2012

Three Econometric Essays on Continuous Time Models

Xiaohu WANG
Singapore Management University, xiaohu.wang.2008@smu.edu.sg

Three Econometric Essays on Continuous Time Models

XIAOHU WANG

SINGAPORE MANAGEMENT UNIVERSITY

2012

Three Econometric Essays on Continuous Time Models

by
Xiaohu Wang

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Peter C.B. Phillips (Supervisor/Co-Chair)
Sterling Professor of Economics and Statistics
Yale University
Distinguished Term Professor of Economics
Singapore Management University

Jun Yu (Supervisor/Co-Chair)
Professor of Economics and Finance
Singapore Management University

Sainan Jin
Associate Professor of Economics
Singapore Management University

Kian Guan Lim
Professor of Finance
Singapore Management University

Singapore Management University
2012

Abstract

Three Econometric Essays on Continuous Time Models

Xiaohu Wang

Multivariate continuous time models are now widely used in economics and finance. Empirical applications typically rely on some process of discretization so that the system may be estimated with discrete data. The Chapter 2 introduces a framework for discretizing linear multivariate continuous time systems that includes the commonly used Euler and trapezoidal approximations as special cases and leads to a general class of estimators for the mean reversion matrix. Asymptotic distributions and bias formulae are obtained for estimates of the mean reversion parameter. Explicit expressions are given for the discretization bias and its relationship to estimation bias in both multivariate and in univariate settings. In the univariate context, we compare the performance of the two approximation methods relative to exact maximum likelihood (ML) in terms of bias and variance for the Vasicek process. The bias and the variance of the Euler method are found to be smaller than the trapezoidal method, which are in turn smaller than those of exact ML. Simulations suggest that for plausible parameter settings the approximation methods work better than ML, the bias formulae are accurate, and for scalar models the estimates obtained from the two approximate methods have smaller bias and variance than exact ML. For the square root process, the Euler method outperforms the Nowman method in terms of both bias and variance. Simulation evidence indicates that the Euler method has smaller bias and variance than exact ML, Nowman's method and the Milstein method.

The Chapter 3 examines the asymptotic properties of the maximum likelihood

(ML) estimate of the mean reversion matrix that is obtained from the corresponding exact discrete model. Both the consistency and the asymptotic distribution are derived in the cases of stationarity and non-stationarity. Special attention is paid to the explicit expressions for the asymptotic covariance matrix, especially in low dimensional cases. This limit theory is facilitated by a new formula for the mapping from the discrete to the continuous system coefficients and its derivatives. An empirical application is conducted on daily realized volatility data on Pound, Euro and Yen exchange rates, illustrating the implementation of the theory.

Recently, with the coming of the financial crisis, the interest of using explosive process to model asset bubbles has been growing tremendously. This underlies the importance of statistic properties of the explosive process. The Chapter 4 develops a double asymptotic limit theory for the persistent parameter (κ) in explosive continuous time models driven by Lévy processes with a large number of time span (N) and a small number of sampling interval (h). The simultaneous double asymptotic theory is derived using a technique in the same spirit as in Phillips and Magdalinos (2007) for the mildly explosive discrete time model. Both the intercept term and the initial condition appear in the limiting distribution. In the special case of explosive continuous time models driven by the Brownian motion, we develop the limit theory that allows for the joint limits where $N \to \infty$ and $h \to 0$ simultaneously, the sequential limits where $N \to \infty$ is followed by $h \to 0$, and the sequential limits where $h \to 0$ is followed by $N \to \infty$. All three asymptotic distributions are the same.

Table of Contents

Acknowledgements

First of all, I am deeply indebted to my wife and parents for their continuous support and encouragement.

I also would like to extend my sincere gratitude to my supervisors, Professor Peter C.B. Phillips and Professor Jun Yu, who steered me out of darkness in the initial stage of PhD study, who has provided me with valuable guidance in all the time of research for and writing of this dissertation. Without their stimulating suggestions, enlightening instruction, impressive kindness and patience, this dissertation can not be completed. Their keen and vigorous academic observation enlightens me not only in this dissertation but also in my future studies.

My special thanks go to two committee members, Professor Sainan Jin and Professor Kian Guan Lim, for their helpful comments and guidance throughout the course of this research.

Finally, I deeply appreciate the contribution to this dissertation made in various ways by my friends and classmates, especially Yonghui Zhang.

Chapter 1 Introduction

Continuous time models, which are specified in terms of stochastic differential equations, have found wide applications in economics and finance. Empirical interest in systems of this type has grown particularly rapidly in recent years with the availability of high frequency financial data. Correspondingly, growing attention has been given to the development of econometric methods of inference. In order to capture causal linkages among variables and allow for multiple determining factors, many continuous systems are specified in multivariate form. The literature is now wide-ranging. Bergstrom (1990) motivated the use of multivariate continuous time models in macroeconomics; Sundaresan (2000) provided a list of multivariate continuous time models, particularly multivariate diffusions, in finance; and Piazzesi (2009) discusses how to use multivariate continuous time models to address various macro-finance issues.

Data in economics and finance are typically available at discrete points in time or over discrete time intervals and many continuous systems are formulated as Markov processes. These two features suggest that the log likelihood function can be expressed as the product of the log transition probability densities (TPD). Consequently, the implementation of maximum likelihood (ML) requires evaluation of TPD. But since the TPD is unavailable in closed form for many continuous systems and several methods have been proposed as approximations.

The simplest approach is to approximate the model using some discrete time system. Both the Euler approximation and the trapezoidal rule have been suggested in the literature. Sargan (1974) and Bergstrom (1984) showed that the ML estimators (MLEs) based on these two approximations converge to the true MLE as the sampling interval $h \to 0$, at least under a linear specification. This property also holds for more general diffusions (Florens-Zmirou, 1989). Of course, the quality of the approximation depends on the size of h. However, the advantage of the approximation approach is that it is computationally simple and often works well when h is small, for example at the daily frequency.

More accurate approximations have been proposed in recent years. The two that have received the most attention are in-fill simulations and closed-form approximations. Studies of in-fill simulations include Pedersen (1995) and Durham and Gallant (2002). For closed-form approximations, seminal contributions include Aït-Sahalia (1999, 2002, 2008), Aït-Sahalia and Kimmel (2007), and Aït-Sahalia and Yu (2006). These approximations have the advantage that they can control the size of the approximation errors even when h is not small. Aït-Sahalia (2008) provides evidence that the closed-form approximation is highly accurate and allows for fast repeated evaluations. Since the approximate TPD takes a complicated form in both these approaches, no closed form expression is available for the MLE. Consequently, numerical optimizations are needed to obtain the MLE.

No matter which of the above methods is used, when the system variable is persistent, the resulting estimator of the speed of mean reversion can suffer from severe bias in finite samples. This problem is well known in scalar diffusions (Phillips and Yu, 2005a, 2005b, 2009a, 2009b) but has also been reported in multivariate models (Phillips and Yu, 2005a and Tang and Chen, 2009). In the scalar case, Tang and Chen (2009) and Yu (2009) give explicit expressions to approximate the bias. To obtain these explicit expressions, the corresponding estimators must have a closed-form expression. That is why explicit bias results are presently available only for the scalar Vasicek model (Vasicek, 1977) and the Cox-Ingersoll-Ross (CIR, 1985) model.

