213	0.9214	0.3661	0.3651	0.3644	0.9213	0.9214	0.3649
0.4	0.4271	0.4271	0.9977	0.9977	0.4081	0.4123	0.4123	0.9977	0.9977	0.4111	0.4095	0.4088	0.9977	0.9977	0.4086
0.6	0.4660	0.4660	1.0000	1.0000	0.4492	0.4535	0.4534	1.0000	1.0000	0.4544	0.4512	0.4507	1.0000	1.0000	0.4516
0.8	0.5024	0.5026	1.0000	1.0000	0.4889	0.4922	0.4922	1.0000	1.0000	0.4932	0.4904	0.4899	1.0000	1.0000	0.4907
1	0.5366	0.5372	1.0000	1.0000	0.5255	0.5281	0.5282	1.0000	1.0000	0.5299	0.5267	0.5263	1.0000	1.0000	0.5273
1.5	0.6111	0.6115	1.0000	1.0000	0.6050	0.6059	0.6061	1.0000	1.0000	0.6074	0.6053	0.6051	1.0000	1.0000	0.6055
2	0.6711	0.6713	1.0000	1.0000	0.6700	0.6689	0.6683	1.0000	1.0000	0.6717	0.6691	0.6687	1.0000	1.0000	0.6695
3	0.7592	0.7594	1.0000	1.0000	0.7628	0.7623	0.7619	1.0000	1.0000	0.7655	0.7638	0.7634	1.0000	1.0000	0.7647
5	0.8617	0.8618	1.0000	1.0000	0.8757	0.8716	0.8713	1.0000	1.0000	0.8752	0.8746	0.8742	1.0000	1.0000	0.8755
8	0.9318	0.9318	1.0000	1.0000	0.9490	0.9451	0.9449	1.0000	1.0000	0.9488	0.9485	0.9484	1.0000	1.0000	0.9498
20	0.9906	0.9904	1.0000	1.0000	0.9980	0.9971	0.9970	1.0000	1.0000	0.9981	0.9980	0.9981	1.0000	1.0000	0.9982
50	0.9995	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6.2: Panel 1.B ($T = 10$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Fixed x_0, $\kappa = 0.01$, $x_0 = \mu = 0$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0000	0.0000	0.1319	0.1318	0.0000	0.0000	0.0000	0.1318	0.1318	0.0000	0.0000	0.0000	0.1318	0.1318	0.0000
-3	0.0015	0.0015	0.2513	0.2512	0.0014	0.0014	0.0013	0.2512	0.2512	0.0014	0.0013	0.0013	0.2512	0.2512	0.0014
-2	0.0124	0.0124	0.3274	0.3274	0.0114	0.0118	0.0117	0.3274	0.3274	0.0120	0.0116	0.0115	0.3274	0.3274	0.0120
-1.5	0.0355	0.0356	0.3687	0.3687	0.0347	0.0344	0.0342	0.3687	0.3687	0.0345	0.0341	0.0342	0.3687	0.3687	0.0348
-1	0.0940	0.0938	0.4116	0.4115	0.0924	0.0922	0.0920	0.4115	0.4115	0.0924	0.0918	0.0919	0.4115	0.4115	0.0929
-0.8	0.1310	0.1308	0.4290	0.4290	0.1291	0.1293	0.1292	0.4290	0.4290	0.1292	0.1288	0.1289	0.4290	0.4290	0.1284
-0.6	0.1749	0.1745	0.4467	0.4466	0.1734	0.1730	0.1729	0.4466	0.4466	0.1718	0.1725	0.1729	0.4466	0.4466	0.1715
-0.4	0.2226	0.2221	0.4644	0.4644	0.2207	0.2208	0.2205	0.4644	0.4644	0.2195	0.2194	0.2206	0.4644	0.4644	0.2192
-0.2	0.2715	0.2710	0.4822	0.4822	0.2703	0.2698	0.2695	0.4822	0.4822	0.2683	0.2677	0.2698	0.4822	0.4822	0.2694
-0.1	0.2957	0.2952	0.4911	0.4911	0.2958	0.2940	0.2935	0.4911	0.4911	0.2920	0.2928	0.2941	0.4911	0.4911	0.2935
-0.01	0.3171	0.3166	0.4991	0.4991	0.3169	0.3155	0.3149	0.4991	0.4991	0.3132	0.3152	0.3155	0.4991	0.4991	0.3153
-0.001	0.3192	0.3188	0.4999	0.4999	0.3191	0.3177	0.3171	0.4999	0.4999	0.3157	0.3163	0.3176	0.4999	0.4999	0.3175
0	0.3194	0.3190	0.5000	0.5000	0.3194	0.3179	0.3173	0.5000	0.5000	0.3159	0.3166	0.3178	0.5000	0.5000	0.3177
0.001	0.3197	0.3193	0.5001	0.5001	0.3195	0.3181	0.3176	0.5001	0.5001	0.3161	0.3168	0.3181	0.5001	0.5001	0.3179
0.01	0.3218	0.3213	0.5009	0.5009	0.3216	0.3198	0.3197	0.5009	0.5009	0.3183	0.3190	0.3201	0.5009	0.5009	0.3199
0.1	0.3427	0.3422	0.5089	0.5089	0.3424	0.3412	0.3408	0.5089	0.5089	0.3399	0.3404	0.3410	0.5089	0.5089	0.3415
0.2	0.3654	0.3647	0.5178	0.5178	0.3655	0.3640	0.3637	0.5178	0.5178	0.3626	0.3637	0.3639	0.5178	0.5178	0.3641
0.4	0.4107	0.4082	0.5356	0.5356	0.4101	0.4079	0.4075	0.5356	0.5356	0.4065	0.4077	0.4077	0.5356	0.5356	0.4077
0.6	0.4505	0.4494	0.5533	0.5534	0.4518	0.4495	0.4490	0.5534	0.5534	0.4489	0.4493	0.4493	0.5534	0.5534	0.4500
0.8	0.4892	0.4882	0.5710	0.5710	0.4904	0.4884	0.4879	0.5710	0.5710	0.4868	0.4883	0.4882	0.5710	0.5710	0.4891
1	0.5204	0.5239	0.5884	0.5885	0.5264	0.5245	0.5238	0.5885	0.5885	0.5241	0.5244	0.5244	0.5885	0.5885	0.5252
1.5	0.6030	0.6017	0.6313	0.6313	0.6059	0.6028	0.6019	0.6313	0.6313	0.6038	0.6027	0.6033	0.6313	0.6313	0.6044
2	0.6663	0.6653	0.6726	0.6726	0.6683	0.6664	0.6654	0.6726	0.6726	0.6673	0.6664	0.6674	0.6726	0.6726	0.6681
3	0.7607	0.7602	0.7487	0.7488	0.7636	0.7586	0.7604	0.7488	0.7488	0.7611	0.7616	0.7627	0.7488	0.7488	0.7629
5	0.8720	0.8717	0.8681	0.8682	0.8735	0.8732	0.8727	0.8682	0.8682	0.8741	0.8735	0.8740	0.8682	0.8682	0.8739
8	0.9467	0.9468	0.9631	0.9632	0.9484	0.9481	0.9480	0.9632	0.9632	0.9493	0.9484	0.9487	0.9632	0.9632	0.9489
20	0.9977	0.9977	1.0000	1.0000	0.9981	0.9981	0.9981	1.0000	1.0000	0.9982	0.9982	0.9981	1.0000	1.0000	0.9982
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6.3: Panel 2.A ($T = 1$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Fixed x_0, $\kappa = 0.01$, $x_0 = \mu = 0.1$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
-3	0.0016	0.0016	0.0000	0.0000	0.0003	0.0005	0.0005	0.0000	0.0000	0.0004	0.0004	0.0004	0.0000	0.0000	0.0004
-2	0.0060	0.0060	0.0000	0.0000	0.0021	0.0027	0.0027	0.0000	0.0000	0.0021	0.0020	0.0022	0.0000	0.0000	0.0021
-1.5	0.0114	0.0114	0.0000	0.0000	0.0049	0.0060	0.0060	0.0000	0.0000	0.0049	0.0045	0.0051	0.0000	0.0000	0.0049
-1	0.0209	0.0209	0.0000	0.0000	0.0108	0.0127	0.0127	0.0000	0.0000	0.0105	0.0113	0.0112	0.0000	0.0000	0.0108
-0.8	0.0263	0.0264	0.0000	0.0000	0.0146	0.0169	0.0168	0.0000	0.0000	0.0147	0.0152	0.0151	0.0000	0.0000	0.0144
-0.6	0.0329	0.0330	0.0000	0.0000	0.0194	0.0221	0.0219	0.0000	0.0000	0.0196	0.0188	0.0199	0.0000	0.0000	0.0192
-0.4	0.0408	0.0409	0.0023	0.0023	0.0258	0.0286	0.0287	0.0023	0.0023	0.0260	0.0263	0.0262	0.0023	0.0023	0.0251
-0.2	0.0501	0.0501	0.0787	0.0787	0.0332	0.0365	0.0365	0.0787	0.0787	0.0333	0.0339	0.0339	0.0787	0.0787	0.0324
-0.1	0.0553	0.0554	0.2398	0.2398	0.0373	0.0411	0.0410	0.2398	0.2398	0.0376	0.0379	0.0383	0.2398	0.2398	0.0369
-0.01	0.0603	0.0603	0.4718	0.4718	0.0417	0.0455	0.0454	0.4718	0.4718	0.0420	0.0425	0.0425	0.4718	0.4718	0.0412
-0.001	0.0609	0.0608	0.4972	0.4972	0.0421	0.0459	0.0459	0.4972	0.4972	0.0424	0.0429	0.0430	0.4972	0.4972	0.0418
0	0.0609	0.0609	0.5000	0.5000	0.0422	0.0460	0.0459	0.5000	0.5000	0.0424	0.0430	0.0430	0.5000	0.5000	0.0418
0.001	0.0610	0.0609	0.5028	0.5028	0.0423	0.0460	0.0460	0.5028	0.5028	0.0424	0.0430	0.0431	0.5028	0.5028	0.0419
0.01	0.0615	0.0615	0.5282	0.5282	0.0427	0.0465	0.0464	0.5282	0.5282	0.0428	0.0435	0.0436	0.5282	0.5282	0.0423
0.1	0.0669	0.0669	0.7602	0.7603	0.0474	0.0513	0.0513	0.7602	0.7603	0.0476	0.0484	0.0483	0.7603	0.7603	0.0470
0.2	0.0732	0.0732	0.9213	0.9214	0.0532	0.0571	0.0569	0.9213	0.9214	0.0531	0.0541	0.0539	0.9214	0.9214	0.0525
0.4	0.0871	0.0869	0.9977	0.9977	0.0655	0.0699	0.0698	0.9977	0.9977	0.0651	0.0665	0.0665	0.9977	0.9977	0.0649
0.6	0.1023	0.1021	1.0000	1.0000	0.0798	0.0843	0.0842	1.0000	1.0000	0.0791	0.0809	0.0809	1.0000	1.0000	0.0791
0.8	0.1190	0.1187	1.0000	1.0000	0.0952	0.1003	0.1003	1.0000	1.0000	0.0947	0.0399	0.0967	1.0000	1.0000	0.0950
1	0.1368	0.1366	1.0000	1.0000	0.1130	0.1178	0.1177	1.0000	1.0000	0.1118	0.1145	0.1143	1.0000	1.0000	0.1125
1.5	0.1854	0.1851	1.0000	1.0000	0.1620	0.1672	0.1667	1.0000	1.0000	0.1631	0.1643	0.1641	1.0000	1.0000	0.1635
2	0.2380	0.2380	1.0000	1.0000	0.2187	0.2227	0.2224	1.0000	1.0000	0.2188	0.2207	0.2206	1.0000	1.0000	0.2186
3	0.3477	0.3475	1.0000	1.0000	0.3423	0.3418	0.3408	1.0000	1.0000	0.3430	0.3423	0.3419	1.0000	1.0000	0.3410
5	0.5460	0.5456	1.0000	1.0000	0.5641	0.5582	0.5577	1.0000	1.0000	0.5634	0.5634	0.5629	1.0000	1.0000	0.5658
8	0.7392	0.7386	1.0000	1.0000	0.7840	0.7697	0.7690	1.0000	1.0000	0.7823	0.7792	0.7787	1.0000	1.0000	0.7812
20	0.9562	0.9562	1.0000	1.0000	0.9879	0.9821	0.9821	1.0000	1.0000	0.9885	0.9873	0.9872	1.0000	1.0000	0.9877
50	0.9977	0.9977	1.0000	1.0000	1.0000	0.9999	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6.4: Panel 2.B ($T = 10$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Fixed x_0, $\kappa = 0.01$, $x_0 = \mu = 0.1$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0000	0.0000	0.1319	0.1318	0.0000	0.0000	0.0000	0.1318	0.1318	0.0000	0.0000	0.0000	0.1318	0.1318	0.0000
-3	0.0005	0.0005	0.2513	0.2512	0.0004	0.0004	0.0004	0.2512	0.2512	0.0004	0.0004	0.0004	0.2512	0.2512	0.0004
-2	0.0025	0.0026	0.3274	0.3274	0.0023	0.0023	0.0022	0.3274	0.3274	0.0022	0.0022	0.0022	0.3274	0.3274	0.0024
-1.5	0.0056	0.0058	0.3687	0.3687	0.0052	0.0053	0.0052	0.3687	0.3687	0.0051	0.0052	0.0051	0.3687	0.3687	0.0054
-1	0.0121	0.0124	0.4116	0.4115	0.0115	0.0115	0.0115	0.4115	0.4115	0.0114	0.0114	0.0114	0.4115	0.4115	0.0120
-0.8	0.0160	0.0165	0.4290	0.4290	0.0152	0.0154	0.0153	0.4290	0.4290	0.0155	0.0153	0.0152	0.4290	0.4290	0.0158
-0.6	0.0213	0.0215	0.4467	0.4466	0.0199	0.0204	0.0203	0.4466	0.4466	0.0204	0.0202	0.0203	0.4466	0.4466	0.0210
-0.4	0.0276	0.0279	0.4644	0.4644	0.0260	0.0267	0.0267	0.4644	0.4644	0.0267	0.0257	0.0265	0.4644	0.4644	0.0274
-0.2	0.0354	0.0357	0.4822	0.4822	0.0333	0.0343	0.0343	0.4822	0.4822	0.0341	0.0341	0.0341	0.4822	0.4822	0.0351
-0.1	0.0398	0.0401	0.4911	0.4911	0.0376	0.0387	0.0388	0.4911	0.4911	0.0381	0.0384	0.0385	0.4911	0.4911	0.0393
-0.01	0.0441	0.0445	0.4991	0.4991	0.0420	0.0430	0.0431	0.4991	0.4991	0.0419	0.0425	0.0426	0.4991	0.4991	0.0436
-0.001	0.0446	0.0449	0.4999	0.4999	0.0424	0.0434	0.0435	0.4999	0.4999	0.0424	0.0430	0.0431	0.4999	0.4999	0.0442
0	0.0446	0.0450	0.5000	0.5000	0.0425	0.0435	0.0436	0.5000	0.5000	0.0425	0.0430	0.0432	0.5000	0.5000	0.0443
0.001	0.0447	0.0450	0.5001	0.5001	0.0426	0.0435	0.0437	0.5001	0.5001	0.0425	0.0431	0.0432	0.5001	0.5001	0.0444
0.01	0.0451	0.0455	0.5009	0.5009	0.0430	0.0440	0.0441	0.5009	0.5009	0.0429	0.0434	0.0437	0.5009	0.5009	0.0447
0.1	0.0499	0.0503	0.5089	0.5089	0.0475	0.0485	0.0489	0.5089	0.5089	0.0473	0.0484	0.0484	0.5089	0.5089	0.0492
0.2	0.0555	0.0560	0.5178	0.5178	0.0534	0.0543	0.0545	0.5178	0.5178	0.0534	0.0540	0.0540	0.5178	0.5178	0.0547
0.4	0.0680	0.0685	0.5356	0.5356	0.0662	0.0646	0.0668	0.5356	0.5356	0.0662	0.0664	0.0665	0.5356	0.5356	0.0671
0.6	0.0822	0.0829	0.5533	0.5534	0.0797	0.0793	0.0810	0.5534	0.5534	0.0797	0.0805	0.0807	0.5534	0.5534	0.0819
0.8	0.0979	0.0988	0.5710	0.5710	0.0954	0.0956	0.0965	0.5710	0.5710	0.0950	0.0962	0.0965	0.5710	0.5710	0.0978
1	0.1151	0.1162	0.5884	0.5885	0.1123	0.1138	0.1137	0.5885	0.5885	0.1121	0.1135	0.1137	0.5885	0.5885	0.1148
1.5	0.1640	0.1651	0.6313	0.6313	0.1613	0.1629	0.1630	0.6313	0.6313	0.1620	0.1627	0.1629	0.6313	0.6313	0.1632
2	0.2195	0.2207	0.6726	0.6726	0.2175	0.2188	0.2191	0.6726	0.6726	0.2178	0.2186	0.2190	0.6726	0.6726	0.2172
3	0.3394	0.3406	0.7487	0.7488	0.3393	0.3397	0.3399	0.7488	0.7488	0.3410	0.3398	0.3398	0.7488	0.7488	0.3383
5	0.5585	0.5597	0.8681	0.8682	0.5624	0.5608	0.5611	0.8682	0.8682	0.5616	0.5613	0.5613	0.8682	0.8682	0.5584
8	0.7740	0.7742	0.9631	0.9632	0.7822	0.7780	0.7777	0.9632	0.9632	0.7791	0.7789	0.7796	0.9632	0.9632	0.7781
20	0.9857	0.9856	1.0000	1.0000	0.9884	0.9876	0.9877	1.0000	1.0000	0.9880	0.9880	0.9881	1.0000	1.0000	0.9885
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6.5: Panel 3.A ($T = 1$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Random $x_0 \sim N(\mu, \sigma^2/(2\kappa))$, $\kappa = 0.01$, $\mu = 0$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0000	0.0000	0.0000	0.0000	0.0099	0.0000	0.0000	0.0000	0.0000	0.0092	0.0000	0.0000	0.0000	0.0000	0.0099
-3	0.0001	0.0001	0.0000	0.0000	0.0158	0.0001	0.0001	0.0000	0.0000	0.0146	0.0001	0.0001	0.0000	0.0000	0.0155
-2	0.0008	0.0009	0.0000	0.0000	0.0210	0.0006	0.0006	0.0000	0.0000	0.0196	0.0005	0.0005	0.0000	0.0000	0.0203
-1.5	0.0025	0.0025	0.0000	0.0000	0.0250	0.0020	0.0020	0.0000	0.0000	0.0233	0.0018	0.0018	0.0000	0.0000	0.0242
-1	0.0084	0.0084	0.0000	0.0000	0.0339	0.0073	0.0072	0.0000	0.0000	0.0315	0.0071	0.0070	0.0000	0.0000	0.0328
-0.8	0.0144	0.0145	0.0000	0.0000	0.0423	0.0130	0.0129	0.0000	0.0000	0.0393	0.0127	0.0127	0.0000	0.0000	0.0412
-0.6	0.0264	0.0265	0.0000	0.0000	0.0581	0.0246	0.0243	0.0000	0.0000	0.0550	0.0241	0.0242	0.0000	0.0000	0.0580
-0.4	0.0536	0.0536	0.0023	0.0023	0.0919	0.0512	0.0505	0.0023	0.0023	0.0870	0.0506	0.0508	0.0023	0.0023	0.0892
-0.2	0.1343	0.1340	0.0787	0.0787	0.1715	0.1310	0.1303	0.0787	0.0787	0.1660	0.1302	0.1297	0.0787	0.0787	0.1651
-0.1	0.2430	0.2432	0.2398	0.2398	0.2658	0.2391	0.2387	0.2398	0.2398	0.2623	0.2382	0.2378	0.2398	0.2398	0.2613
-0.01	0.4202	0.4208	0.4718	0.4718	0.4232	0.4161	0.4154	0.4718	0.4718	0.4239	0.4152	0.4145	0.4718	0.4718	0.4227
-0.001	0.4406	0.4411	0.4972	0.4972	0.4427	0.4364	0.4360	0.4972	0.4972	0.4439	0.4355	0.4348	0.4972	0.4972	0.4427
0	0.4428	0.4434	0.5000	0.5000	0.4449	0.4387	0.4382	0.5000	0.5000	0.4461	0.4377	0.4370	0.5000	0.5000	0.4450
0.001	0.4451	0.4456	0.5028	0.5028	0.4471	0.4410	0.4404	0.5028	0.5028	0.4484	0.4400	0.4393	0.5028	0.5028	0.4472
0.01	0.4653	0.4658	0.5282	0.5282	0.4669	0.4612	0.4609	0.5282	0.5282	0.4687	0.4602	0.4596	0.5282	0.5282	0.4676
0.1	0.6361	0.6365	0.7602	0.7603	0.6472	0.6323	0.6319	0.7602	0.7603	0.6541	0.6315	0.6310	0.7603	0.7603	0.6526
0.2	0.7445	0.7454	0.9213	0.9214	0.7733	0.7412	0.7405	0.9213	0.9214	0.7825	0.7404	0.7403	0.9214	0.9214	0.7814
0.4	0.8390	0.8394	0.9977	0.9977	0.8816	0.8365	0.8358	0.9977	0.9977	0.8887	0.8360	0.8360	0.9977	0.9977	0.8822
0.6	0.8803	0.8807	1.0000	1.0000	0.9218	0.8785	0.8781	1.0000	1.0000	0.9263	0.8781	0.8780	1.0000	1.0000	0.9199
0.8	0.9041	0.9045	1.0000	1.0000	0.9414	0.9028	0.9024	1.0000	1.0000	0.9460	0.9025	0.9024	1.0000	1.0000	0.9405
1	0.9200	0.9205	1.0000	1.0000	0.9530	0.9191	0.9188	1.0000	1.0000	0.9575	0.9189	0.9189	1.0000	1.0000	0.9526
1.5	0.9446	0.9451	1.0000	1.0000	0.9684	0.9442	0.9440	1.0000	1.0000	0.9716	0.9441	0.9443	1.0000	1.0000	0.9676
2	0.9590	0.9593	1.0000	1.0000	0.9760	0.9590	0.9589	1.0000	1.0000	0.9784	0.9590	0.9591	1.0000	1.0000	0.9748
3	0.9750	0.9751	1.0000	1.0000	0.9837	0.9755	0.9755	1.0000	1.0000	0.9853	0.9756	0.9757	1.0000	1.0000	0.9826
5	0.9886	0.9886	1.0000	1.0000	0.9901	0.9895	0.9894	1.0000	1.0000	0.9910	0.9897	0.9897	1.0000	1.0000	0.9895
8	0.9954	0.9952	1.0000	1.0000	0.9937	0.9964	0.9964	1.0000	1.0000	0.9943	0.9966	0.9965	1.0000	1.0000	0.9935
20	0.9995	0.9995	1.0000	1.0000	0.9975	0.9999	0.9999	1.0000	1.0000	0.9977	0.9999	0.9999	1.0000	1.0000	0.9974
50	1.0000	1.0000	1.0000	1.0000	0.9990	1.0000	1.0000	1.0000	1.0000	0.9991	1.0000	1.0000	1.0000	1.0000	0.9990
100	1.0000	1.0000	1.0000	1.0000	0.9995	1.0000	1.0000	1.0000	1.0000	0.9995	1.0000	1.0000	1.0000	1.0000	0.9995

Table 6.6: Panel 3.B ($T = 10$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Random $x_0 \sim N(\mu, \sigma^2/(2\kappa))$, $\kappa = 0.01$, $\mu = 0$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0000	0.0000	0.1319	0.1318	0.0265	0.0000	0.0000	0.1318	0.1318	0.0270	0.0000	0.0000	0.1318	0.1318	0.0260
-3	0.0002	0.0002	0.2513	0.2512	0.0418	0.0002	0.0002	0.2512	0.2512	0.0417	0.0002	0.0001	0.2512	0.2512	0.0407
-2	0.0019	0.0019	0.3274	0.3274	0.0548	0.0018	0.0018	0.3274	0.3274	0.0538	0.0019	0.0017	0.3274	0.3274	0.0527
-1.5	0.0063	0.0065	0.3687	0.3687	0.0639	0.0062	0.0062	0.3687	0.3687	0.0623	0.0062	0.0062	0.3687	0.3687	0.0611
-1	0.0235	0.0237	0.4116	0.4115	0.0850	0.0232	0.0231	0.4115	0.4115	0.0824	0.0231	0.0230	0.4115	0.4115	0.0811
-0.8	0.0411	0.0410	0.4290	0.4290	0.1053	0.0406	0.0407	0.4290	0.4290	0.1026	0.0405	0.0406	0.4290	0.4290	0.1013
-0.6	0.0736	0.0731	0.4467	0.4466	0.1432	0.0730	0.0731	0.4466	0.4466	0.1421	0.0729	0.0729	0.4466	0.4466	0.1404
-0.4	0.1342	0.1338	0.4644	0.4644	0.2081	0.1336	0.1334	0.4644	0.4644	0.2091	0.1334	0.1332	0.4644	0.4644	0.2077
-0.2	0.2394	0.2393	0.4822	0.4822	0.2931	0.2387	0.2388	0.4822	0.4822	0.2950	0.2385	0.2385	0.4822	0.4822	0.2944
-0.1	0.3102	0.3084	0.4911	0.4911	0.3441	0.3075	0.3076	0.4911	0.4911	0.3460	0.3073	0.3076	0.4911	0.4911	0.3458
-0.01	0.3760	0.3740	0.4991	0.4991	0.3991	0.3732	0.3732	0.4991	0.4991	0.4013	0.3730	0.3730	0.4991	0.4991	0.4016
-0.001	0.3805	0.3804	0.4999	0.4999	0.4052	0.3797	0.3798	0.4999	0.4999	0.4074	0.3795	0.3797	0.4999	0.4999	0.4077
0	0.3812	0.3812	0.5000	0.5000	0.4059	0.3804	0.3805	0.5000	0.5000	0.4081	0.3803	0.3804	0.5000	0.5000	0.4084
0.001	0.3819	0.3819	0.5001	0.5001	0.4065	0.3812	0.3813	0.5001	0.5001	0.4087	0.3810	0.3811	0.5001	0.5001	0.4091
0.01	0.3885	0.3884	0.5009	0.5009	0.4127	0.3877	0.3877	0.5009	0.5009	0.4149	0.3875	0.3878	0.5009	0.5009	0.4153
0.1	0.4512	0.4510	0.5089	0.5089	0.4779	0.4525	0.4505	0.5089	0.5089	0.4804	0.4502	0.4505	0.5089	0.5089	0.4813
0.2	0.5135	0.5134	0.5178	0.5178	0.5553	0.5148	0.5128	0.5178	0.5178	0.5574	0.5126	0.5132	0.5178	0.5178	0.5588
0.4	0.6113	0.6108	0.5356	0.5356	0.7052	0.6106	0.6108	0.5356	0.5356	0.7075	0.6106	0.6111	0.5356	0.5356	0.7094
0.6	0.6804	0.6803	0.5533	0.5534	0.7949	0.6800	0.6806	0.5534	0.5534	0.7971	0.6799	0.6801	0.5534	0.5534	0.7980
0.8	0.7309	0.7308	0.5710	0.5710	0.8479	0.7306	0.7313	0.5710	0.5710	0.8499	0.7305	0.7311	0.5710	0.5710	0.8509
1	0.7694	0.7693	0.5884	0.5885	0.8805	0.7691	0.7700	0.5885	0.5885	0.8817	0.7691	0.7695	0.5885	0.5885	0.8821
1.5	0.8349	0.8349	0.6313	0.6313	0.9208	0.8348	0.8349	0.6313	0.6313	0.9201	0.8348	0.8349	0.6313	0.6313	0.9206
2	0.8763	0.8759	0.6726	0.6726	0.9393	0.8763	0.8764	0.6726	0.6726	0.9382	0.8763	0.8765	0.6726	0.6726	0.9389
3	0.9249	0.9249	0.7487	0.7488	0.9585	0.9251	0.9253	0.7488	0.7488	0.9571	0.9251	0.9251	0.7488	0.7488	0.9576
5	0.9677	0.9676	0.8681	0.8682	0.9744	0.9680	0.9681	0.8682	0.8682	0.9733	0.9680	0.9680	0.8682	0.8682	0.9738
8	0.9891	0.9890	0.9631	0.9632	0.9838	0.9893	0.9893	0.9632	0.9632	0.9830	0.9894	0.9896	0.9632	0.9632	0.9834
20	0.9997	0.9997	1.0000	1.0000	0.9934	0.9997	0.9998	1.0000	1.0000	0.9931	0.9998	0.9998	1.0000	1.0000	0.9933
50	1.0000	1.0000	1.0000	1.0000	0.9973	1.0000	1.0000	1.0000	1.0000	0.9972	1.0000	1.0000	1.0000	1.0000	0.9973
100	1.0000	1.0000	1.0000	1.0000	0.9987	1.0000	1.0000	1.0000	1.0000	0.9986	1.0000	1.0000	1.0000	1.0000	0.9987