The present Chapter 2 focuses on extending existing bias formulae to the multivariate continuous system case. We partly confine our attention to linear systems so that explicit formulae are possible for approximating the estimation bias of the mean reversion matrix. It is known from previous work that bias in the mean reversion parameter has some robustness to specification changes in the diffusion function (Tang and Chen, 2009), which gives this approach a wider relevance. Understanding the source of the mean reversion bias in linear systems can also be helpful in more general situations where there are nonlinearities.

The Chapter 2 develops a framework for studying estimation in multivariate continuous time models with discrete data. In particular, we show how the estimator that is based on the Euler approximation and the estimator based on the trapezoidal approximation can be obtained by taking Taylor expansions to the first and second orders. Moreover, the uniform framework simplifies the derivation of the asymptotic bias order of the ordinary least squares estimator and the two stage least squares estimator of Bergstrom (1984). Asymptotic theory is provided under long time span asymptotics and explicit formulae for the matrix bias approximations are obtained. The bias formulae are decomposed into the discretization bias and the estimation bias. Simulations reveal that the bias formulae work well in practice. The results are specialized to the scalar case, giving two approximate estimators of the mean reversion parameter which are shown to work well relative to the exact MLE.

The results confirm that bias can be severe in multivariate continuous time models for parameter values that are empirically realistic, just as it is in scalar models. Specializing our formulae to the univariate case yields some useful alternative bias expressions. Simulations are reported that detail the performance of the bias formulae in the multivariate setting and in the univariate setting.

Although the approximations by using some discrete time system can help to simplify the estimation of continuous time model, they also bring some drawbacks. For example, all estimators from approximate discrete time model are not consistent. Therefore, it is still necessary to study the estimator from exact discrete time model.

For multivariate continuous time models with a linear drift function, an exact discrete time vector autoregressive (VAR) model can be obtained. When the diffusion function is constant, the VAR model is Gaussian and hence can be estimated by ordinary least squares (OLS) or ML. When the diffusion function has the level effect, the VAR model becomes non-Gaussian but can be estimated by generalized least squares. The asymptotic theory for VAR estimation is standard; see, for example, Mann and Wald (1943) for the stationary case and Phillips and Durlauf (1986) for the unit root case. It is known that the mean reversion matrix in the continuous time model is the logarithmic transformation of the autoregressive (AR) coefficient matrix. Under the identification condition, this relation is bijective. It is this bijective and measurable relationship that will be used find the asymptotic theory of the estimated mean reversion matrix.

It appears that the delta method, when applied to the principal value of the logarithm of the VAR coefficient matrix, can be used to find the limit distribution of the estimated mean reversion matrix. Unfortunately, this straightforward application of the delta method leads to a covariance matrix that is practically difficult to use. The standard limit distribution is available for the estimated VAR coefficient matrix. But to utilize this distribution, the standard matrix calculus formula implies that the mean reversion matrix is expressed as an infinite polynomial of the VAR coefficient matrix. As a result, the covariance matrix involves an infinite polynomial which must be truncated in practice and hence the calculation of the asymptotic covariance is difficult to implement. This situation is in the sharp contrast to the univariate setup where the delta method is easily applied.

The Chapter 3 contributes to the literature in three ways. First, under regular conditions, we derive the asymptotic distribution of the estimated mean reversion matrix whose covariance matrix is very easy to calculate. We do this by using a new result obtained in the linear algebra literature, which enables us to relate the mean reversion matrix to the VAR coefficient matrix as a polynomial function of finite order. Second, we derive the asymptotic theory for the estimated mean reversion matrix not only for the stationary case but also for the non-stationary case. Third, we provide the joint limit distribution of the estimated mean reversion matrix and its eigenvalues. The theory is established in the context of the multivariate diffusion model of an arbitrary dimension but with a linear drift and a constant diffusion. We focus on this model simply because the asymptotic theory is well developed for the exact discrete time model. However, our theory continues to work for models with a more complicated diffusion function. As long as the asymptotic theory for the exact discrete time model is known, our method is applicable.

All Continuous time models concerned in Chapter 2 and Chapter 3 are driven by the Brownian motion, and, therefore, can be called as diffusion processes. An important property of diffusion processes is that, under some smoothness condition on the drift function and the diffusion function, the sample path is continuous everywhere. This restriction is often found to be too strong in applications. There are different ways to introduce discontinuity into the continuous time models. For example, Poisson processes, which allow for a finite number of jumps in a finite time interval, have been used to model jumps in finance (Merton, 1976). In recent years, however, strong evidence of the presence of infinite activity jumps have been documented in finance; see, for example, Aït-Sahalia and Jacod (2011). Consequently, continuous time Lévy processes have become increasingly popular to model discontinuity in financial time series. Not surprisingly, various Lévy processes have been developed in the asset pricing literature (see, for example, Barndorff-Nielsen (1998), Madan, Carr and Chang (1998), Carr and Wu (2003)).

Independent to the development in continuous time modeling, there has been a long-standing interest in statistics for developing the asymptotic theory for explosive processes. Two of the earliest studies are White (1958) and Anderson (1959) where the asymptotic distribution of the autoregressive (AR) coefficient was derived when the root is larger than unity. Phillips and Magdalinos (2007, PM hereafter) has provided an asymptotic theory and an invariance principle for mildly explosive processes where the root is moderately deviated from unity. Magdalinos (2011) extended the result to the case where the error is serially dependent. Anu and Horvath (2007) extended the result to the case where the error is infinite. In economics, there has recently been a growing interest on using explosive processes to model asset price bubbles. Phillips et al (2011) has developed a recursive method to detect bubbles in the discrete time AR model. Phillips and Yu (2011) applied the method to analyze the bubble episodes in various markets in the U.S. and documented the bubble migration mechanism during the subprime crisis.

All the above cited studies on explosiveness focus exclusively on discrete time models. Explosive behavior can also be described using continuous time models. Let T, h, N be the sample size, the sampling interval, and the time span of the data, respectively. Obviously $T = N/h$. While the asymptotic theory in discrete time models always corresponds to the scheme of $T \to \infty$, how to develop the asymptotic theory in continuous time is less a clear cut because $T \to \infty$ is achievable from different ways. In the literature, three alternative sampling schemes have been discussed (see, for example, Jeong and Park (2011) and Zhou and Yu (2011)), namely:

$$N \to \infty, \ \ h \text{ is fixed}; \qquad (A1)$$

$$N \to \infty, \ \ h \to 0; \qquad (A2)$$

$$h \to 0, \ \ N \text{ is fixed}. \qquad (A3)$$

The main purpose of the Chapter 4 is to develop the double asymptotic theory under scheme (A2) for explosive continuous time models driven by Lévy processes, in which $N \to \infty$ and $h \to 0$ simultaneously. In the special case of Brownian motion driven continuous time models, three alternative double asymptotics are considered. In the first case, $N \to \infty$ and $h \to 0$ simultaneously. In the second case, a sequential asymptotic treatment is considered, i.e., $N \to \infty$ is followed by $h \to 0$. In the third case, another sequential asymptotic treatment is considered wherein, $h \to 0$ is followed by $N \to \infty$. We show that the asymptotic distributions under these three treatments are the same. Different from PM, in our double asymptotic distribution, the initial condition, either fixed or random, appears in the limiting distribution.