Table 6.7: Panel 4.A ($T = 1$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Random $x_0 \sim N(\mu, \sigma^2/(2\kappa))$, $\kappa = 0.01$, $\mu = 0.1$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
-3	0.0016	0.0016	0.0000	0.0000	0.0003	0.0005	0.0005	0.0000	0.0000	0.0003	0.0004	0.0004	0.0000	0.0000	0.0003
-2	0.0061	0.0060	0.0000	0.0000	0.0021	0.0027	0.0027	0.0000	0.0000	0.0021	0.0020	0.0023	0.0000	0.0000	0.0021
-1.5	0.0114	0.0114	0.0000	0.0000	0.0049	0.0061	0.0060	0.0000	0.0000	0.0049	0.0045	0.0053	0.0000	0.0000	0.0049
-1	0.0210	0.0210	0.0000	0.0000	0.0109	0.0128	0.0129	0.0000	0.0000	0.0109	0.0113	0.0115	0.0000	0.0000	0.0109
-0.8	0.0264	0.0264	0.0000	0.0000	0.0148	0.0170	0.0171	0.0000	0.0000	0.0147	0.0152	0.0155	0.0000	0.0000	0.0147
-0.6	0.0331	0.0330	0.0000	0.0000	0.0197	0.0222	0.0223	0.0000	0.0000	0.0196	0.0189	0.0205	0.0000	0.0000	0.0197
-0.4	0.0410	0.0410	0.0023	0.0023	0.0258	0.0287	0.0287	0.0023	0.0023	0.0258	0.0265	0.0269	0.0023	0.0023	0.0258
-0.2	0.0503	0.0504	0.0787	0.0787	0.0334	0.0367	0.0368	0.0787	0.0787	0.0335	0.0341	0.0344	0.0787	0.0787	0.0335
-0.1	0.0556	0.0556	0.2398	0.2398	0.0378	0.0413	0.0412	0.2398	0.2398	0.0379	0.0384	0.0389	0.2398	0.2398	0.0379
-0.01	0.0606	0.0605	0.4718	0.4718	0.0421	0.0457	0.0457	0.4718	0.4718	0.0422	0.0427	0.0433	0.4718	0.4718	0.0422
-0.001	0.0611	0.0610	0.4972	0.4972	0.0426	0.0462	0.0462	0.4972	0.4972	0.0426	0.0432	0.0437	0.4972	0.4972	0.0426
0	0.0612	0.0611	0.5000	0.5000	0.0426	0.0462	0.0463	0.5000	0.5000	0.0427	0.0434	0.0438	0.5000	0.5000	0.0427
0.001	0.0612	0.0611	0.5028	0.5028	0.0427	0.0463	0.0463	0.5028	0.5028	0.0427	0.0435	0.0438	0.5028	0.5028	0.0427
0.01	0.0618	0.0617	0.5282	0.5282	0.0432	0.0467	0.0468	0.5282	0.5282	0.0432	0.0437	0.0443	0.5282	0.5282	0.0432
0.1	0.0672	0.0671	0.7602	0.7603	0.0479	0.0516	0.0516	0.7602	0.7603	0.0479	0.0487	0.0491	0.7603	0.7603	0.0479
0.2	0.0735	0.0736	0.9213	0.9214	0.0535	0.0574	0.0573	0.9213	0.9214	0.0535	0.0544	0.0548	0.9214	0.9214	0.0536
0.4	0.0874	0.0873	0.9977	0.9977	0.0661	0.0702	0.0700	0.9977	0.9977	0.0662	0.0668	0.0674	0.9977	0.9977	0.0662
0.6	0.1028	0.1026	1.0000	1.0000	0.0805	0.0847	0.0847	1.0000	1.0000	0.0805	0.0813	0.0818	1.0000	1.0000	0.0805
0.8	0.1194	0.1194	1.0000	1.0000	0.0965	0.1008	0.1007	1.0000	1.0000	0.0966	0.0974	0.0979	1.0000	1.0000	0.0966
1	0.1373	0.1374	1.0000	1.0000	0.1141	0.1183	0.1183	1.0000	1.0000	0.1142	0.1150	0.1154	1.0000	1.0000	0.1142
1.5	0.1861	0.1862	1.0000	1.0000	0.1642	0.1678	0.1676	1.0000	1.0000	0.1643	0.1649	0.1653	1.0000	1.0000	0.1643
2	0.2387	0.2388	1.0000	1.0000	0.2210	0.2234	0.2230	1.0000	1.0000	0.2211	0.2214	0.2217	1.0000	1.0000	0.2211
3	0.3485	0.3485	1.0000	1.0000	0.3433	0.3426	0.3422	1.0000	1.0000	0.3434	0.3430	0.3436	1.0000	1.0000	0.3434
5	0.5467	0.5463	1.0000	1.0000	0.5658	0.5589	0.5581	1.0000	1.0000	0.5659	0.5641	0.5648	1.0000	1.0000	0.5659
8	0.7397	0.7395	1.0000	1.0000	0.7825	0.7701	0.7695	1.0000	1.0000	0.7826	0.7796	0.7800	1.0000	1.0000	0.7826
20	0.9562	0.9561	1.0000	1.0000	0.9885	0.9822	0.9823	1.0000	1.0000	0.9885	0.9873	0.9872	1.0000	1.0000	0.9885
50	0.9977	0.9977	1.0000	1.0000	1.0000	0.9999	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6.8: Panel 4.B ($T = 10$): $\Pr(T(\hat{\kappa} - \kappa) \leq w)$, Random $x_0 \sim N(\mu, \sigma^2/(2\kappa))$, $\kappa = 0.01$, $\mu = 0.1$, $\sigma = 0.1$

	Monthly					Weekly					Daily				
w	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}	p	p_{edf}	p_{exp}	p_{mix}	p_{inf}
-5	0.0000	0.0000	0.1319	0.1318	0.0000	0.0000	0.0000	0.1318	0.1318	0.0000	0.0000	0.0000	0.1318	0.1318	0.0000
-3	0.0005	0.0005	0.2513	0.2512	0.0004	0.0004	0.0004	0.2512	0.2512	0.0004	0.0004	0.0004	0.2512	0.2512	0.0004
-2	0.0026	0.0026	0.3274	0.3274	0.0023	0.0024	0.0024	0.3274	0.3274	0.0023	0.0023	0.0024	0.3274	0.3274	0.0023
-1.5	0.0076	0.0060	0.3687	0.3687	0.0054	0.0055	0.0055	0.3687	0.3687	0.0054	0.0055	0.0054	0.3687	0.3687	0.0054
-1	0.0127	0.0129	0.4116	0.4115	0.0118	0.0121	0.0120	0.4115	0.4115	0.0119	0.0120	0.0118	0.4115	0.4115	0.0119
-0.8	0.0168	0.0172	0.4290	0.4290	0.0159	0.0162	0.0161	0.4290	0.4290	0.0159	0.0161	0.0158	0.4290	0.4290	0.0160
-0.6	0.0223	0.0226	0.4467	0.4466	0.0211	0.0215	0.0213	0.4466	0.4466	0.0211	0.0213	0.0210	0.4466	0.4466	0.0212
-0.4	0.0290	0.0292	0.4644	0.4644	0.0276	0.0280	0.0279	0.4644	0.4644	0.0276	0.0271	0.0275	0.4644	0.4644	0.0277
-0.2	0.0371	0.0373	0.4822	0.4822	0.0355	0.0360	0.0358	0.4822	0.4822	0.0356	0.0358	0.0353	0.4822	0.4822	0.0357
-0.1	0.0417	0.0419	0.4911	0.4911	0.0401	0.0406	0.0405	0.4911	0.4911	0.0401	0.0403	0.0398	0.4911	0.4911	0.0402
-0.01	0.0462	0.0465	0.4991	0.4991	0.0445	0.0450	0.0451	0.4991	0.4991	0.0446	0.0335	0.0443	0.4991	0.4991	0.0447
-0.001	0.0467	0.0470	0.4999	0.4999	0.0450	0.0455	0.0455	0.4999	0.4999	0.0451	0.0450	0.0447	0.4999	0.4999	0.0452
0	0.0467	0.0470	0.5000	0.5000	0.0450	0.0456	0.0456	0.5000	0.5000	0.0451	0.0341	0.0448	0.5000	0.5000	0.0452
0.001	0.0468	0.0471	0.5001	0.5001	0.0451	0.0456	0.0456	0.5001	0.5001	0.0452	0.0451	0.0448	0.5001	0.5001	0.0453
0.01	0.0473	0.0475	0.5009	0.5009	0.0456	0.0461	0.0461	0.5009	0.5009	0.0456	0.0455	0.0453	0.5009	0.5009	0.0457
0.1	0.0522	0.0525	0.5089	0.5089	0.0504	0.0508	0.0511	0.5089	0.5089	0.0505	0.0411	0.0501	0.5089	0.5089	0.0506
0.2	0.0580	0.0584	0.5178	0.5178	0.0562	0.0568	0.0571	0.5178	0.5178	0.0563	0.0524	0.0559	0.5178	0.5178	0.0564
0.4	0.0710	0.0712	0.5356	0.5356	0.0691	0.0677	0.0703	0.5356	0.5356	0.0692	0.0799	0.0689	0.5356	0.5356	0.0694
0.6	0.0857	0.0859	0.5533	0.5534	0.0837	0.0828	0.0848	0.5534	0.5534	0.0838	0.0840	0.0834	0.5534	0.5534	0.0840
0.8	0.1019	0.1022	0.5710	0.5710	0.0999	0.0996	0.1010	0.5710	0.5710	0.1000	0.1003	0.0994	0.5710	0.5710	0.1002
1	0.1196	0.1200	0.5884	0.5885	0.1176	0.1183	0.1187	0.5885	0.5885	0.1178	0.1180	0.1172	0.5885	0.5885	0.1180
1.5	0.1697	0.1704	0.6313	0.6313	0.1679	0.1686	0.1689	0.6313	0.6313	0.1681	0.1683	0.1674	0.6313	0.6313	0.1684
2	0.2260	0.2267	0.6726	0.6726	0.2248	0.2253	0.2259	0.6726	0.6726	0.2250	0.2251	0.2244	0.6726	0.6726	0.2253
3	0.3466	0.3468	0.7487	0.7488	0.3468	0.3469	0.3473	0.7488	0.7488	0.3470	0.3470	0.3459	0.7488	0.7488	0.3473
5	0.5649	0.5644	0.8681	0.8682	0.5678	0.5671	0.5678	0.8682	0.8682	0.5680	0.5677	0.5670	0.8682	0.8682	0.5683
8	0.7778	0.7774	0.9631	0.9632	0.7833	0.7818	0.7820	0.9632	0.9632	0.7834	0.7827	0.7818	0.9632	0.9632	0.7835
20	0.9860	0.9859	1.0000	1.0000	0.9885	0.9879	0.9880	1.0000	1.0000	0.9885	0.9883	0.9882	1.0000	1.0000	0.9885
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Chapter 7

On Efficiency Properties of an R-square Coefficient Based on Final Prediction Error

7.1 Introduction

In a recent work, Rousson and Goşoniu (2007) proposed a new version of the coefficient of determination – R^2_{FPE}, which measures the percentage of variation in newly observed dependent variable explained by the fitted model. One of the most exciting advantages of R^2_{FPE} is that it can be used as a model selection criterion which is capable to choose a model with the best prediction ability. Also, the newly proposed R^2_{FPE} can not only overcome the prominent limitation of using R^2 – inflation, but also avoid the problem of selecting a overfitted model with some irrelevant explanatory variables caused by using R^2_a. In addition, as Rousson and Goşoniu (2007) mentioned, R^2_{FPE} and AIC are asymptotically equivalent. The empirical analysis in their paper provided the evidence that using R^2_{FPE} as a model selection criterion is perfectly consistent with using AIC, and is closest with the criterion BIC than R^2 and R^2_a. Thus, R^2_{FPE} can be simultaneously devoted to both aims of goodness-of-fit measure and model selection, which is practical in extensive empirical work. However, Rousson and Goşoniu (2007) didn't study the efficiency properties of R^2_{FPE}. Motivated by the aforementioned facts, the goal of the present chapter is to reveal the bias and MSE properties of this newly proposed goodness-of-fit based on the final prediction error, and compare it with R^2 and R^2_a.

This chapter first shows that the exact bias of R^2_{FPE} is always smaller than that of R^2 and R^2_a, and the variance of R^2_{FPE} is always higher than the other two without specifying the distribution of disturbances. Second, we conduct the analysis of the large-sample asymptotic expansions of the biases and MSEs with $i.i.d$ non-normal disturbances. The large-sample approximate biases give the identical results to the exact forms, that is, the approximate bias of R^2_{FPE} is always smaller than the other two measures. However, the results of the approximate MSEs are more complicated. In normal case, the approximate MSE of R^2_{FPE} is higher than those of R^2 and R^2_a. When the disturbances are non-normally distributed, the superiority of R^2_{FPE} in efficiency will be held under some conditions. These efficiency results developed show that the FPE based R-square is useful for the models which have low values of the population goodness of fit measures (for example in cross section models) or for models with high goodness of fit measures (for example in time series models).

The structure of this chapter is as follows. In the next section, R^2_{FPE} will be introduced and its efficiency property in comparison with R^2, R^2_a will be conducted. In addition, a small numerical analysis is presented as well. The last section is concluding remarks.

7.2 Efficiency Properties of R-Square Coefficients

Let us start with the following linear regression model

$$y = \alpha\iota + X\beta + u \tag{7.1}$$

where the dependent variable y is a $n \times 1$ vector, ι is a $n \times 1$ unity vector, X is a $n \times p$ matrix with n observations and p explanatory variables, regression parameters β is a $p \times 1$ vector, α is a scalar, and the disturbances u has $n \times 1$ dimension with zero mean vector and $\sigma^2 I_n$ variance-covariance matrix.

We have the goodness-of-fit measure R^2 as follows

$$R^2 = \frac{y'MX(X'MX)^{-1}X'My}{y'My} \tag{7.2}$$

where $M = I_n - n^{-1}\iota\iota'$. It is well known that R^2 can also be written in the form of SSR and SST as $R^2 = 1 - SSR/SST$, where $SSR = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, $SST = \sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2$. And the adjusted version of R^2 (R_a^2) can be written in terms of R^2 as $R_a^2 = (1+r)R^2 - r$, $r \equiv \frac{p}{n-p-1} > 0$. Then it is easy to verify that $R_a^2 \leq R^2$, since $R^2 - R_a^2 = r(1 - R^2)$, and $0 \leq R^2 \leq 1, r > 0$.

In the following, I will briefly state how Rousson and Goşoniu (2007) obtained R^2_{FPE}. If one is interested in the prediction ability of a model, it is good to consider the "mean squared prediction error" denoted by $MSPE$. Suppose that $(x_{i1}, x_{i2, \cdots}, x_{ip}, Z_i)$, $i = 1, \cdots, n$, is a new observed sample which has the same sample size with the original one. Notice that Z_i is the "new observed value" and $\hat{Y}_i$ is the predicted value, so the "mean squared prediction error" can be defined as $MSPE = \sum_{i=1}^{n} E(Z_i - \hat{Y}_i)^2/n = \sigma_u^2(n+p+1)/n$. When a model doesn't include any explanatory variable ($p = 0, \sigma_u^2 = \sigma_Y^2$), one can obtain the mean squared prediction error $MSPE_0 = \sigma_Y^2(n+1)/n$. In order to get R^2_{FPE}, Rousson and Gosoniu (2007) use the unbiased estimators FPE

and FPE_0 to substitute $MSPE$ and $MSPE_0$, respectively. Then R^2_{FPE} is given by $R^2_{FPE} = 1 - \frac{MSPE}{MSPE_0} = 1 - \frac{FPE}{FPE_0}$,where $FPE = s^2_u(n+p+1)/n$ and $FPE_0 = s^2_Y(n+1)/n$. Now it is easy to rewrite R^2_{FPE} in terms of SSR and SST, and further in terms of R^2, and R^2_a as follows:

$$\begin{aligned} R^2_{FPE} &= 1 - \frac{SSR}{SST} \cdot \frac{n-1}{n-p-1} \cdot \frac{n+p+1}{n+1} \\ &= \frac{(n+p+1)R^2_a - p}{n+1} \qquad (7.3) \\ &= \frac{(n-1)(n+p+1)R^2 - 2pn}{(n-p-1)(n+1)} \qquad (7.4) \end{aligned}$$

Now one can easily show that $R^2_{FPE} \leq R^2_a \leq R^2$ as Rousson and Gosoniu (2007) pointed out. Examing the formulas of R^2_{FPE},one can obviously see that R^2_{FPE} provides the measure of ability of predicting newly observed sample by using fitted model.

As Cramer (1987) has shown, the 'population' measure of goodness-of-fit has the following form

$$p \lim_{n\to\infty} R^2 = (1 + \frac{\sigma^2}{n^{-1}\beta' X' M X \beta})^{-1} \equiv \theta, \ 0 < \theta \leq 1 \qquad (7.5)$$

and R^2_a has the same probability limit. Thus, we can easily obtain that R^2_{FPE} has the same probability limit with R^2, i.e., $p\lim_{n\to\infty} R^2_{FPE} = p\lim_{n\to\infty} \frac{(n-1)(n+p+1)R^2 - 2pn}{(n-p-1)(n+1)} = \theta$.

In the following, we will examine the efficiency propertities of R^2, R^2_a and R^2_{FPE} without assuming the distribution of disturbances. Since $R^2_{FPE} \leq R^2$ and $R^2_{FPE} \leq R^2_a$, if moments exist, we have

$$ER^2_{FPE} - \theta \leq ER^2 - \theta \Rightarrow B(R^2_{FPE}) \leq B(R^2), \qquad (7.6)$$

$$ER^2_{FPE} - \theta \leq ER^2_a - \theta \Rightarrow B(R^2_{FPE}) \leq B(R^2_a). \qquad (7.7)$$

Thus, the bias of R^2_{FPE} is always smaller than those of R^2 and R^2_a.

Furthermore, from (7.3) and (7.4), it is also straightforward to show

$$V(R^2_{FPE}) \geq V(R^2_a) \geq V(R^2).$$

So the variance of R^2_{FPE} is always higher than the other two measures.

To simplify notation, R^2_{FPE} can be rewritten as follows

$$R^2_{FPE} = R^2_a - \xi(1 - R^2_a) \tag{7.8}$$

where $\xi \equiv p/(n+1), 0 < \xi < 1$. From (7.8), we have

$$(R^2_a - \theta) - (R^2_{FPE} - \theta) = \xi(1 - R^2_a). \tag{7.9}$$

Hence, one can obtain the distance between MSE of R^2_a and that of R^2_{FPE} as follows

$$\begin{aligned} D_1 &= M(R^2_a) - M(R^2_{FPE}) \\ &= \xi[2(1-\theta)E(1-R^2_a) - (2+\xi)E(1-R^2_a)^2]. \end{aligned} \tag{7.10}$$

From (7.10) we have $M(R^2_{FPE}) \leq M(R^2_a)$ provided $D_1 \geq 0$, that is

$$\xi + 2 \leq \frac{2(1-\theta)E(1-R^2_a)}{E(1-R^2_a)^2}. \tag{7.11}$$

Similarly, the difference between $M(R^2)$ and $M(R^2_{FPE})$ can be written as

$$\begin{aligned} D_2 &= M(R^2) - M(R^2_{FPE}) \\ &= \frac{4nr}{n+1}(1-\theta)E(1-R^2) - \frac{4nr(n+1)+4n^2r^2}{(n+1)^2}E(1-R^2)^2 \end{aligned} \tag{7.12}$$

It, therefore, follows that $M(R^2_{FPE}) \leq M(R^2)$ provided $D_2 \geq 0$, that is

$$1 + \frac{nr}{n+1} \leq \frac{(1-\theta)E(1-R^2)}{E(1-R^2)^2}, \tag{7.13}$$

where $E(1-R^2) \geq E(1-R^2)^2$.

It is clear that $B(R^2_{FPE}) \leq B(R^2_a)$, $B(R^2_{FPE}) \leq B(R^2)$, and $V(R^2) \leq V(R^2_a) \leq V(R^2_{FPE})$ from the previous part of the chapter. However, it is not staightforward to compare $M(R^2_{FPE})$ with $M(R^2_a)$ and $M(R^2)$. From (7.11) and (7.13), R^2_{FPE} may perform well in the sense of having relatively lower MSE in some cases, while it may have higher MSE in other cases.

Using lagre-sample approximations, further, we investigate the bias and MSE properties of R^2_{FPE}. We still do not impose restriction on the distribution of disturbances, but assume that the disturbances are *i.i.d* and have first fourth moments as 0, $\sigma^2, \sigma^3\gamma_1$, and $\sigma^4(\gamma_2 + 3)$, where γ_1 and γ_2 are Pearson's measures of skewness and kurtosis, respectively. The large sample asymptotic results are listed in the following theorem, and the derivation is briefly sketched.

Theorem 25 *The large sample asymptotic approximations for the bias of R^2_{FPE} up to order $O(n^{-1})$ is given by*

$$B(R^2_{FPE}) = \frac{(1-\theta)}{n}[-p + \theta(2\theta - 1) + \theta(1-\theta)\gamma_2] \tag{7.14}$$

and the differences among the mean squared errors of three verisons of R-square up to order $O(n^{-2})$ are given by

$$D_1 = M(R^2_a) - M(R^2_{FPE}) = \frac{2p(1-\theta)^2}{n^2}[-\frac{p}{2} + \theta(4\theta - 5) + \theta(1-2\theta)\gamma_2] \tag{7.15}$$

$$D_2 = M(R^2) - M(R^2_{FPE}) = \frac{4p(1-\theta)^2\theta}{n^2}[4\theta - 5 + (1-2\theta)\gamma_2]. \tag{7.16}$$

Proof. Along the lines of Srivastava, Srivastava and Ullah (1995) , the following can be obtained

$$R^2_{FPE} - \theta = (1-\theta)(g_{-1/2} + g_{-1} - \frac{2p}{n}) + O_p(n^{-3/2}).$$

where $g_{-1/2} \equiv \frac{1}{\sigma^2}a_{-1/2}$, $g_{-1} \equiv \frac{1}{\sigma^2}(a_{-1/2} \cdot b_{-1/2} + a_{-1})$, and a_{-r}, b_{-r}, g_{-r} are denoted as

$O_p(n^{-r})$, specifically,

$$\begin{aligned}
a_{-1/2} &= \frac{1}{n}2(1-\theta)u'MX\beta - \theta v \\
a_{-1} &= \frac{1}{n}u'MX(X'MX)^{-1}X'Mu + n^{-2}\theta(u'\iota)^2 \\
b_{-1/2} &= -\frac{1-\theta}{n\sigma^2}(nv + 2u'MX\beta)
\end{aligned}$$

and $v \equiv (\frac{u'u}{n} - \sigma^2)$. By using the method by Ullah (2004, p. 187) , one can obtain $E\left(g_{-1/2}\right)$, $E\left(g_{-1}\right)$, and $E\left(g_{-1/2}^2\right)$. Hence, the bias of R^2_{FPE} up to order $O(n^{-1})$ is given by (7.14). By some algebra, we can have

$$\begin{aligned}
D_1 &= M(R_a^2) - M(R^2_{FPE}) \\
&= \frac{2p(1-\theta)^2}{n}E(g_{-1/2} + g_{-1} - g_{-1/2}^2 - \frac{3p}{2n}) + O(n^{-2}).
\end{aligned}$$

Then (7.15) can hereby be obtained. Similarly, (7.16) can be derived. ■

Remark 1 When $\gamma_2 = 0$, the theorem gives the results for normal case. From the above theorem, we notice that the skewness doesn't affect the bias of R^2_{FPE}, but the kurtosis does. Note that $\partial B(R^2_{FPE})/\partial\gamma_2 = \frac{\theta(1-\theta)^2}{n}$, so $B(R^2_{FPE})$ increases as γ_2 increases. Also, $B(R^2_{FPE})$ is a monotonically decreasing function of p with $\partial B(R^2_{FPE})/\partial p = -\frac{(1-\theta)}{n}$.

Remark 2 Srivastava et al. (1995) derived the large sample asymptotic approximations for $B(R^2)$ and $B(R_a^2)$ as $B(R^2) = \frac{(1-\theta)}{n}[p + \theta(2\theta-1) + \theta(1-\theta)\gamma_2]$, $B(R_a^2) = \frac{\theta(1-\theta)}{n}[(2\theta-1) + (1-\theta)\gamma_2]$. Hence, it is easy to compare the difference between the bias of R^2_{FPE} and those of R^2 and R_a^2.

$$B(R^2) - B(R^2_{FPE}) = \frac{2(1-\theta)p}{n} \geq 0 \tag{7.17}$$

$$B(R_a^2) - B(R^2_{FPE}) = \frac{(1-\theta)p}{n} \geq 0 \tag{7.18}$$

Obviously, the approximate bias $B(R^2_{FPE})$ is always smaller than the other two for the cases with $i.i.d$ disturbances, which is consistent with the results in the exact biases which apply all the cases without assuming the distribution of the disturbances.

Remark 3 From (7.15) and (7.16), one can see that the number of regressors in a model plays a very important role on the sign of D_1. However, it doesn't matter with the sign of D_2. When $\gamma_2 = 0$ (normal case), $D_1 = M(R^2_a) - M(R^2_{FPE}) < 0$, and $D_2 = M(R^2) - M(R^2_{FPE}) < 0$. For $\gamma_2 \neq 0$ (non-normal case), it is not straightforward to see whether $M(R^2_{FPE})$ is lower than $M(R^2)$ and $M(R^2_a)$. We need further numerical analysis.

Remark 4 To see the marginal effect of the number of regressors and kurtosis on the mean square differences in (7.15) and (7.16), we examine the following first partial derivatives.

$$\frac{\partial D_1}{\partial p} = -\frac{2(1-\theta)^2}{n^2}[p + \theta(4\theta - 5) + \theta(1-2\theta)\gamma_2] \tag{7.19}$$

$$\frac{\partial D_1}{\partial \gamma_2} = \frac{2p(1-\theta)^2\theta(1-2\theta)}{n^2} \tag{7.20}$$

$$\frac{\partial D_2}{\partial p} = \frac{4(1-\theta)^2\theta}{n^2}[4\theta - 5 + (1-2\theta)\gamma_2] \tag{7.21}$$

$$\frac{\partial D_2}{\partial \gamma_2} = \frac{4p(1-\theta)^2\theta(1-2\theta)}{n^2} \tag{7.22}$$

From (7.19) and (7.21), obviously, the sign of these derivatives are uncertain. Therefore, we cannot have a clear picture of the impact on mean square differences by adding or removing a regressor. From (7.20) and (7.22), if $\theta > 0.5$, both D_1 and D_2 are decreasing function of γ_2. When $\theta < 0.5$ D_1 and D_2 increases as γ_2 increases.

Table 7.1 on page 164 gives the numerical analysis based on (7.16). The positive sign "+" denotes the relative efficiency gain of R^2_{FPE} over R^2 in terms of approximate MSE, and the negative sign "-" denotes the relative efficiency loss of the former over the latter. For $\gamma_2 = -2$ and θ is high (≥ 0.9), R^2_{FPE} performs better than R^2 in the sense of having lower MSE. Also, for the cases with $\gamma_2 > 0$ and low population fit ($\theta < 0.5$), R^2_{FPE} has lower MSE than R^2 in most cases However, in other cases, the negative sign implies the efficiency loss R^2_{FPE} over R^2.

The results of numerical analysis on comparing the MSE of R^2_{FPE} with that of R^2_a is similar to the above. The difference from before is that the number of regressors (p) plays a role. The details of the results are omitted here. Generally, for large positive kurtosis ($\gamma_2 > 0$) and low population fit ($\theta < 0.5$), or negative kurtosis ($\gamma_2 < 0$) and high population fit ($\theta > 0.9$), R^2_{FPE} performs better than R^2_a in the sense of having lower MSE. For the cases which have zero kurtosis or $0 < \gamma_2 < 10$ and $0.5 \leq \theta < 0.9$, R^2_a tends to be more effcient.

Based on the above analytical results and their calculations we find out that the FPE based goodness-of-fit measures have better efficiency compared to R^2 and R^2_a in terms of having lower MSE. The results suggest that for the models with low values of fits as well as high values the FPE based goodness-of-fit measure is better to use in practice. Since the fit values in many cross section based empirical studies are found to be low (below 0.4) ,and in many time series based empirical studies these values are high (greater than 0.9), we find that FPE based goodness-of-fit measure is useful in both contexts.

7.3 Concluding Remarks

To sum up, the exact bias of R^2_{FPE} is always less than the exact biases of R^2 and R^2_a for all the distributions, and the asymptotic approximations of biases have identical results with those of exact biases. The exact MSE of R^2_{FPE} can behave better only when the conditions in (7.11) and (7.13) are satisfied. However, these conditions are not meaningful since they do depend on the exact moments of R^2 and R^2_a. In view of this we develope the approximate MSE expressions for the MSE of R^2_{FPE}. The efficiency results developed show that the FPE based R-square is useful to consider in the case of cross sectional models with low values of goodness of fit measures as well as for the time series models which tend to have high goodness of fit values. These results, along with the finding of Rousson and Goşoniu (2007) that the FPE based R-square perform well in the model selection, strengthen the usefulness of using this goodness of fit in practice.

Table 7.1: Relative Efficiency Gain/Loss of R^2_{FPE} over R^2

γ_2	$\theta = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
-2	-	-	-	-	-	-	-	-	+	+
0	-	-	-	-	-	-	-	-	-	-
1	-	-	-	-	-	-	-	-	-	-
10	+	+	+	-	-	-	-	-	-	-
20	+	+	+	+	-	-	-	-	-	-
30	+	+	+	+	-	-	-	-	-	-

Chapter 8

Conclusions

This dissertation develops a new set of theoretical results under nonparametric/semiparametric models and continuous time models. Chapters 2-4 discuss our new estimation method and its empirical application within nonparametric and semiparametric framework. Both simulation results and empirical findings show the usefulness of the newly proposed method in practice. Chapter 5-7 are developed within finite sample framework. Chapters 5-6 examine the finite sample properties of the mean reversion parameter estimator ($\widehat{\kappa}$) in continuous time models. Bias approximations of $\widehat{\kappa}$ and its bias corrected estimators are given in chapter 5. The exact distribution of $\widehat{\kappa}$ is evaluated accurately in chapter 6. In chapter 7, we study the efficiency properties of the coefficient of determination (R^2_{FPE}) based on final prediction error.

More specifically, In the second chapter, we propose a two-step estimator of nonparametric regression function with general parametric error covariance for multivariate case and single nonparametric regression. The asymptotic theorem for both mean and slope estimators are established. A small set of Monte Carlo studies shows the relative efficiency gain of the newly proposed estimator in comparison with LLLS and some other two-step estimator in nonparametric regression with either AR(2) errors or heteroskedastic errors. The theoretical results can be widely applied to a general single nonparametric regression analysis.

Chapter 3 systematically develops a new set of results for seemingly unrelated regression (SUR) analysis within nonparametric and semiparametric framework. The properties of LLLS and local linear weighted least squares (LLWLS) estimators in nonparametric SUR are studied as well. To obtain a more efficient estimation, we develop a two-step estimator for the system and establish its asymptotic theorems under both unconditional and conditional error variance-covariance cases. The procedures of estimation for various nonparametric and semiparametric SUR models are proposed, such as, the NP SUR model with error components, partially linear semiparametric model, model with nonparametric autocorrelated errors, additive nonparametric model, varying coefficient model, and the model with endogeneity. These specification have widely practical use in empirical analysis. In addition, two nonparametric goodness-of-fit measures for the system are given. A small set of Monte Carlo simulations shows the relative efficiency gain of the newly developed two-step estimator over LLLS, LLWLS, and a class of two-step estimator.