Chapter 2 Bias in Estimating Multivariate and Univariate Diffusions

2.1 Introduction

Continuous time models, which are specified in terms of stochastic differential equations, have found wide applications in economics and finance. Empirical interest in systems of this type has grown particularly rapidly in recent years with the availability of high frequency financial data. Correspondingly, growing attention has been given to the development of econometric methods of inference. In order to capture causal linkages among variables and allow for multiple determining factors, many continuous systems are specified in multivariate form. The literature is now wide-ranging. Bergstrom (1990) motivated the use of multivariate continuous time models in macroeconomics; Sundaresan (2000) provided a list of multivariate continuous time models, particularly multivariate diffusions, in finance; and Piazzesi (2009) discusses how to use multivariate continuous time models to address various macro-finance issues.

Data in economics and finance are typically available at discrete points in time or over discrete time intervals and many continuous systems are formulated as Markov processes. These two features suggest that the log likelihood function can be expressed as the product of the log transition probability densities (TPD). Consequently, the implementation of maximum likelihood (ML) requires evaluation of TPD. But since the TPD is unavailable in closed form for many continuous systems and several methods have been proposed as approximations.

The simplest approach is to approximate the model using some discrete time system. Both the Euler approximation and the trapezoidal rule have been suggested in the literature. Sargan (1974) and Bergstrom (1984) showed that the ML estimators (MLEs) based on these two approximations converge to the true MLE as the sampling interval $h \to 0$, at least under a linear specification. This property also holds for more general diffusions (Florens-Zmirou, 1989). Of course, the quality of the approximation depends on the size of h. However, the advantage of the approximation approach is that it is computationally simple and often works well when h is small, for example at the daily frequency.

More accurate approximations have been proposed in recent years. The two that have received the most attention are in-fill simulations and closed-form approximations. Studies of in-fill simulations include Pedersen (1995) and Durham and Gallant (2002). For closed-form approximations, seminal contributions include Aït-Sahalia (1999, 2002, 2008), Aït-Sahalia and Kimmel (2007), and Aït-Sahalia and Yu (2006). These approximations have the advantage that they can control the size of the approximation errors even when h is not small. Aït-Sahalia (2008) provides evidence that the closed-form approximation is highly accurate and allows for fast repeated evaluations. Since the approximate TPD takes a complicated form in both these approaches, no closed form expression is available for the MLE. Consequently, numerical optimizations are needed to obtain the MLE.

No matter which of the above methods is used, when the system variable is persistent, the resulting estimator of the speed of mean reversion can suffer from severe bias in finite samples. This problem is well known in scalar diffusions (Phillips and Yu, 2005a, 2005b, 2009a, 2009b) but has also been reported in multivariate models (Phillips and Yu, 2005a and Tang and Chen, 2009). In the scalar case, Tang and Chen (2009) and Yu (2009) give explicit expressions to approximate the bias. To obtain these explicit expressions, the corresponding estimators must have a closed-form expression. That is why explicit bias results are presently available only for the scalar Vasicek model (Vasicek, 1977) and the Cox-Ingersoll-Ross (CIR, 1985) model.

The present paper focuses on extending existing bias formulae to the multivariate continuous system case. We partly confine our attention to linear systems so that explicit formulae are possible for approximating the estimation bias of the mean reversion matrix. It is known from previous work that bias in the mean reversion parameter has some robustness to specification changes in the diffusion function (Tang and Chen, 2009), which gives this approach a wider relevance. Understanding the source of the mean reversion bias in linear systems can also be helpful in more general situations where there are nonlinearities.

The paper develops a framework for studying estimation in multivariate continuous time models with discrete data. In particular, we show how the estimator that is based on the Euler approximation and the estimator based on the trapezoidal approximation can be obtained by taking Taylor expansions to the first and second orders. Moreover, the uniform framework simplifies the derivation of the asymptotic bias order of the ordinary least squares estimator and the two stage least squares estimator of Bergstrom (1984). Asymptotic theory is provided under long time span asymptotics and explicit formulae for the matrix bias approximations are obtained. The bias formulae are decomposed into the discretization bias and the estimation bias. Simulations reveal that the bias formulae work well in practice. The results are specialized to the scalar case, giving two approximate estimators of the mean reversion parameter which are shown to work well relative to the exact MLE.

The results confirm that bias can be severe in multivariate continuous time models for parameter values that are empirically realistic, just as it is in scalar models. Specializing our formulae to the univariate case yields some useful alternative bias expressions. Simulations are reported that detail the performance of the bias formulae in the multivariate setting and in the univariate setting.

The rest of the paper is organized as follows. Section 2 introduces the model and the setup and reviews four existing estimation methods. Section 3 outlines our unified framework for estimation, establishes the asymptotic theory, and provides explicit expressions for approximating the bias in finite samples. Section 4 discusses the relationship between the new estimators and two existing estimators in the literature, and derives a new bias formula in the univariate setting. Section 5 compares the performance of the estimator based on the Euler scheme relative to that the method proposed by Nowman (1997) in the context of the square root process and a diffusion process with a linear drift but a more general diffusion. Simulations are reported in Section 6. Section 7 concludes and the Appendix collects together proofs of the main results.

2.2 The Model and Existing Methods

We consider an M-dimensional multivariate diffusion process of the form (cf. Phillips, 1972):

$$dX(t) = (A(\theta)X(t) + B(\theta))dt + \zeta(dt), \ \ X(0) = X_0, \tag{2.2.1}$$

where $X(t) = (X_1(t), \cdots, X_M(t))'$ is an M-dimensional continuous time process, $A(\theta)$ and $B(\theta)$ are $M \times M$ and $M \times 1$ matrices, whose elements depend on unknown parameters $\theta = (\theta_1, \cdots, \theta_K)$ that need to be estimated, $\zeta(dt)$ $(:= (\zeta_1(dt), \cdots, \zeta_M(dt)))$ is a vector random process with uncorrelated increments and covariance matrix Σdt. The particular model receiving most attention in finance is when $\zeta(dt)$ is a vector of Brownian increments (denoted by $dW(t)$) with covariance Σdt, viz.,

$$dX(t) = (A(\theta)X(t) + B(\theta))dt + dW(t), \ \ X(0) = X_0, \tag{2.2.2}$$

corresponding to a multivariate version of the Vasicek model (Vasicek, 1977).

Although the process follows a continuous time stochastic differential equation system, observations are available only at discrete time points, say at n equally spaced points $\{th\}_{t=0}^{n}$, where h is the sampling interval and is taken to be fixed. In practice, h might be very small, corresponding to high-frequency data. In this paper, we use $X(t)$ to represent a continuous time process and X_t to represent a discrete time process. When there is no confusion, we simply write X_{th} as X_t.

Bergstrom (1990) provided arguments why it is useful for macro-economists and policy makers like central bankers to formulate models in continuous time even when discrete observations only are available. In finance, early fundamental work by Black and Scholes (1973) and much of the ensuing literature such as Duffie and Kan (1996) successfully demonstrated the usefulness of both scalar and multivariate diffusion models in the development of financial asset pricing theory.

Phillips (1972) showed that the exact discrete time model corresponding to (2.2.1) is given by

$$X_t = \exp\{A(\theta)h\}X_{t-1} - A^{-1}(\theta)[\exp\{A(\theta)h\} - I]B(\theta) + \varepsilon_t. \tag{2.2.3}$$

where $\varepsilon_t = (\varepsilon_1, \cdots, \varepsilon_M)'$ is a martingale difference sequence (MDS) with respect to the natural filtration and

$$E(\varepsilon_t \varepsilon_t^{'}) = \int_0^h \exp\{A(\theta)s\}\Sigma\exp\{A(\theta)'s\}ds := G.$$

Letting $F(\theta) := \exp\{A(\theta)h\}$ and $g(\theta) := -A^{-1}(\theta)[\exp\{A(\theta)h\} - I]B(\theta)$, we have the system

$$X_t = F(\theta)X_{t-1} + g(\theta) + \varepsilon_t, \tag{2.2.4}$$

which is a vector auto-regression (VAR) model of order 1 with $\text{MDS}(0, G)$ innovations.

In general, identification of θ from the implied discrete model (2.2.3) generating discrete observations $\{X_{th}\}$ is not automatically satisfied. The necessary and sufficient condition for identifiability of θ in model (2.2.3) is that the correspondence between θ and $[F(\theta), g(\theta)]$ be one-to-one, since (2.2.3) is effectively a reduced form for the discrete observations. Phillips (1973) studied the identifiability of $(A(\theta), \Sigma)$ in (2.2.3) in terms of the identifiability of the matrix $A(\theta)$ in the matrix exponential $F = \exp(A(\theta)h)$ under possible restrictions implied by the structural functional dependence $A = A(\theta)$ in (2.2.1). In general, a one-to–one correspondence between

$A(\theta)$ and F, requires the structural matrix $A(\theta)$ to be restricted. This is because if $A(\theta)$ satisfies $\exp\{A(\theta)h\} = F$ and some of its eigenvalues are complex, $A(\theta)$ is not uniquely identified. In fact, adding to each pair of conjugate complex eigenvalues the imaginary numbers $2ik\pi$ and $-2ik\pi$ for any integer k, leads to another matrix satisfying $\exp\{Ah\} = F$. This phenomenon is well known as aliasing in the signal processing literature. When restrictions are placed on the structural matrix $A(\theta)$ identification is possible. Phillips (1973) gave a rank condition for the case of linear homogeneous relations between the elements of a row of A. A special case is when $A(\theta)$ is triangular. Hansen and Sargent (1983) extended this result by showing that the reduced form covariance structure $G > 0$ provides extra identifying information about A, reducing the number of potential aliases.

To deal with the estimation of (2.2.1) using discrete data and indirectly (because it was not mentioned) the problem of identification, two approximate discrete time models were proposed in earlier studies. The first is based on the Euler approximation given by

$$\int_{(t-1)h}^{th} A(\theta)X(r)dr \approx A(\theta)hX_{t-1},$$

which leads to the approximate discrete time model

$$X_t - X_{t-1} = A(\theta)hX_{t-1} + B(\theta)h + u_t. \tag{2.2.5}$$

The second, proposed by Bergstrom (1966), is based on the trapezoidal approximation

$$\int_{(t-1)h}^{th} A(\theta)X(r)dr \approx \frac{1}{2}A(\theta)h(X_t + X_{t-1}),$$

which gives rise to the approximate nonrecursive discrete time model

$$X_t - X_{t-1} = \frac{1}{2}A(\theta)h(X_t + X_{t-1}) + B(\theta)h + v_t. \tag{2.2.6}$$

The discrete time models are then estimated by standard statistical methods, namely OLS for the Euler approximation and systems estimation methods such as two-stage or three-stage least squares for the trapezoidal approximation. As explained by Lo (1988) in the univariate context, such estimation strategies inevitably suffer from discretization bias. The size of the discretization bias depends on the sampling interval, h, and does not disappear even if $n \to \infty$. The bigger is h, the larger is the discretization bias. Sargan (1974) showed that the asymptotic discretization bias of the two-stage and three-stage least squares estimators for the trapezoidal approximation is $O(h^2)$ as $h \to 0$. Bergstrom (1984) showed that the asymptotic discretization bias of the OLS estimator for the Euler approximation is $O(h)$.

For the more general multivariate diffusion

$$dX(t) = \kappa(\mu - X(t))dt + \Sigma(X(t);\psi)dW(t), \ \ X(0) = X_0, \tag{2.2.7}$$

where W is standard Brownian motion, two other approaches have been used to approximate the continuous time model (2.2.7). The first, proposed by Nowman (1997), approximates the diffusion function within each unit interval, $[(i-1)h, ih)$ by its left end point value leading to the approximate model

$$dX(t) = \kappa(\mu - X(t))dt + \Sigma(X_{(i-1)h};\psi)dW(t) \quad \text{for } t \in [(i-1)h, ih). \tag{2.2.8}$$

Since (2.2.8) is a multivariate Vasicek model within each unit interval, there is a corresponding exact discrete model as in (2.2.3). This discrete time model, being an approximation to the exact discrete time model of (2.2.7), facilitates direct Gaussian estimation.

To reduce the approximation error introduced by the Euler scheme, Milstein (1978) suggested taking the second order term in a stochastic Taylor series expansion when approximating the drift function and the diffusion function. Integrating (2.2.7) gives

$$\int_{(i-1)h}^{ih} dX(t) = \int_{(i-1)h}^{ih} \kappa(\mu - X(t))dt + \int_{(i-1)h}^{ih} \Sigma(X(t);\psi)dW(t). \tag{2.2.9}$$

By Itô's lemma, the linearity of the drift function in (2.2.7), and using tensor summation notation for repeated indices (p,q), we obtain

$$d\mu(X(t);\theta) = \frac{\partial\mu(X(t);\theta)}{\partial X_p} dX_p(t),$$

and

$$d\Sigma(X(t);\psi) = \frac{\partial\Sigma(X(t);\psi)}{\partial X_p} dX_p(t) + \frac{1}{2}\frac{\partial^2\Sigma(X(t);\psi)}{\partial X_p \partial X_q'} dX_p(t) dX_q(t), \qquad (2.2.10)$$

where $\mu_j(X(t);\theta)$ is the j^{th} element of the (linear) drift function $\kappa(\mu - X(t))$, Σ_{pq} is the $(p,q)^{th}$ element of Σ and X_p is the p^{th} element of X. These expressions lead to the approximations

$$\mu(X(t);\theta) \simeq \mu(X_{(i-1)h};\theta),$$

and

$$\Sigma(X(t);\psi) \simeq \Sigma(X_{(i-1)h};\theta) + \frac{\partial\Sigma(X_{(i-1)h};\psi)}{\partial X_p}\Sigma_{pq}(X_{(i-1)h};\psi)\int_{(i-1)h}^{t} dW_q(\tau).$$