Chapter 4 presents the practical use of the new methods developed in chapter 2 and 3. We apply nonparametric model and two-step estimation to a real data on return to public capital in U.S. There are some interesting findings in the empirical analysis: First, the average returns of public capital on states' private economic growth are statistically significant and positive. Second, in general, the returns to the public capital are positive. However, a few states, for instances, Wyoming, South Dakota, North Dakota, New Mexico, Montana, have negative returns to the public capital, which are consistent with some recent studies under nonparametric framework. Third, the mean returns to the public capital across all the 48 states changes over the period of 1970-1986. The returns to public capital increased sharply during recessions, started decreasing when the economy stepped into recovering, and fluctuated in small magnitudes during normal time. Note that the last two interesting findings can be only obtained by nonparametric analysis in this real data setting.

The theoretical results in chapters 5-7 are developed within the finite sample framework. Chapter 5 considers the bias of the mean reversion estimator ($\widehat{\kappa}$) in the continuous time Lévy processes. The bias of $\widehat{\kappa}$ is approximated and the bias expressions are obtained for the Lévy-based Ornstein-Uhlenbeck (OU) process. The approximate bias of $\widehat{\kappa}$ under normality is also derived as a special case. The bias expressions indicate that both the skewness and the kurtosis of the Lévy measure affect the bias when the time span is not very large and the sampling frequency is not very high. The initial condition, the long term mean (μ), and the volatility parameter (σ) also enter the bias expressions. A bias corrected estimator of κ is proposed. Monte Carlo studies show the good performance of our newly proposed bias corrected estimators.

It is found that the true distribution of the MLE of κ can be severely skewed in finite samples and that the asymptotic results in general may provide misleading results. In chapter 6, we evaluate the exact distribution of the MLE under different scenarios: known or unknown drift term, fixed or random start-up value, and zero or positive κ. The numerical calculations demonstrate the remarkably reliable performance of our newly proposed exact approach.

Chapter 7 studies the efficiency properties of the coefficient of determination (R^2_{FPE}) based on final prediction error and compares it with conventional goodness-of-fit measures (R^2, R^2_a) in linear regression models with both normal and non-normal disturbances. The theoretical results and a small set of numerical analysis show R^2_{FPE}

is a useful tool as a model selection and goodness-of-fit measure in both cross-sectional analysis and time series analysis.

Appendix A

Mathematical Derivations

I. Derivations in Chapter 2

Proof of Theorem 3

Following MY, we can readily show that $\widehat{\delta}_{SUW,h_2}(x)$ is asymptotically equivalent to the following infeasible estimator

$$\widehat{\delta}_{SUW,h_2}(x) \equiv \left(\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{R}_x^*\right)^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{Z}^* \tag{A.1}$$

where $\mathbf{Z}^* \equiv P^{-1}\mathbf{Y} + \left(H^{-1} - P^{-1}\right)\mathbf{m} = H^{-1}\mathbf{m} + \epsilon^*$. By the second order Taylor expansion around x for elements in $\mathbf{m}$, we have

$$\widehat{\delta}_{SUW,h_2}(x) = \delta(x) + \left(\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{R}_x^*\right)^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\left\{H^{-1}\mathbf{B}_x + \epsilon^*\right\} + o_p\left(h_2^2\right)$$

where $\mathbf{B}_x$ is a $n \times 1$ column vector whose ith element is given by

$$b_{x,i} = \frac{1}{2}\left(X_i - x\right)' m^{(2)}(x)\left(X_i - x\right),$$

and $m^{(2)}(x)$ is the $q\times q$ Hessian matrix of $m(x)$. It follows that

$$\begin{aligned}
&\sqrt{nh_2^q}D_{h_2}\left(\widehat{\delta}_{SUW,h_2}(x)-\delta(x)\right)\\
&=\sqrt{nh_2^q}D_{h_2}\left(\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{R}_x^*\right)^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}H^{-1}\mathbf{B}_x\\
&\quad+\sqrt{nh_2^q}D_{h_2}\left(\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{R}_x^*\right)^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\epsilon^*+o_p(1)\\
&\equiv B_{SUW}+V_{SUW}+o_p(1)\,,\ \text{say,}
\end{aligned}\tag{A.2}$$

where the definitions of the bias term B_{SUW} and the variance term V_{SUW} are self-evident. Note that $E(\epsilon^*\epsilon^{*\prime})=I_{n\times n}$.

To calculate the asymptotic bias, let $S_n\equiv n^{-1}D_{h_2}^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}\mathbf{R}_x^*D_{h_2}^{-1}$. It is easy to show that

$$\begin{aligned}
S_n &= n^{-1}\sum_{i=1}^n\begin{pmatrix}\upsilon_{ii}^2 & \upsilon_{ii}^2\frac{(X_i-x)'}{h_2}\\ \upsilon_{ii}^2\frac{X_i-x}{h_2} & \upsilon_{ii}^2\frac{(X_i-x)(X_i-x)'}{h_2^2}\end{pmatrix}K_{h_2}(X_i-x)\\
&\xrightarrow{p}\begin{pmatrix}\omega_f^*(x,\theta_0) & 0\\ 0 & \omega_f^*(x,\theta_0)\kappa_{21}I_q\end{pmatrix}.
\end{aligned}\tag{A.3}$$

Similarly,

$$\begin{aligned}
&\frac{1}{n}D_{h_2}^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}H^{-1}\mathbf{B}_x\\
&=\frac{1}{n}\begin{pmatrix}\sum_{i=1}^n\upsilon_{ii}^2\mathbf{K}_{x,h_2}b_{x,i}\\ \sum_{i=1}^n\upsilon_{ii}^2\frac{X_i-x}{h_2}\mathbf{K}_{x,h_2}b_{x,i}\end{pmatrix}\\
&=\begin{pmatrix}\frac{\omega_f^*(x,\theta_0)\kappa_{21}h_2^2}{2}\sum_{j=1}^q\frac{\partial^2 m(x)}{\partial x_j^2}\\ \mathbf{0}_{q\times1}\end{pmatrix}+o_p\left(h_2^2\right).
\end{aligned}$$

It follows that $B_{SUW}=\sqrt{nh_2^q}S_n^{-1}\frac{1}{n}D_{h_2}^{-1}\mathbf{R}_x^{*\prime}\mathbf{K}_{x,h_2}H^{-1}\mathbf{B}_x=\begin{pmatrix}\sqrt{nh_2^q}\frac{\kappa_{21}h_2^2}{2}\sum_{j=1}^q\frac{\partial^2 m(x)}{\partial x_j^2}\\ \mathbf{0}_{q\times1}\end{pmatrix}$ $+\,o_p\left(h_2^2\right)$.

Next, by (A.2)-(A.3) we have

$$\begin{aligned} V_{SUW} &= \sqrt{nh_2^q} S_n^{-1} \frac{1}{n} D_{h_2}^{-1} \mathbf{R}_x^{*\prime} \mathbf{K}_{x,h_2} \epsilon^* \\ &= \begin{pmatrix} \omega_f^*(x,\theta_0) & 0 \\ 0 & \omega_f^*(x,\theta_0)\kappa_{21} I_q \end{pmatrix}^{-1} (1+o_p(1)) \\ &\quad \times \sqrt{n^{-1}h_2^q} D_{h_2}^{-1} \sum_{i=1}^{n} \upsilon_{ii} K_{h_2}(X_i - x) \begin{pmatrix} \epsilon_i^* \\ (X_i - x)\epsilon_i^* \end{pmatrix}, \end{aligned}$$

where ϵ_i^* is the ith element of ϵ^*. Applying the Liapounov central limit theorem yields $V_{SUW} \xrightarrow{d} N(0, \Omega_{SUW})$. This completes the proof of the theorem.

II. Derivations in Chapter 3

Proof of Theorem 4

For the LLLS estimator, we can write

$$\begin{aligned} D\left[\hat{\delta}(x) - \delta(x)\right] &= D(Z'(x)K(x)Z(x))^{-1}Z'(x)K(x)\mathbf{B}_x \\ &\quad + D(Z'(x)K(x)Z(x))^{-1}Z'(x)K(x)\mathbf{u} + o_p(1), \end{aligned}$$

where $\mathbf{B}_x = \begin{pmatrix} \mathbf{B}_{x_1}, & \ldots & \mathbf{B}_{x_M} \end{pmatrix}$ is a $NM \times 1$ column vector, $\mathbf{B}_{x_i}$ is a $M \times 1$ column vector whose jth element is given by $b_{x_{i,j}} = \frac{1}{2}(X_{ij} - x_i)' m^{(2)}(x_i)(X_{ij} - x_i)$, and $m^{(2)}(x_i)$ is the $q_i \times q_i$ Hessian matrix of $m(x_i)$. The ith $LLLS$ can be written as

$$\begin{aligned} D_i\left[\hat{\delta}_i(x_i) - \delta_i(x_i)\right] &= D_i(Z_i'(x_i)K(x_i)Z(x_i))^{-1}Z'(x_i)K(x_i)\mathbf{B}_{x_i} \\ &\quad + D_i(Z_i'(x_i)K(x_i)Z(x_i))^{-1}Z'(x_i)K(x_i)\mathbf{u}_i + o_p(1). \end{aligned}$$

The bias of the ith $LLLS$ is $B_{i,LLLS} = S_i^{-1}\frac{1}{N}D_{h_i}^{-1}Z'(x_i)K(x_i)\mathbf{B}_{x_i}$. It can be shown that

$$S_i = N^{-1}\sum_{j=1}^{N}\begin{pmatrix} 1 & \frac{(X_{ij}-x_i)'}{h_i} \\ \frac{X_{ij}-x_i}{h_i} & \frac{(X_{ij}-x_i)(X_{ij}-x_i)'}{h_i^2} \end{pmatrix} K_{h_i}(X_{ij}-x_i)$$

$$\xrightarrow{p} \begin{pmatrix} \bar{f}_i(x_i,\theta_0) & 0 \\ 0 & \bar{f}_i(x_i,\theta_0)\kappa_{21}I_{q_i} \end{pmatrix}. \quad \text{(A.4)}$$

Then we can prove that

$$\frac{1}{N}D_{h_i}^{-1}Z'(x_i)K(x_i)\mathbf{B}_{x_i} = \frac{1}{N}\begin{pmatrix} \sum_{j=1}^{N} K_{h_i}(X_{ij}-x_i)\, b_{x_{i,j}} \\ \sum_{j=1}^{N} \frac{X_{ij}-x_i}{h_i} K_{h_i}(X_{ij}-x_i)\, b_{x_{i,j}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\bar{f}_i(x_i,\theta_0)\kappa_{21}h_i^2}{2}\sum_{s=1}^{q_i}\frac{\partial^2 m_i(x_i)}{\partial x_{i,s}^2} \\ \mathbf{0}_{q_i\times 1} \end{pmatrix} + o_p\left(h_i^2\right).$$

It follows that $B_{i,LLLS} = S_i^{-1}\frac{1}{N}D_{h_i}^{-1}Z'(x_i)K(x_i)\mathbf{B}_{x_i} = \begin{pmatrix} \frac{\kappa_{21}h_i^2}{2}\sum_{s=1}^{q_i}\frac{\partial^2 m_i(x_i)}{\partial x_{i,s}^2} \\ \mathbf{0}_{q_i\times 1} \end{pmatrix} + o_p\left(h_i^2\right).$

Next, we have

$$V_i = \sqrt{Nh_i^{q_i}}S_i^{-1}\frac{1}{N}D_{h_i}^{-1}Z'(x_i)K(x_i)\mathbf{u}_i$$

$$= \frac{1+o_p(1)}{\bar{f}_i(x_i,\theta_0)}\sqrt{Nh_i^{q_i}}\sum_{j=1}^{N}K_{h_i}(X_{ij}-x_i)\begin{pmatrix} \mathbf{u}_i \\ \frac{X_{ij}-x_i}{h_i}\mathbf{u}_i \end{pmatrix}.$$

Then it is easy to obtain $E(V_i)=0$, $E(V_iV_i')=\Omega_{i,LLLS}$, and $E(V_iV_{i'}')=o_p(1)$ which is smaller order than $E(V_iV_i')$. Applying the Liapounov central limit theorem yields $V \xrightarrow{d} N(0,\Omega_{LLLS})$. This completes the proof of the theorem 2.1.

Proof of Theorem 5

Similarly, we can have

$$D\left[\hat{\delta}_{2-step}(x)-\delta(x)\right] = D(R^{*\prime}(x)K(x)R^*(x))^{-1}R^{*\prime}(x)K(x)H^{-1}\mathbf{B}_x$$
$$+D(R^{*\prime}(x)K(x)R^*(x))^{-1}R^{*\prime}(x)K(x)\mathbf{v}+o_p(1).$$

Note that $E(\mathbf{v}\mathbf{v}') = I_{NM\times NM}$. To calculate the asymptotic bias, let

$$S_i^* \equiv N^{-1} D_{h_i}^{-1} R_i^{*\prime}(x_i) K(x_i) R_i^*(x_i) D_{h_i}^{-1}.$$

It is easy to show that

$$\begin{aligned} S_i^* &= N^{-1}\sum_{j=1}^{N}\begin{pmatrix} \upsilon_{(i-1)N+j}^2 & \upsilon_{(i-1)N+j}^2\frac{(X_{ij}-x_i)'}{h_i} \\ \upsilon_{(i-1)N+j}^2\frac{X_{ij}-x_i}{h_i} & \upsilon_{(i-1)N+j}^2\frac{(X_{ij}-x_i)(X_{ij}-x_i)'}{h_i^2}\end{pmatrix} K_{h_i}\left(X_{ij}-x_i\right) \\ &\overset{p}{\to} \begin{pmatrix} \omega_{f,i}^*\left(x,\theta_0\right) & 0 \\ 0 & \omega_{f,i}^*\left(x,\theta_0\right)\kappa_{21} I_{q_i}\end{pmatrix}. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{1}{N} D_{h_i}^{-1} R_i^{*\prime}(x_i) K(x_i) H_i^{-1}\mathbf{B}_{x_i} &= \frac{1}{N}\begin{pmatrix} \sum_{j=1}^{N}\upsilon_{(i-1)N+j}^2 K_{h_i}\left(X_{ij}-x_i\right) b_{x,i} \\ \sum_{j=1}^{N}\upsilon_{(i-1)N+j}^2\frac{X_{ij}-x_i}{h_i} K_{h_i}\left(X_{ij}-x_i\right) b_{x,i}\end{pmatrix} \\ &= \begin{pmatrix} \frac{\omega_{f,i}^*(x,\theta_0)\kappa_{21}h_i^2}{2}\sum\limits_{s=1}^{q_i}\frac{\partial^2 m_i(x_i)}{\partial x_{i,s}^2} \\ \mathbf{0}_{q_i\times 1}\end{pmatrix} + o_p\left(h_i^2\right). \end{aligned}$$

It follows that

$$\begin{aligned} B_{i,2-step} &= S_i^{*-1}\frac{1}{N} D_{h_i}^{-1} R^{*\prime}(x) K(x) H_i^{-1}\mathbf{B}_{x_i} \\ &= \begin{pmatrix} \frac{\kappa_{21}h_i^2}{2}\sum\limits_{s=1}^{q_i}\frac{\partial^2 m_i(x_i)}{\partial x_{i,s}^2} \\ \mathbf{0}_{q_i\times 1}\end{pmatrix} + o_p\left(h_i^2\right). \end{aligned}$$

Next, we have

$$\begin{aligned} V_{i,2-step} &= S_i^{*-1}\frac{1}{N} D_{h_i}^{-1} R_i^{*\prime}(x_i) K(x_i)\mathbf{v}_i \\ &= \frac{1+o_p(1)}{\omega_{f,i}^*\left(x,\theta_0\right)}\sqrt{N^{-1}h_i^{q_i}}\sum_{j=1}^{N}\upsilon_{(i-1)N+j}^2 K_{h_i}\left(X_{ij}-x_i\right)\begin{pmatrix}\mathbf{v}_i \\ \frac{X_{ij}-x_i}{h_i}\mathbf{v}_i\end{pmatrix}, \end{aligned}$$

where $\mathbf{v}_i$ is the ith element of $\mathbf{v}$. Applying the Liapounov central limit theorem yields $V_{2-step}\overset{d}{\to} N\left(0,\Omega_{2-step}\right)$. This completes the proof of the theorem.