Using these approximations in (2.2.9) we find

$$\begin{aligned} X_{ih} - X_{(i-1)h} &= \int_{(i-1)h}^{ih} \kappa(\mu - X(t))dt + \int_{(i-1)h}^{ih} \Sigma(X(t);\psi)dW(t) \\ &\simeq \mu(X_{(i-1)h};\theta)h + \Sigma(X_{(i-1)h};\psi)\int_{(i-1)h}^{ih} dW(t) \\ &+ \frac{\partial\Sigma(X_{(i-1)h};\psi)}{\partial X_p}\Sigma_{pq}(X_{(i-1)h};\psi)\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t} dW_q(\tau)dW(t). \end{aligned} \qquad (2.2.11)$$

The multiple (vector) stochastic integral in (2.2.11) reduces as follows:

$$\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t} dW_q(\tau)dW_p(t) = \int_{(i-1)h}^{ih}\left(W_q(t)-W_{q(i-1)h}\right)dW_p(t)$$
$$= \begin{cases} \frac{1}{2}\left\{\left(W_{qih}-W_{q(i-1)h}\right)^2 - h\right\} & p=q \\ \int_{(i-1)h}^{ih}\left(W_q(t)-W_{q(i-1)h}\right)dW_p(t) & p\neq q \end{cases}. \tag{2.2.12}$$

The approximate model under a Milstein second order discretization is then

$$X_{ih}-X_{(i-1)h} \simeq \mu(X_{(i-1)h};\theta)h + \Sigma(X_{(i-1)h};\psi)\left(W_{ih}-W_{(i-1)h}\right)$$
$$+ \frac{\partial\Sigma(X_{(i-1)h};\psi)}{\partial X_p}\Sigma_{pq}(X_{(i-1)h};\psi)\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t} dW_q(\tau)dW_p(t). \tag{2.2.13}$$

In view of the calculation (2.2.12), when the model is scalar the discrete approximation has the simple form (c.f., Phillips and Yu, 2009)

$$X_{ih}-X_{(i-1)h} \simeq \left[\mu(X_{(i-1)h};\theta) - \frac{1}{2}\sigma'(X_{(i-1)h};\psi)\sigma(X_{(i-1)h};\psi)\right]h$$
$$+ \sigma(X_{(i-1)h};\psi)\left(W_{ih}-W_{(i-1)h}\right)$$
$$+ \sigma'(X_{(i-1)h};\psi)\sigma(X_{(i-1)h};\psi)\frac{1}{2}\left(W_{ih}-W_{(i-1)h}\right)^2. \tag{2.2.14}$$

Since $\frac{1}{2}\left\{\left(W_{qih}-W_{q(i-1)h}\right)^2 - h\right\}$ has mean zero, the net contribution to the drift from the second order term is zero.

In the multivariate Vasicek model, $\Sigma(X(t);\psi)=\Sigma$, and the Milstein approximation (2.2.13) reduces to

$$X_{ih}-X_{(i-1)h} \simeq \mu(X_{(i-1)h};\theta)h + \Sigma(X_{(i-1)h};\psi)\left(W_{ih}-W_{(i-1)h}\right).$$

Thus, for the multivariate Vasicek model, the Milstein and Euler schemes are equivalent.

2.3 Estimation Methods, Asymptotic Theory and Bias

In this paper, following the approach of Phillips (1972), we estimate θ directly from the exact discrete time model (2.2.3). In particular, we first estimate $F(\theta)$ and θ from (2.2.3), assuming throughout that $A(\theta)$ and θ are identifiable and that all the eigenvalues in $A(\theta)$ have negative real parts. The latter condition ensures that X_t is stationary and is therefore mean reverting. The exact discrete time model (2.2.3) in this case is a simple VAR(1) model which has been widely studied in the discrete time series literature. We first review some relevant results from this literature.

Let $Z_t = [X_t', 1]'$. The OLS estimator of $H = [F, g]$ is

$$\hat{H} = [\hat{F}, \hat{g}] = \left[n^{-1} \sum_{t=1}^{n} X_t Z_{t-1}' \right] \cdot \left[n^{-1} \sum_{t=1}^{n} Z_{t-1} Z_{t-1}' \right]^{-1}. \qquad (2.3.1)$$

If we have prior knowledge that $B(\theta) = 0$ and hence $g = 0$, the OLS estimator of F is:

$$\hat{F} = \left[n^{-1} \sum_{t=1}^{n} X_t X_{t-1}' \right] \cdot \left[n^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' \right]^{-1}, \qquad (2.3.2)$$

for which the standard theory first order limit theory (e.g., Fuller (1976, p.340) and Hannan (1970, p.329)) is well known.

Lemma 2.3.1 *For the stationary VAR(1) model (2.2.4), if h is fixed and $n \to \infty$, we have*

(a) $\hat{F} \xrightarrow{p} F$;

(b) $\sqrt{n}\{Vec(\hat{F}) - Vec(F)\} \xrightarrow{d} N(0, (\Gamma(0))^{-1} \otimes G)$,

where $\Gamma(0) = Var(X_t) = \sum_{i=0}^{\infty} F^i \cdot G \cdot F'^i$ *and* $G = E(\varepsilon_t \varepsilon_t')$

Under different but related conditions, Yamamoto and Kunitomo (1984) and Nicholls and Pope (1988) derived explicit bias expressions for the OLS estimator $\hat{F}$. The proof of the following lemma is given in Yamamoto and Kunitomo (1984, theorem 1).

Lemma 2.3.2 *(Yamamoto and Kunitomo (1984)) Assume:*

(A1) X_t is a stationary VAR(1) process whose error term is iid $(0,G)$ with G nonsingular;

(A2) For some $s_0 \geq 16$, $E|\varepsilon_{ti}|^{s_0} < \infty$, for all $i = 1,\cdots,M$;

(A3) $E\left\|\left[n^{-1}\sum_{t=1}^{n} Z_{t-1}Z'_{t-1}\right]^{-1}\right\|^2$ is bounded, where the operator $\|\cdot\|$ is defined by

$$\| Q \| = \sup_{\beta}(\beta' Q' Q \beta)^{1/2}(\beta'\beta \leq 1),$$

for any vector β;

Under (A1)-(A3) if $n \to \infty$, the bias of OLS estimator of F in the VAR(1) model with an unknown intercept is

$$BIAS(\hat{F}) = -n^{-1}G\sum_{k=0}^{\infty}\{F'^{k} + F'^{k}tr(F^{k+1}) + F'^{2k+1}\}D^{-1} + O(n^{-\frac{3}{2}}), \qquad (2.3.3)$$

where

$$D = \sum_{i=0}^{\infty} F^{i} G F'^{i},$$

and the bias of the OLS estimator of F for the VAR(1) model with a known intercept is

$$BIAS(\hat{F}) = -\frac{1}{n}G\sum_{k=0}^{\infty}\{F'^{k}tr(F^{k+1}) + F'^{2k+1}\}D^{-1} + O(n^{-\frac{3}{2}}). \qquad (2.3.4)$$

We now derive a simplified bias formulae in the two models which facilitates the calculation of the bias formulae in continuous time models.