III. Derivations in Chapter 5

Proof. The following outlines the proof of Theorem 21 and Theorem 23 in Chapter 5. We have

$$\widehat{\kappa} = -\frac{\ln \hat{\phi}}{h}$$

Then, by higher order taylor expansion

$$\widehat{\kappa} = \kappa - \frac{\hat{\phi}-\phi}{h\phi} + \frac{1}{2h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^2 - \frac{1}{3h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^3 + \frac{1}{4h}\left(\frac{\hat{\phi}-\phi}{\phi}\right)^4 + o_p(n^{-2}h^{-1}).$$

Take expectation on both side, we can have

$$E(\widehat{\kappa}-\kappa) = -\frac{E\left(\hat{\phi}-\phi\right)}{h\phi} + \frac{E\left(\hat{\phi}-\phi\right)^2}{2h\phi^2} - \frac{E\left(\hat{\phi}-\phi\right)^3}{3h\phi^3} + \frac{E\left(\hat{\phi}-\phi\right)^4}{4h\phi^4} + o(n^{-2}h^{-1}).$$

To obtain the bias approximation of $\widehat{\kappa}$, we need derive $E\left(\hat{\phi}-\phi\right)^3$, and $E\left(\hat{\phi}-\phi\right)^4$. The approximate bias $E\left(\hat{\phi}-\phi\right)$ and MSE $E\left(\hat{\phi}-\phi\right)^2$ are given in Bao (2007). The following proof follows Bao (2007), and here we use the similar notation for simplicity. First, we write the pure model as

$$y_t = \phi y_{t-1} + \varepsilon_t$$

and the intercept model as

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t.$$

We write the OLS estimator in matrix form

$$\hat{\phi} = \phi + \frac{y_{-1}' A \varepsilon}{y_{-1}' A y_{t-1}} = \phi + \frac{N}{D}$$

where $y_{-1} = (y_0, y_1, \ldots, y_{n-1})'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$. For pure model

$$y_{-1} = y_D + c\varepsilon = y_0 F + c\varepsilon, \text{ and } A = I, \text{ where } y_D = y_0 F.$$

For Intercept model

$$y_{-1} = y_D + c\varepsilon = y_0 F + \alpha c\iota + c\varepsilon, \text{ and } A = M = I - n^{-1}\iota\iota', \text{ where } y_D = y_0 F + \alpha c\iota.$$

For both models,

$$F = \begin{pmatrix} 1 \\ \phi \\ \vdots \\ \phi^{n-1} \end{pmatrix}, \ c = \begin{pmatrix} 0 & 0 & \dots & \dots & 0 & 0 \\ 1 & 0 & \ddots & \dots & \vdots & \vdots \\ \phi & 1 & 0 & \ddots & \vdots & \vdots \\ \phi^2 & \phi & 1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ \phi^{n-2} & \phi^{n-3} & \dots & \phi & 1 & 0 \end{pmatrix}$$

$$\begin{aligned} N &= r_D'\varepsilon + \varepsilon' Ac\varepsilon, \\ D &= r_D' r_D + 2r_D' c\varepsilon + \varepsilon' c' Ac\varepsilon, \\ r_D &= Ay_D. \end{aligned}$$

By Nagar (1959) expansion, we can have

$$\begin{aligned} \hat{\phi} - \phi &= \frac{N}{D} = \frac{N}{D - E(D) + E(D)} \\ &= \frac{N}{E(D)} \left\{1 + [D - E(D)] E^{-1}(D)\right\}^{-1} \\ &= \frac{N}{E(D)} - \frac{N[D - E(D)]}{E^2(D)} + \frac{N[D - E(D)]^2}{E^3(D)} - \frac{N[D - E(D)]^3}{E^4(D)} + o_p(n^{-2}). \end{aligned}$$

Therefore, we can obtain the first to fourth moments of $\hat{\phi}$ as the followings

$$\begin{aligned} E(\hat{\phi} - \phi) &= 4\frac{E(N)}{E(D)} - 6\frac{E(ND)}{[E(D)]^2} + 4\frac{E(ND^2)}{[E(D)]^3} - \frac{E(ND^3)}{[E(D)]^4} + o(n^{-2}), \\ E[(\hat{\phi} - \phi)^2] &= 6\frac{E(N^2)}{[E(D)]^2} - 8\frac{E(N^2D)}{[E(D)]^3} + 3\frac{E(N^2D^2)}{[E(D)]^4} + o(n^{-2}), \\ E[(\hat{\phi} - \phi^3] &= 4\frac{E(N^3)}{[E(D)]^3} - 3\frac{E(N^3D)}{[E(D)]^4} + o(n^{-2}), \\ E[(\hat{\phi} - \phi)^4] &= \frac{E(N^4)}{[E(D)]^4} + o(n^{-2}). \end{aligned}$$

We need derive $E(N^3)$, $E(N^3D)$, $E(N^4)$, where

$$
\begin{aligned}
N^3 &= (r_D'\varepsilon)^3 + 3(r_D'\varepsilon)^2(\varepsilon'Ac\varepsilon) + 3(r_D'\varepsilon)(\varepsilon'Ac\varepsilon)^2 + (\varepsilon'Ac\varepsilon)^3 \\
N^3D &= (r_D'\varepsilon)^3 r_D' r_D + 2(r_D'\varepsilon)^3(r_D'c\varepsilon) + (r_D'\varepsilon)^3(\varepsilon'c'Ac\varepsilon) + 3(r_D'\varepsilon)^2(\varepsilon'Ac\varepsilon)r_D'r_D \\
&\quad +6(r_D'\varepsilon)^2(\varepsilon'Ac\varepsilon)(r_D'c\varepsilon) + 3(r_D'\varepsilon)^2(\varepsilon'Ac\varepsilon)(\varepsilon'c'Ac\varepsilon) + 3(r_D'\varepsilon)(\varepsilon'Ac\varepsilon)^2 r_D'r_D \\
&\quad +6(r_D'\varepsilon)(\varepsilon'Ac\varepsilon)^2(r_D'c\varepsilon) + 3(r_D'\varepsilon)(\varepsilon'Ac\varepsilon)^2(\varepsilon'c'Ac\varepsilon) \\
&\quad +(\varepsilon'Ac\varepsilon)^3 r_D'r_D + 2(\varepsilon'Ac\varepsilon)^3(r_D'c\varepsilon) + (\varepsilon'Ac\varepsilon)^3(\varepsilon'c'Ac\varepsilon) \\
N^4 &= (r_D'\varepsilon)^4 + 4(r_D'\varepsilon)^3(\varepsilon'Ac\varepsilon) + 6(r_D'\varepsilon)^2(\varepsilon'Ac\varepsilon)^2 + 4(r_D'\varepsilon)(\varepsilon'Ac\varepsilon)^3 + (\varepsilon'Ac\varepsilon)^4
\end{aligned}
$$

Notice that we only need keep the terms of at least $O(n)$ in $E(N^3)$, $O(n^2)$ in $E(N^3D)$ and $E(N^4)$. Using Ullah (2004), Bao and Ullah (2010) and by tedious calculations, for pure model, we have

$$
\begin{aligned}
E(N^3) &= \sigma_0^6\left(\gamma_1^2\beta_3 + 6\rho\beta_2^2\right)n + o(n) \\
E(N^3D) &= \sigma_0^8(12\rho\beta_2^3 + \beta_3\beta_2\gamma_1^2)n^2 + o(n^2) \\
E(N^4) &= 3\sigma_0^8\beta_2^2 n^2 + o(n^2)
\end{aligned}
$$

and

$$
\begin{aligned}
E(D) &= E(r_D'r_D + 2r_D'c\varepsilon + \varepsilon'c'c\varepsilon) = r_D'r_D + 2r_D'cE(\varepsilon) + E(\varepsilon'c'c\varepsilon) \\
&= y_0^2\beta_2 + \sigma_0^2(n\beta_2 - \beta_2^2) + o(n^{-1})
\end{aligned}
$$

where $\beta_i = (1-\rho^i)^{-1}$. And

$$
\begin{aligned}
[E(D)]^{-1} &= \left(n\beta_2\sigma_0^2\right)^{-1}[1 - x + x^2 + \cdots], \\
[E(D)]^{-2} &= \left(n\beta_2\sigma_0^2\right)^{-2}[1 - 2x + 3x^2 + \cdots], \\
[E(D)]^{-3} &= \left(n\beta_2\sigma_0^2\right)^{-3}[1 - 3x + 6x^2 + \cdots], \\
[E(D)]^{-4} &= \left(n\beta_2\sigma_0^2\right)^{-4}[1 - 4x + 10x^2 + \cdots],
\end{aligned}
$$

where $x = \frac{1}{n}[y_0^2/\sigma_0^2 - \beta_2 + o(n^{-1})]$. Since $\varepsilon_i \sim (0,1)$, $\sigma = 1$, $nh = T$, $\phi = e^{-\kappa h}$, $\sigma_0 = \sigma\sqrt{(1-e^{-\kappa h})/2\kappa}$, we have

$$\begin{aligned} E[\left(\hat{\phi}-\phi\right)^3] &= 4\frac{E(N^3)}{[E(D)]^3} - 3\frac{E(N^3D)}{[E(D)]^4} + o(n^{-2}) \\ &= n^{-2}[\beta_2^{-3}(\gamma_1^2\beta_3 - 12\rho\beta_2)] + o(n^{-2}), \\ E[\left(\hat{\phi}-\phi\right)^4] &= \frac{E(N^4)}{[E(D)]^4} + o(n^{-2}) \\ &= 3n^{-2}\beta_2^{-2} + o(n^{-2}). \end{aligned}$$

Substitute the above results into $E(\widehat{\kappa}-\kappa)$, we can have Theorem 21. For Intercept model, we have $y_{-1} = y_D + c\varepsilon = y_0F + \alpha c\iota + c\varepsilon$, $A = M = I - T^{-1}\iota\iota'$, and $r_D = Ay_D = My_0F + M\alpha c\iota = y_0MF + \alpha Mc\iota$. More specifically,

$$y_0MF = y_0\begin{pmatrix} 1-\frac{1}{n}\beta_1(1-\phi^n) \\ \phi-\frac{1}{n}\beta_1(1-\phi^n) \\ \vdots \\ \phi^{n-1}-\frac{1}{n}\beta_1(1-\phi^n) \end{pmatrix},$$

$$\alpha Mc\iota = \alpha\beta_1\begin{pmatrix} -\frac{n-1}{n}+\frac{1}{n}\beta_1(\phi-\phi^n) \\ (1-\phi)-\frac{n-1}{n}+\frac{1}{n}\beta_1(\phi-\phi^n) \\ \vdots \\ (1-\phi^{n-1})-\frac{n-1}{n}+\frac{1}{n}\beta_1(\phi-\phi^n) \end{pmatrix}.$$

Then we can simplify the expression of r_D as

$$r_D = \begin{pmatrix} (y_0-\alpha\beta_1)\phi^0+\lambda \\ (y_0-\alpha\beta_1)\phi+\lambda \\ \vdots \\ (y_0-\alpha\beta_1)\phi^{n-1}+\lambda \end{pmatrix},$$

where

$$
\begin{aligned}
\lambda &= \alpha\beta_1 - \frac{1}{n}\left[y_0\beta_1(1-\phi^n) + \alpha\beta_1(n-1) - \alpha\beta_1^2(\phi-\phi^n)\right] \\
&= -\frac{1}{n}\left[y_0\beta_1(1-\phi^n) - \alpha\beta_1 - \alpha\beta_1^2(\phi-\phi^n)\right] \\
&= O(n^{-1}).
\end{aligned}
$$

Using Ullah (2004), Bao and Ullah (2010) and by tedious calculations, for intercept model, we obtain

$$
\begin{aligned}
E(N^3) &= E[(r_D'\varepsilon)^3] + 3E[(r_D'\varepsilon)^2(\varepsilon' A c\varepsilon)] + 3E[(r_D'\varepsilon)(\varepsilon' A c\varepsilon)^2] + E[(\varepsilon' A c\varepsilon)^3] \\
&= \sigma_0^6\left(\gamma_1^2\beta_3 + 6\phi\beta_2^2 - 3\beta_1\beta_2\right)n + o(n),
\end{aligned}
$$

$$
\begin{aligned}
E(N^3D) &= r_D' r_D E(N^3) \\
&\quad +2E[(r_D'\varepsilon)^3(r_D' c\varepsilon)] + E[(r_D'\varepsilon)^3(\varepsilon' c' A c\varepsilon)] + 6E[(r_D'\varepsilon)^2(\varepsilon' A c\varepsilon)(r_D' c\varepsilon)] \\
&\quad +3E[(r_D'\varepsilon)^2(\varepsilon' A c\varepsilon)(\varepsilon' c' A c\varepsilon)] + 6E[(r_D'\varepsilon)(\varepsilon' A c\varepsilon)^2(r_D' c\varepsilon)] \\
&\quad +3E[(r_D'\varepsilon)(\varepsilon' A c\varepsilon)^2(\varepsilon' c' A c\varepsilon)] \\
&\quad +2E[(\varepsilon' A c\varepsilon)^3(r_D' c\varepsilon)] + E[(\varepsilon' A c\varepsilon)^3(\varepsilon' c' A c\varepsilon)] \\
&= \sigma_0^8(-3\beta_1\beta_2^2 + 12\phi\beta_2^3 + \beta_3\beta_2\gamma_1^2)n^2 + o(n^2),
\end{aligned}
$$

$$
\begin{aligned}
E(N^4) &= E[(r_D'\varepsilon)^4] + 4E[(r_D'\varepsilon)^3(\varepsilon' A c\varepsilon)] \\
&\quad +6E[(r_D'\varepsilon)^2(\varepsilon' A c\varepsilon)^2] + 4E[(r_D'\varepsilon)(\varepsilon' A c\varepsilon)^3] + E[(\varepsilon' A c\varepsilon)^4] \\
&= 12\sigma_0^8\beta_2^2 n^2 + o(n^2),
\end{aligned}
$$

and

$$
\begin{aligned}
E(D) &= r_D' r_D + 2r_D' c E(\varepsilon) + E(\varepsilon' c' A c\varepsilon) \\
&= n\sigma_0^2\beta_2 + \beta_2\left[y_0^2 + \alpha^2\beta_1^2 - 2\alpha\beta_1 y_0 - \sigma_0^2(\phi^2 + 2\phi + 2)\beta_2\right] + o(1).
\end{aligned}
$$

Further, we obtain

$$
\begin{aligned}
E[(\hat{\phi}-\phi)^3] &= 4\frac{E(N^3)}{[E(D)]^3} - 3\frac{E(N^3D)}{[E(D)]^4} + o(n^{-2}) \\
&= n^{-2}[\beta_2^{-3}(\beta_3\gamma_1^2 - 3\beta_1\beta_2 - 12\phi\beta_2^2)] + o(n^{-2}), \\
E[(\hat{\phi}-\phi)^4] &= \frac{E(N^4)}{[E(D)]^4} + o(n^{-2}) \\
&= 12n^{-2}\beta_2^{-2} + o(n^{-2}).
\end{aligned}
$$

Theorem 23 can be obtained by substituting the above results into $E(\widehat{\kappa}-\kappa)$. ■

IV. Derivations in Chapter 6

Derivation Part (i)

This appendix verifies that various conditions for (6.26)–(6.28), (6.31), (6.32)–(6.35) to be valid. For notational convenience, let $a = 2\mathrm{i}(u+v-u\phi)$, $b = -[1+\phi^2+2\mathrm{i}(u\phi-v)]/(\phi+\mathrm{i}u) \neq 0$, $b_1 = (b+\sqrt{b^2-4})/2$, $b_2 = (b-\sqrt{b^2-4})/2$, $c = -(\phi+\mathrm{i}u)$, and $d = \mathrm{i}(u+2v-2u\phi)$.