Lemma 2.3.3 *Assume (A1)-(A3) hold, h is fixed and $n \to \infty$. The bias of the least squares estimator for F in the VAR(1) is given by*

$$B_n = E(\hat{F}) - F = -\frac{b}{n} + O(n^{-\frac{3}{2}}). \qquad (2.3.5)$$

When the model has a unknown intercept,

$$b = G[(I-C)^{-1} + C(I-C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1}, \qquad (2.3.6)$$

where $C = F'$, $\Gamma(0) = Var(X_t) = \sum_{t=0}^{\infty} F^i \cdot G \cdot F'^i$, $G = E(\varepsilon_t \varepsilon_t')$, and $Spec(C)$ denotes the set of eigenvalues of C. When the model has a known intercept,

$$b = G[C(I - C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda (I - \lambda C)^{-1}]\Gamma(0)^{-1}. \tag{2.3.7}$$

Remark 2.3.1 *The alternative bias formula (2.3.5) is exactly the same as that given by Nicholls and Pope (1988) for the Gaussian case, although here the expression is obtained without Gaussianity and in a simpler way. If the bias is calculated to a higher order, Bao and Ullah (2009) showed that skewness and excess kurtosis of the error distribution figure in the formulae. In a related contribution, Ullah et al (2010) obtain the second order bias in the mean reversion parameter for a (scalar) continuous time Lévy process.*

We now develop estimators for A. To do so we use the matrix exponential expression

$$F = e^{Ah} = \sum_{i=0}^{\infty} \frac{(Ah)^i}{i!} = I + Ah + H = I + Ah + O(h^2) \text{ as } h \to 0. \tag{2.3.8}$$

Rearranging terms we get

$$A = \frac{1}{h}(F - I) - \frac{1}{h}H = \frac{1}{h}(F - I) + O(h) \text{ as } h \to 0, \tag{2.3.9}$$

which suggest the following simple estimator of A

$$\hat{A} = \frac{1}{h}(\hat{F} - I), \tag{2.3.10}$$

where $\hat{F}$ is the OLS estimator of F. We now develop the asymptotic distribution for $\hat{A}$ and the bias in $\hat{A}$.

Theorem 2.3.1 *Assume X_t follows Model (2.2.1) and that all characteristic roots of the coefficient matrix A have negative real parts. Let $\{X_{th}\}_{t=1}^{n}$ be the available data and suppose A is estimated by (2.3.10) with $\hat{F}$ defined by (2.3.1). When h is fixed,* *as* $n \to +\infty$, *we have*

$$\hat{A} - A \xrightarrow{p} \frac{1}{h}(F - I - Ah) = \frac{1}{h}H = O(h) \text{ as } h \to 0, \tag{2.3.11}$$

where $H = F - I - Ah$, *and*

$$h\sqrt{n}Vec\left[\hat{A} - \frac{1}{h}(F - I)\right] \xrightarrow{d} N(0, \Gamma(0)^{-1} \otimes G), \tag{2.3.12}$$

where $\Gamma(0) = Var(X_t) = \sum_{i=0}^{\infty} F^i G F'^i$, $G = E(\varepsilon_t \varepsilon_t')$.

Theorem 2.3.2 *Assume that* X_t *follows Model (2.2.2) where* $W(t)$ *is a vector Brownian Motion with covariance matrix* Σ *and that all characteristic roots of the coefficient matrix* A *have negative real parts. Let* $\{X_{th}\}_{t=1}^{n}$ *be the available data and suppose* A *is estimated by (2.3.10) with* $\hat{F}$ *defined by (2.3.1). When* h *is fixed and* $n \to \infty$, *the bias formula is:*

$$BIAS(\hat{A}) = E(\hat{A} - A) = \frac{1}{h}H + \frac{-b}{T} + o(T^{-1}), \tag{2.3.13}$$

where $H = F - I - Ah$, *and* $T = nh$ *is the time span of the data. If* $B(\theta)$ *is unknown, then*

$$b = G[(I - C)^{-1} + C(I - C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda (I - \lambda C)^{-1}]\Gamma(0)^{-1}, \tag{2.3.14}$$

where $\Gamma(0) = Var(X_t) = \sum_{i=0}^{\infty} F^i \cdot G \cdot F'^i$, $G = E(\varepsilon_t \varepsilon_t')$, *and* $Spec(C)$ *is the set of eigenvalues of* C. *If* $B(\theta)$ *is known, then*

$$b = G[C(I - C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda (I - \lambda C)^{-1}]\Gamma(0)^{-1}. \tag{2.3.15}$$

Remark 2.3.2 *Expression (2.3.11) extends the result in equation (32) of Lo (1988) to the multivariate case. According to Theorem 2.3.2, the bias of the estimator (2.3.10) can be decomposed into two parts, the discretization bias and the estima-*

tion bias, which take the following forms:

$$\textit{discretization bias} = \frac{H}{h} = \frac{F-I-Ah}{h} = O(h) \textit{ as } h \to 0,$$

$$\textit{estimation bias} = \frac{-b}{T} + o(T^{-1}).$$

It is difficult to determine the signs of the discretization bias and the estimation bias in a general multivariate case. However, in the univariate case, the signs are opposite to each other as shown in Section 4.2.

Higher order approximations are possible. For example, we may take the matrix exponential series expansion to the second order to produce a more accurate estimate using

$$\begin{aligned}
F = e^{Ah} &= \sum_{i=0}^{\infty} \frac{(Ah)^i}{i!} \\
&= I + Ah + \frac{Ah}{2}\left[(e^{Ah} - I) + \frac{-A^2h^2}{3!} + \frac{-2A^3h^3}{4!} + \ldots + \frac{-(n-2)A^{n-1}h^{n-1}}{n!} + \cdots\right] \\
&= I + Ah + \frac{Ah}{2}[F-I] + \eta \\
&= I + Ah + \frac{Ah}{2}[F-I] + O(h^3) \text{ as } h \to 0. \qquad (2.3.16)
\end{aligned}$$

Consequently,

$$\begin{aligned}
A &= \frac{2}{h}(F-I)(F+I)^{-1} - \frac{2}{h}\eta(F+I)^{-1} = \frac{2}{h}(F-I)(F+I)^{-1} + \nu \\
&= \frac{2}{h}(F-I)(F+I)^{-1} + O(h^2) \text{ as } h \to 0. \qquad (2.3.17)
\end{aligned}$$

After neglecting terms smaller than $O\left(h^2\right)$, we get the alternative estimator

$$\hat{A} = \frac{2}{h}(\hat{F} - I)(\hat{F} + I)^{-1}. \qquad (2.3.18)$$

Theorem 2.3.3 *Assume that X_t follows Model (2.2.1) and that all characteristic roots of the coefficient matrix A have negative real parts. Let $\{X_{th}\}_{t=1}^{n}$ be the available data and A is estimated by (2.3.18) with $\hat{F}$ defined by (2.3.1). When h is fixed,*

$n \to +\infty$, we have

$$\hat{A} - A \xrightarrow{p} \frac{2}{h}(F-I)(F+I)^{-1} - A = O(h^2) \text{ as } h \to 0,$$

and

$$h\sqrt{n}Vec\left[\hat{A} - \frac{2}{h}(F-I)(F+I)^{-1}\right] \xrightarrow{d} N(0,\Psi),$$

where

$$\Psi = 16\Upsilon[\Gamma(0)^{-1} \otimes G]\Upsilon', \quad \Upsilon = (F'+I)^{-1} \otimes (F+I)^{-1}.$$