For (6.28), a necessary condition is $n + a\boldsymbol{\iota}'_{n-1}\boldsymbol{D}_{n-1}^{-1}\boldsymbol{\iota}_{n-1} \neq 0$. Note that $\phi > 0$, so $c \neq 0$, and define $\boldsymbol{D}_n^* = \boldsymbol{D}_n/c$. Then $n + a\boldsymbol{\iota}'_{n-1}\boldsymbol{D}_{n-1}^{-1}\boldsymbol{\iota}_{n-1} = n + (a/c)\boldsymbol{\iota}'_{n-1}\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}$. It can be easily shown that the symmetric $\boldsymbol{D}_n^{*-1}$ has elements

$$
d_n^{*(ij)} = (-1)^{i+j}\left|\boldsymbol{D}_{i-1}^*\right|\left|\boldsymbol{D}_{n-j}^*\right| / \left|\boldsymbol{D}_n^*\right|, \quad i \leq j, \tag{A.5}
$$

where $|\boldsymbol{D}_k^*| = (k+1)(b/2)^k$ when $b = \pm 2$, $(b_1^{k+1} - b_2^{k+1})/(b_1 - b_2)$ when $b \neq \pm 2$. So when $b = 2$ (corresponding to $\phi = -1$ and $v = 0$) or -2 (corresponding to $\phi = 1$ and

$v = 0)$,

$$\begin{aligned}
\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} &= \frac{2\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}(-1)^{i+j}i\left(\frac{b}{2}\right)^{i-1}(n-j)\left(\frac{b}{2}\right)^{n-j-1}}{n(\frac{b}{2})^{n-1}} \\
&\quad+\frac{\sum_{i=1}^{n-1}i\left(\frac{b}{2}\right)^{i-1}(n-i)\left(\frac{b}{2}\right)^{n-i-1}}{n(\frac{b}{2})^{n-1}} \\
&= \begin{cases} \frac{1}{n}\left[\frac{(n-1)^3}{3}-\frac{(n-1)^2}{4}-\frac{n-1}{3}+\frac{1}{8}-\frac{1}{8}(-1)^{n-1}\right] & b=2 \\ \frac{1}{n}\left[-\frac{(n-1)^4}{12}-\frac{2(n-1)^3}{3}+\frac{(n-1)^2}{12}+\frac{2(n-1)}{3}\right] & b=-2 \end{cases} \qquad \text{(A.6)}
\end{aligned}$$

when $b \neq \pm 2$,

$$\begin{aligned}
\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} &= \frac{1}{(b_1-b_2)(b_1^n-b_2^n)} \\
&\quad\cdot\left[2\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\left(b_1^i-b_2^i\right)\left(b_1^{n-j}-b_2^{n-j}\right)+\sum_{i=1}^{n-1}\left(b_1^i-b_2^i\right)\left(b_1^{n-i}-b_2^{n-i}\right)\right] \\
&= \frac{\left(b_1^{n+1}-b_2^{n+1}\right)(n-1)+\left(b_1b_2^n-b_1^nb_2\right)(n+1)}{(b_1-b_2)^2\left(b_1^n-b_2^n\right)} \\
&\quad+\frac{2}{(b_1-b_2)\left(b_1^n-b_2^n\right)(b_1-1)(b_2-1)} \\
&\quad\cdot\left\{(b_1-b_2)\left(\frac{1-{b_1}^{n-1}}{1-b_1}-\frac{1-b_2^{n-1}}{1-b_2}\right)\right. \\
&\quad\quad+(1-b_2)\left[\frac{b_1^nb_2-b_1^2b_2^{n-1}}{b_1-b_2}-b_1^n(n-2)\right] \\
&\quad\quad\left.-(1-b_1)\left[\frac{b_1b_2^n-b_1^{n-1}b_2^2}{b_1-b_2}+b_2^n(n-2)\right]\right\}. \qquad \text{(A.7)}
\end{aligned}$$

For any positive integer n, we can verify that

$$n + a\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{-1}\boldsymbol{\iota}_{n-1} = n + (a/c)\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} \neq 0$$

under the above two cases. The determinant formula (6.25) shows that $\boldsymbol{D}_n$ is always nonsingular, so (6.27) is valid. $|\boldsymbol{D}_n| \neq 0$ and $n + a\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{-1}\boldsymbol{\iota}_{n-1} \neq 0$ ensure that $\boldsymbol{\Delta}_{n-1}$

is nonsingular, and both (6.26) and (6.28) are valid. Further note that

$$\begin{aligned}\boldsymbol{a}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{a}_{n-1} &= -\frac{u}{n^2c}\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}+c\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{n-1,n-1}\\ &\quad +2\frac{\mathrm{i}u}{n}\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}\\ &\quad -\frac{a}{nc^2+ac\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}}\left[c^2\left(\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}\right)^2\right.\\ &\quad -\frac{u^2}{n^2}\left(\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}\right)^2\\ &\quad \left.+2\frac{\mathrm{i}u}{n}c\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}\right], \end{aligned} \tag{A.8}$$

where $\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}$ is given by (A.6) and (A.7), $\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{n-1,n-1}$ is the lower-right element of $\boldsymbol{D}_{n-1}^{*-1}$, and $\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}$ is the sum of the last column of $\boldsymbol{D}_{n-1}^{*-1}$. In particular, given (A.5),

$$\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{n-1,n-1}=\frac{\left|\boldsymbol{D}_{n-2}^{*}\right|}{\left|\boldsymbol{D}_{n-1}^{*}\right|}=\begin{cases}\frac{2(n-1)}{nb} & b=\pm 2\\ \frac{b_1^{n-1}-b_2^{n-1}}{b_1^n-b_2^n} & b\neq\pm 2\end{cases}, \tag{A.9}$$

and

$$\begin{aligned}\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} &= \sum_{i=1}^{n-1}d_{n-1}^{*(i,n-1)}\\ &= \sum_{i=1}^{n-1}(-1)^{1+i}\frac{\left|\boldsymbol{D}_{n-i-1}^{*}\right|}{\left|\boldsymbol{D}_{n-1}^{*}\right|}\\ &= \begin{cases}\frac{1-n}{2} & b=-2\\ \frac{2n+(-1)^n-1}{4n} & b=2\\ \frac{\left[b_1^n+(-1)^nb_1\right](1+b_2)-\left[b_2^n+(-1)^nb_2\right](1+b_1)}{(1+b_1)(1+b_2)(b_1^n-b_2^n)} & b\neq\pm 2\end{cases} \end{aligned} \tag{A.10}$$

By substitution, we can verify that $1-\boldsymbol{a}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{a}_{n-1}\neq 0$ for any positive integer n. This condition, together with a nonsingular $\boldsymbol{\Delta}_{n-1}$, ensures that the inverse formula (6.31) is valid.

With $\boldsymbol{\Delta}_{n-1}$ being nonsingular, (6.33) is valid. Plugging the expression for

$\boldsymbol{\Delta}_{n-1}^{-1}$ from (6.28) leads to

$$\boldsymbol{b}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{b}_{n-1} = \frac{1}{c}\boldsymbol{b}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{b}_{n-1} - \frac{a(\boldsymbol{b}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1})^2}{nc^2 + ac\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}}, \tag{A.11}$$

where

$$\begin{aligned}
\boldsymbol{b}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{b}_{n-1} &= \frac{d^2}{n^2}\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{\boldsymbol{n-1}}^{\boldsymbol{*-1}}\boldsymbol{\iota}_{n-1} + c^2\boldsymbol{e}_{1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{1,n-1} + \frac{2cd}{n}\boldsymbol{e}_{1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}, \\
\boldsymbol{b}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} &= \frac{d}{n}\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} + c\boldsymbol{e}_{1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1},
\end{aligned}$$

in which $\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}$ is given by (A.6) and (A.7), $\boldsymbol{e}_{1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{1,n-1}$ is the top-left element of $\boldsymbol{D}_{n-1}^{*-1}$, and $\boldsymbol{e}_{1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}$ is the sum of the first row of $\boldsymbol{D}_{n-1}^{*-1}$. From (A.5), we see that

$$\begin{aligned}
\boldsymbol{e}_{1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{1,n-1} &= \boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{n-1,n-1}, \\
\boldsymbol{e}_{1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} &= \boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1},
\end{aligned}$$

given by (A.9) and (A.10), respectively. Upon substitution, we can verify that $1+2\mathrm{i}(u\phi-v)+2\mathrm{i}(v-u\phi)/n-\boldsymbol{b}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{b}_{n-1} \neq 0$ for any positive integer n. This condition, together with a nonsingular $\boldsymbol{D}_{n-1}$, ensures that both (6.32) and (6.34) are valid.

Finally, given the definitions of $\boldsymbol{a}_n^*$ and $\boldsymbol{\Delta}_n^*$, we write

$$\boldsymbol{a}_n^{*\prime}\boldsymbol{\Delta}_n^{*-1}\boldsymbol{a}_n^* = \boldsymbol{a}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{a}_{n-1} + \frac{\left(\boldsymbol{a}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{b}_{n-1} - \frac{\mathrm{i}u}{n}\right)^2}{1 + 2\mathrm{i}(u\phi - v) + \frac{2\mathrm{i}(v-u\phi)}{n} - \boldsymbol{b}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{b}_{n-1}}, \tag{A.12}$$

where $\boldsymbol{a}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{a}_{n-1}$ is given by (A.8), $\boldsymbol{b}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{b}_{n-1}$ is given by (A.11), and

$$\begin{aligned}
\boldsymbol{a}_{n-1}'\boldsymbol{\Delta}_{n-1}^{-1}\boldsymbol{b}_{n-1} = & -\frac{u(u+2v-2u\phi)}{n^2c}\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} + c\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{e}_{1,n-1} \\
& + \frac{a}{n}\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1} \\
& - \frac{a}{nc^2 + ac\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}}\Bigg[c^2(\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1})^2 \\
& - \frac{u(u+2v-2u\phi)}{n^2}(\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1})^2 \\
& + \frac{ac}{n}\boldsymbol{e}_{n-1,n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}_{n-1}'\boldsymbol{D}_{n-1}^{*-1}\boldsymbol{\iota}_{n-1}\Bigg].
\end{aligned}$$

Note that $\boldsymbol{\iota}'_{n-1}\boldsymbol{D}^{*-1}_{n-1}\boldsymbol{\iota}_{n-1}$ is given by (A.6) and (A.7), $\boldsymbol{e}'_{n-1,n-1}\boldsymbol{D}^{*-1}_{n-1}\boldsymbol{e}_{n-1,n-1}$ is given by (A.9), $\boldsymbol{e}'_{n-1,n-1}\boldsymbol{D}^{*-1}_{n-1}\boldsymbol{\iota}_{n-1}$ is given by (A.10). By substitution, we can verify that $1-\boldsymbol{a}^{*\prime}_n\boldsymbol{\Delta}^{*-1}_n\boldsymbol{a}^*_n \neq 0$ for any positive integer n. This condition, together with a nonsingular $\boldsymbol{\Delta}^*_n$, ensures that the inverse formula (6.35) is valid.

Derivation Part (ii)

This appendix gives various derivatives that are needed to evaluate the PDF function. From (6.24),

$$\partial\left|\boldsymbol{R}_n\right|/\partial v = \frac{\partial\left|\boldsymbol{D}_{n-1}\right|}{\partial v} - (\phi + \mathrm{i}u)^2\,\frac{\partial\left|\boldsymbol{D}_{n-2}\right|}{\partial v}, \tag{A.13}$$

where $\partial\left|\boldsymbol{D}_n\right|/\partial v$ might be derived analytically from

$$\left|\boldsymbol{D}_n\right| = \prod_{i=1}^{n}\left[1 + \phi^2 + 2\mathrm{i}(u\phi - v) - 2\left(\phi + \mathrm{i}u\right)\cos\left(\pi i/(n+1)\right)\right].$$

A computational less demanding way is to use

$$\begin{aligned}\frac{\partial\left|\boldsymbol{D}_n\right|}{\partial v} &= \left|\boldsymbol{D}_n\right|\operatorname{tr}\left(\boldsymbol{D}_n^{-1}\frac{\partial\boldsymbol{D}_n}{\partial v}\right) \\ &= -2\mathrm{i}\left|\boldsymbol{D}_n\right|\operatorname{tr}\left(\boldsymbol{D}_n^{-1}\right),\end{aligned} \tag{A.14}$$

where $\left|\boldsymbol{D}_n\right|$ is given by (6.25), and from (6.30)

$$\operatorname{tr}\left(\boldsymbol{D}_n^{-1}\right) = -\frac{\sum_{i=1}^{n} U_{i-1}\left(b\right)U_{n-i}\left(b\right)}{(\phi + \mathrm{i}u)U_n\left(b\right)}, \tag{A.15}$$

where $b = -[1 + \phi^2 + 2\mathrm{i}(u\phi - v)]/[2\left(\phi + \mathrm{i}u\right)] \neq 0$, as defined in Appendix A section IV part (i).

Similarly,

$$\frac{\partial\left|\boldsymbol{S}_n\right|}{\partial v} = \left|\boldsymbol{S}_n\right|\operatorname{tr}\left(\boldsymbol{S}_n^{-1}\frac{\partial\boldsymbol{S}_n}{\partial v}\right), \tag{A.16}$$

where $\left|\boldsymbol{S}_n\right|$ is given by (6.26), and we can verify that

$$\frac{\partial\boldsymbol{S}_n}{\partial v} = \begin{pmatrix} -2\mathrm{i}(\boldsymbol{I}_{n-1} - \frac{1}{n}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}'_{n-1}) & \boldsymbol{0}_{n-1} \\ \boldsymbol{0}'_{n-1} & 0 \end{pmatrix}, \tag{A.17}$$

which can be used directly, together with (6.31) and (6.26), in evaluating $\partial\left|\boldsymbol{S}_n\right|/\partial v = \left|\boldsymbol{S}_n\right| \operatorname{tr}\left(\boldsymbol{S}_n^{-1}\partial\boldsymbol{S}_n/\partial v\right)$ and $\partial\boldsymbol{S}_n^{-1}/\partial v = -\boldsymbol{S}_n^{-1}(\partial\boldsymbol{S}_n/\partial v)\boldsymbol{S}_n^{-1}$.

Next, from the definition of $\boldsymbol{T}_{n+1}$, we write

$$\frac{\partial\boldsymbol{T}_{n+1}}{\partial v} = \begin{pmatrix} -2\mathrm{i}+\frac{2\mathrm{i}}{n} & \frac{2\mathrm{i}}{n}\boldsymbol{\iota}_{n-1}' & 0 \\ \frac{2\mathrm{i}}{n}\boldsymbol{\iota}_{n-1} & -2\mathrm{i}(\boldsymbol{I}_{n-1}-\frac{1}{n}\boldsymbol{\iota}_{n-1}\boldsymbol{\iota}_{n-1}') & \mathbf{0}_{n-1} \\ 0 & \mathbf{0}_{n-1}' & 0 \end{pmatrix}, \tag{A.18}$$

which can be used directly, together with (6.35) and (6.32), in evaluating $\partial\left|\boldsymbol{T}_{n+1}\right|/\partial v = \left|\boldsymbol{T}_{n+1}\right| \operatorname{tr}\left(\boldsymbol{T}_{n+1}^{-1}\partial\boldsymbol{T}_{n+1}/\partial v\right)$ and $\partial\boldsymbol{T}_{n+1}^{-1}/\partial v = -\boldsymbol{T}_{n+1}^{-1}(\partial\boldsymbol{T}_{n+1}/\partial v)\boldsymbol{T}_{n+1}^{-1}$.

Derivation Part (iii)

Given the characteristic functions (6.11), (6.13), (6.18), (6.22), we need to implement numerical integration to calculate (6.5) via (6.7). This can be straightforwardly implemented using Matlab's `quadgk` command. One caveat to note is that the square root function in the complex domain is not continuous. One choice is to follow Perron (1989) to identify explicitly the discontinuous points by grid search and then integrate by parts. The search, however, might be inefficient and time-consuming. Instead, we use the following algorithm so that the integrand function for `quadgk` is always continuous. Let $g(t) = \sqrt{a(t)+\mathrm{i}b(t)}$ denote the integrand function in question with $t \in [l,u]$. `quadgk` requires the integrand function to accept a vector $(t_1, t_2, \cdots, t_n)$ and returns a vector of output. Let $\theta_i = \arg\left(a(t_i)+\mathrm{i}b(t_i)\right) \in [-\pi,\pi]$ and denote $a_i = a(t_i)$, $b_i = b(t_i)$, and $g_i = g(t_i)$.

1. Start with t_1 and set $g_1 = \texttt{sqrt}(a_1+\mathrm{i}b_1)$. Set $k=0$.

2. Beginning with t_2, if $a_i < 0$, $a_{i-1} < 0$, and $b_i b_{i-1} <= 0$, set $k = k+1$; otherwise, k is unchanged. Set $g_i = \sqrt{a_i^2+b_i^2}\left(\cos\left(\theta_i^*/2\right)+\mathrm{i}\sin\left(\theta_i^*/2\right)\right)$, where $\theta_i^* = \theta_i + 2k\pi$.

Bibliography

[1] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusion: A closed-form approximation approach. *Econometrica*, 70:223–262, 2002.

[2] M.M. Ali. Distribution of the least squares estimator in a first-order autoregressive model. *Econometric Reviews*, 21:89–119, 2002.

[3] Y. At-Sahalia and J. Jacod. Fisher's information for discretely sampled lévy processes. *Econometrica*, 76:727–761, 2008.

[4] R. B. Avery. Error components and seemingly unrelated regressions. *Econometrica*, 45, 1:199–208, 1977.

[5] B. H. Baltagi and N. Pinnoi. Public capital stock and state productivity growth: Further evidence from an error components model. *Empirical Economics*, 20:351–359, 1995.

[6] F.M. Bandi and P.C.B. Phillips. Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71:241–283, 2003.

[7] F.M. Bandi and P.C.B. Phillips. A simple approach to the parametric estimation of potentially nonstationary diffusions. *Journal of Econometrics*, 137:354–395, 2007.

[8] Y. Bao. The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. *Econometric Theory*, 23:1013–1021, 2007.

[9] Y. Bao and A. Ullah. The second-order and mean squared error of estimators in time-series models. *Journal of Econometrics*, 140:650–669, 2007.

[10] Y. Bao and A. Ullah. Expectation of quadratic forms in normal and nonnormal variable with econometric applications. *Journal of Statistical Planning and Inference*, 140:1193–1205, 2010.

[11] O.E. Barndorff-Nielsen. Processes of normal inverse gaussian type. *Finance and Stochastics*, 2:41–68, 1998.

[12] D. S. Berstein. *Matrix Mathematics: Theory, Facts, and Formulas.* Princeton University Press, New Jersey, 2009.

[13] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economics*, 81:637–654, 1973.

[14] B.W. Brown and J.T.Hewitt. Asymptotic likelihood theory for diffusion processes. *Journal of Applied Probability*, 12:228–238, 1975.

[15] P. Carr and L. Wu. The finite moment log stable process and option pricing. *Journal of Finance*, 58:753–778, 2003.

[16] J. Cox, J. Ingersoll, and S. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.

[17] J.S. Cramer. Mean and variance of r^2 in small and moderate samples. *Journal of Econometrics*, 35:253–266, 1987.

[18] C.M. da Fonseca and J. Petronilho. Explicit inverses of some tridiagonal matrices. *Linear Algebra and its Applications*, 325:7–21, 2001.

[19] M. Das. *Theory and Empirical Evidence on the Ecology Versus Ecocnomy Debate: A Nonparametric Estimation of the SURE Model and an Application to a General Equilibrium Model of Trade and Environment.* Phd thesis, University of California, Riverside, 2005.

[20] R.B. Davies. Numerical inversion of a characteristic function. *Biomeirika*, 60:416, 1973.

[21] R.B. Davies. Algorithm as 155: The distribution of a linear combination of χ^2 random variables. *The American Statistician*, 29:323–333, 1980.

[22] P. Evans and G. Karras. Are government activities productive? evidence from a panel of u.s. states. *Review of Economics and Statistics*, 76, 1:1–11, 1994.

[23] J. Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38:481–482, 1951.

[24] W. H. Greene. *Econometric Analysis.* Prentice Hall, New Jersey, 2007.

[25] J. Gurland. Inversion formula for the distribution of ratios. *Annals of Mathematical Statistics*, 19:228–237, 1948.

[26] F. Hayashi. *Econometrics.* Princeton University Press, Princeton, New Jersey, 2000.

[27] D. Henderson and A. Ullah. Nonparametric random effect estimator. *Economics Letters*, 88:403–407, 2005.

[28] D. Henderson and A. Ullah. Nonparametric estimation in a one-way error component model: A monte carlo analysis. Working paper, Economics Department, University of California, Riverside, 2008.

[29] D. Holtz-Eakin. Public-sector capital and the productivity puzzle. *Review of Economics and Statistics*, 76:12–21, 1994.

[30] Y. Hong and H. Li. Nonparametric specification testing for continuous time model with application to spot interest rates. *Review of Financial Studies*, 18:37–84, 2005.

[31] G.Y. Hu and R.F. O'Connell. Analytical inversion of symmetric tridiagonal matrices. *ournal of Physics A: Mathematical and General*, 29:1511–1513, 1996.

[32] L. Hurwicz. Least squares bias in time series. In T.C. Koopmans, editor, *Statistical Inference in Dynamic Models*, pages 365–383. Wiley, New York, 1950.

[33] J.P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48:419–426, 1961.

[34] N. Kakwani. The unbiasedness of zellner's seemingly unrelated regression equations estimators. *Journal of the American Statistical Association*, 62:141–142, 1967.

[35] M.G. Kendall. Note on bias in the estimation of autocorrelation. *Biometrika*, 41:403–404, 1954.

[36] G. Koop, D. Poier, and J. Tobias. Semiparametric bayesian inference in multiple equation models. *Journal of Applied Econometrics*, 20,6:723–747, 2005.

[37] Q. Li and J. S. Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton, New Jersey, 2007.

[38] X. Lin and R. J. Carroll. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95:520–534, 2000.

[39] O. B. Linton and E. Mammen. Nonparametric transformation to white noise. *Journal of Econometrics*, 142:241–264, 2008.

[40] Z.-H. Lu. The numerical evaluation of the probability density function of a quadratic form in normal variables. *Computational Statistics and Data Analysis*, 51:1986–1996, 2006.

[41] Z.-H. Lu and M.L. King. Improving the numerical technique for computing the accumulated distribution of a quadratic form in normal variables. *Econometric Reviews*, 21:149–165, 2002.

[42] D. Madan, P. Carr P, and E. Chang. The variance gamma processes and option pricing. *European Finance Review*, 2:79–105, 1999.

[43] D. Madan and E. Seneta. The variance gamma (v.g.) model for share market returns. *Journal of Business*, 63:511–524, 1990.

[44] C. Martins-Filho and F. Yao. Nonparametric regression estimation with general parametric error covariance. *Journal of Multivariate Analysis*, 100:309–333, 2009.

[45] R.C. Merton. Optimum consumption and portfolio rules in a continuous time model. *Journal of Economic Theory*, 3:373–413, 1971.

[46] T. Muir. *Exercise on determinants, in Mathematical Questions and Solutions*. Hodggon, London, 1884.

[47] A. H. Munnell. How does public infrastructure affect regional economic performance? *New England Economic Review*, pages 11–32, 1990.

[48] A.L. Nagar. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica*, 27:575–595, 1959.

[49] A. Pagan and A. Ullah. *Nonparametric Econometrics.* Cambridge University Press, New York, 1999.

[50] P. Perron. The calculation of the limiting distribution of the least-squares estimator in a near-integrated model. *Econometric Theory*, 5:241–255, 1989.

[51] P.C.B. Phillips and J. Yu. Jackknifing bond option prices. *Review of Financial Studies*, 18:707–742, 2005.

[52] P.C.B. Phillips and J. Yu. Simulation-based estimation of contingent-claims prices. *Review of Financial Studies*, 22:3669–3705, 2009a.

[53] P.C.B. Phillips and J. Yu. A two-stage realized volatility approach to estimation of diffusion processes with discrete data. *Journal of Econometrics*, 150:139–150, 2009b.

[54] P.C.B. Phillips and J. Yu. Maximum likelihood and gaussian estimation of continuous time models in finance. In T. Mikosch, J. Kreiβ, R. A. Davis, and T. G. Andersen, editors, *Handbook of Financial Time Series*, pages 497–530. Springer-Verlag, New York, 2009c.

[55] P. Protter. *Stochastic Integration and Differential Equations: A New Approach.* Springer, Heidelberg, 1990.

[56] L.A. Roberts. On the existence of moments of ratios of quadratic forms. *Econometric Theory*, 11:750–774, 1995.

[57] P. M. Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56, 4:931–954, 1988.

[58] V. Rousson and N.F. Goşoniu. An r-square coefficient based on final prediction error. *Statistical Methodology*, 4:331–340, 2007.

[59] A. F. Ruckstuhl, A. H. Welsh, and R. J. Carroll. Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica*, 10:51–71, 2000.

[60] K. Singleton. Estimation of affine pricing models using the empirical characteristic function. *Journal of Econometrics*, 102:111–141, 2001.

[61] M. Smith and R. Kohn. Nonparametric seemingly unrelated regression. *Journal of Econometrics*, 98:257–281, 2000.

[62] A.K. Srivastava, V.K. Srivastava, and A. Ullah. The coefficient of determination and its adjusted version in linear regression models. *Econometric Reviews*, 14(2):229–240, 1995.

[63] L. Su, I. Murtazashvili, and A. Ullah. Local linear gmm estimation of functional coefficient iv models with application to the estimation of rate of return to schooling. Working paper, Unversity of California, Riverside, 2011.

[64] L. Su and A. Ullah. More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. *Econometric Theory*, 22:98–126, 2006.

[65] L. Su and A. Ullah. More efficient estimation of nonparametric panel data models with random effects. *Economics Letters*, 96:375–380, 2007.

[66] L. Su, A. Ullah, and Y. Wang. Nonparametric regression estimation with general parametric error covariance: A more efficient two-step estimator. Working paper, Unversity of California, Riverside, 2011.

[67] S.M. Sundaresan. Continuous-time methods in finance: A review and an assessment. *Journal of Finance*, 55:1569–1622, 2000.

[68] C.Y. Tang and S.X. Chen. Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149:65–81, 2009.

[69] W. Taylor. Small sample properties of a class of two stage aitken estimators. *Econometrica*, 45:497–508, 1977.

[70] A.K. Tsui and M.M. Ali. Approximations to the distribution of the least squares estimator in a first-order stationary autoregressive model. *Communications in Statistics-Simulation*, 21:463–484, 1992.

[71] A.K. Tsui and M.M. Ali. Exact distributions, density functions and moments of the least squares estimator in a first-order autoregressive model. *Computational Statistics and Data Analysis*, 17:433–454, 1994.

[72] A. Ullah. *Finite Sample Econometrics*. Oxford University Press, New York, 2004.

[73] A. Ullah and N. Roy. Nonparametric and semiparametric econometrics of panel data. In A. Ullah and D. E. A. Giles, editors, *Handbook of Applied Economics Statistics*. Marcel Dekker, New York, 1998.

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

[75] Y. D. Wang, W. S. Guo, and B. Brown. Spline smoothing for bivariate data with application between hormones. *Statistica Sinica*, 10, 2:377–397, 2000.

[76] A. H. Welsh and T. W. Yee. Local regression for vector responses. *Journal of Statistical Planning and Inference*, 136:3007–3031, 2006.

[77] Z. Xiao, O. B. Linton, R. J. Carroll, and E. Mammen. More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of American Statistical Association*, 98:980–992, 2003.

[78] L. Yang, W. Härdle, and J. Nielsen. Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis*, 20:579–604, 1999.

[79] F. Yao and A. Ullah. A nonparametric r^2 test for the presence of relevant variables. Working paper, Economics, Department, West Virginia University, 2011.

[80] J. Yu. Bias in the estimation of the mean reversion parameter in continuous time models. Working paper, Singapore Management University, 2009.

[81] J. Yu. Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics*, forthcoming, 2011.

[82] A. Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57, 298:348–368, 1962.

[83] Q. Zhou and J. Yu. Asymptotic distributions of the least square estimator for diffusion processes. Working paper, Singapore Management University, 2010.