Theorem 2.3.4 *Assume that X_t follows (2.2.2) where $W(t)$ is a vector Brownian motion with covariance matrix Σ and that all characteristic roots of the coefficient matrix A have negative real parts. Let $\{X_{th}\}_{t=1}^n$ be the available data and suppose A is estimated by (2.3.18) with $\hat{F}$ defined by (2.3.1). When h is fixed, $n \to \infty$, and $T = hn$, the bias formula is:*

$$BIAS(\hat{A}) = -\nu - \frac{4}{T}(I+F)^{-1}b(I+F)^{-1} - \frac{4}{h}L(I+F)^{-1} + o(T^{-1}), \qquad (2.3.19)$$

where $\nu = A - \frac{2}{h}(F-I)(F+I)^{-1}$, $\Delta = [I_M \otimes (I+F)^{-1}] \cdot \Gamma(0)^{-1} \otimes G \cdot [I_M \otimes (I+F)^{-1}]'$, and L is a $M \times M$ matrix whose ij^{th} element is given by

$$L_{ij} = \frac{1}{n}\sum_{s=1}^{M} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s},$$

with e_i being a column vector of dimension M^2 whose i^{th} element is 1 and other elements are 0. If $B(\theta)$ is an unknown vector, then

$$b = G[(I-C)^{-1} + C(I-C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1}.$$

If $B(\theta)$ is a known vector, then

$$b = G[C(I-C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1}.$$

Remark 2.3.3 *Theorem 2.3.4 shows that the bias of the estimator (2.3.18) can be decomposed into a discretization bias and an estimation bias as follows:*

$$\textit{discretization bias} \; = \; -\nu = \frac{2}{h}(F-I)(F+I)^{-1} - A = O(h^2) \textit{ as } h \to 0,$$

$$\textit{estimation bias} = -\frac{4}{T}(I+F)^{-1} b (I+F)^{-1} - \frac{4}{h} L (I+F)^{-1} + o(T^{-1}).$$

As before, it is difficult to determine the signs of the discretization bias and estimation bias in a general multivariate case. However, in the univariate case, the signs are opposite each other as reported in Section 4.2.

Remark 2.3.4 *The estimator (2.3.10) is based on a first order Taylor expansion whereas the estimator (2.3.18) is based on a second order expansion, so it is not surprising that (2.3.18) has a smaller discretization bias than (2.3.10). It is not as easy to compare the magnitudes of the two estimation biases. In the univariate case, however, we show in Section 4.2 that the estimator (2.3.18) has a larger estimation bias than the estimator (2.3.10).*

2.4 Relations to Existing Results

2.4.1 The Euler and Trapezoidal Approximations

The estimators given above include as special cases the two estimators obtained from the Euler approximation and the trapezoidal approximation. Consequently, both the asymptotic and the bias properties are applicable to these two approximation models and the simple framework above unifies some earlier theory on the estimation of approximate discrete time models.

The Euler approximate discrete time model is of the form:

$$X_t - X_{t-1} = AhX_{t-1} + Bh + u_t.$$

The OLS estimator of A is given by

$$[\widehat{I+Ah}, \widehat{Bh}] := \left[n^{-1}\sum_{t=1}^{n} X_t Z_{t-1}^{'}\right]\left[n^{-1}\sum_{t=1}^{n} Z_{t-1} Z_{t-1}^{'}\right]^{-1} =: [\hat{F}, \hat{g}].$$

If B is known apriori and assumed zero without loss of generality, then the OLS estimator of A is

$$[\widehat{I+Ah}] = \left[n^{-1}\sum_{t=1}^{n} X_t X_{t-1}^{'}\right]\left[n^{-1}\sum_{t=1}^{n} X_{t-1} X_{t-1}^{'}\right]^{-1} =: [\hat{F}],$$

where Z_{t-1}, $\hat{F}$, $\hat{g}$ are defined in the same way as before. Hence,

$$\hat{A} = \frac{1}{h}[\hat{F} - I].$$

This is precisely the estimator given by (2.3.10) based on a first order expansion of the matrix exponential $\exp(Ah)$ in h.

The trapezoidal approximate discrete time model is of the form

$$X_t - X_{t-1} = \frac{1}{2}Ah(X_t + X_{t-1}) + Bh + \nu_t. \quad (2.4.1)$$

If $B = 0$, the approximate discrete model becomes

$$X_t - X_{t-1} = \frac{1}{2}Ah(X_t + X_{t-1}) + \nu_t. \quad (2.4.2)$$

Note that (2.4.2) is a simultaneous equations model, as emphasized by Bergstrom (1966,1984). We show that the two stage least squares estimator of A from (2.4.1) is equivalent to the estimator given by (2.3.18) based on a second order expansion of $\exp(Ah)$ in h. To save space, we focus on the approximate discrete time model with known $B = 0$. The result is easily extended to the case of unknown B.

The two stage least squares estimator of Bergstrom (1984) takes the form

$$\hat{A} = \left[\sum_{t=1}^{n} \frac{1}{h}(X_t - X_{t-1})V_t'\right]\left[\sum_{t=1}^{n} \frac{1}{2}(X_t + X_{t-1})V_t'\right]^{-1}, \tag{2.4.3}$$

where

$$V_t = \frac{1}{2}(X_t^* + X_{t-1}), \tag{2.4.4}$$

$$X_t^* = \left[\sum_{t=1}^{n} X_t X_{t-1}'\right]\left[\sum_{t=1}^{n} X_{t-1} X_{t-1}'\right]^{-1} X_{t-1}. \tag{2.4.5}$$

Theorem 2.4.1 *The two stage least squares estimator suggested in Bergstrom (1984) has the following form*

$$\hat{A} = \frac{2}{h}[\hat{F} - I][\hat{F} + I]^{-1},$$

and is precisely the same estimator as that given by (2.3.18) based on a second order expansion of $\exp(Ah)$ *in* h.

2.4.2 Bias in univariate models

The univariate diffusion model considered in this section is the OU process:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma dW(t),\ X(0) = 0, \tag{2.4.6}$$

where $W(t)$ is a standard scalar Brownian motion. The exact discrete time model corresponding to (2.4.6) is

$$X_t = \phi X_{t-1} + \mu(1 - e^{-\kappa h}) + \sigma\sqrt{\frac{1 - e^{-2\kappa h}}{2\kappa}}\varepsilon_t, \tag{2.4.7}$$

where $\phi = e^{-\kappa h}$, $\varepsilon_t \sim iid\ N(0,1)$ and h is the sampling interval.

The ML estimator of κ (conditional on X_0) is given by

$$\hat{\kappa} = -\ln(\hat{\phi})/h, \tag{2.4.8}$$

where

$$\hat{\phi} = \frac{n^{-1}\Sigma X_t X_{t-1} - n^{-2}\Sigma X_t \Sigma X_{t-1}}{n^{-1}\Sigma X_t^2 - n^{-2}(\Sigma X_{t-1})^2},$$

and $\hat{\kappa}$ exists provided $\hat{\phi} > 0$. Tang and Chen (2009) analyzed the asymptotic properties and derived the finite sample variance formula and the bias formula, respectively,

$$Var(\hat{\kappa}) = \frac{1-\phi^2}{Th\phi^2} + o(T^{-1}),$$

$$E(\hat{\kappa}) - \kappa = \frac{1}{T}\left(\frac{5}{2} + e^{\kappa h} + \frac{e^{2\kappa h}}{2}\right) + o(T^{-1}). \tag{2.4.9}$$

When μ is known (assumed to be 0), the exact discrete model becomes

$$X_t = \phi X_{t-1} + \delta\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_t, \tag{2.4.10}$$

and the ML estimator of κ is $\hat{\kappa} = -\ln(\hat{\phi})/h$, where $\hat{\phi} = \Sigma X_t X_{t-1}/\Sigma X_{t-1}^2$. In this case, Yu (2009) derived the following bias formula under stationary initial conditions

$$E(\hat{\kappa}) - \kappa = \frac{1}{2T}(3 + e^{2\kappa h}) - \frac{2(1-e^{-2n\kappa h})}{Tn(1-e^{-2\kappa h})} + o(T^{-1}). \tag{2.4.11}$$

When the initial condition is $X(0) = 0$, the bias formula becomes

$$E(\hat{\kappa}) - \kappa = \frac{1}{2T}(3 + e^{2\kappa h}) + o(T^{-1}). \tag{2.4.12}$$

Since the MLE is based on the exact discrete time model, there is no discretization bias in (2.4.7) and (2.4.10). The bias in $\hat{\kappa}$ is induced entirely by estimation and is always positive.

We may link our results for multivariate systems to the univariate model. For example, $\kappa = -A$ in (2.4.6) and the first order Taylor series expansion (i.e., the

Euler method) gives the estimator

$$\widehat{\kappa}_1 = \frac{1}{h}[1-\hat{\phi}]. \tag{2.4.13}$$

In this case the results obtained in Theorems 2.3.1 and 2.3.2 may be simplified as in the following two results.

Theorem 2.4.2 *Assuming $\kappa > 0$, when h is fixed, and $n \to \infty$, we have*

$$\hat{\kappa}_1 - \kappa \xrightarrow{p} -\frac{\exp(-\kappa h) - 1 + \kappa h}{h} = O(h) \ as \ \ h \to 0,$$

and

$$h\sqrt{n}\left[\hat{\kappa}_1 - \frac{1-\exp(-\kappa h)}{h}\right] \xrightarrow{d} N(0, 1-\exp(-2\kappa h)). \tag{2.4.14}$$

For the OU process with an unknown mean,

$$BIAS(\hat{\kappa}_1) = -\frac{H}{h} + \frac{1+3\exp(-\kappa h)}{T} + o(T^{-1}), \tag{2.4.15}$$

For the OU process with a known mean,

$$BIAS(\hat{\kappa}_1) = -\frac{H}{h} + \frac{2\exp(-\kappa h)}{T} + o(T^{-1}), \tag{2.4.16}$$

where $\frac{1+3\exp(-\kappa h)}{T} + o(T^{-1})$ and $\frac{2\exp(-\kappa h)}{T} + o(T^{-1})$ are the estimation biases in the two models, respectively. In both models, the discretization bias has the following form:

$$\frac{-H}{h} = -\frac{\exp(-\kappa h) - 1 + \kappa h}{h}. \tag{2.4.17}$$

Remark 2.4.1 *From (2.4.14) the asymptotic variance for $\hat{\kappa}_1$ is*

$$AsyVar(\hat{\kappa}_1) = \frac{1-\exp(-2\kappa h)}{Th}.$$

Remark 2.4.2 *The estimation bias is always positive in both models. If $\kappa h \in (0,3]$*

which is empirically realistic, the discretization bias may be written as

$$
\begin{aligned}
\frac{-H}{h} &= -\kappa^2 h \sum_{i=2}^{\infty} \frac{(-\kappa h)^{i-2}}{i!} \\
&= -\kappa^2 h \sum_{j=2,4,\cdots} \frac{(-\kappa h)^{j-2}}{(j+1)!} (j+1-\kappa h) \\
&< 0.
\end{aligned}
$$

This means that the discretization bias has sign opposite to that of the estimation bias.

Remark 2.4.3 *For the unknown mean model, if $T < h(1+3\phi)/(\kappa h+\phi-1)$, the estimation bias is larger than the discretization bias in magnitude because this condition is equivalent to*

$$\frac{1+3\exp(-\kappa h)}{T} > \frac{\kappa h+\exp(-\kappa h)-1}{h}.$$

Further

$$
\begin{aligned}
h(1+3\phi)/(\kappa h+\phi-1) &= \frac{h(1+3(1-\kappa h+O(h^2)))}{\frac{1}{2}\kappa^2 h^2 - \frac{1}{6}\kappa^3 h^3 + O(h^4)} \\
&= \frac{2}{\kappa^2 h}(4-3\kappa h+O(h^2))) \left(1-\frac{1}{3}\kappa h+O(h^2)\right)^{-1} \\
&= \frac{2}{\kappa^2 h}(4-3\kappa h+O(h^2))) \left(1+\frac{1}{3}\kappa h+O(h^2)\right) \\
&= \frac{8}{\kappa^2 h}(1+O(h)).
\end{aligned}
$$

In empirically relevant cases, $8/(\kappa^2 h)$ is likely to take very large values, thereby requiring very large values of T before the estimation bias is smaller than the discretization bias. For example, if $\kappa = 0.1$ and $h = 1/12$, $T > 9{,}600$ years are needed for the bias to be smaller. The corresponding result for the known mean case is $2h\phi/(\kappa h+\phi-1) = \left(4/(\kappa^2 h)\right)(1+O(h))$ and again large values of T are required to reduce the relative magnitude of the estimation bias.

Similarly, the second order expansion (i.e. the trapezoidal method) gives the estimator

$$\hat{\kappa}_2 = -\hat{A} = -\frac{2}{h}[\hat{F} - I][\hat{F} + I]^{-1} = \frac{2(1-\hat{\phi})}{h(1+\hat{\phi})}, \tag{2.4.18}$$

for which we have the following result.

Theorem 2.4.3 *Assuming $\kappa > 0$, when h is fixed, and $n \to \infty$, we have*

$$\hat{\kappa}_2 - \kappa \xrightarrow{p} \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))} - \kappa = O(h^2) \text{ as } h \to 0,$$

and

$$h\sqrt{n}\left[\hat{\kappa}_2 - \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))}\right] \xrightarrow{d} N\left(0, \frac{16(1-\exp(-\kappa h))}{(1+\exp(-\kappa h))^3}\right). \tag{2.4.19}$$

For the OU process with an unknown mean,

$$BIAS(\hat{\kappa}_2) = \nu + \frac{8}{T(1+\exp(-\kappa h))} + o(T^{-1}). \tag{2.4.20}$$

For the OU process with a known mean,

$$BIAS(\hat{\kappa}_2) = \nu + \frac{4}{T(1+\exp(-\kappa h))} + o(T^{-1}), \tag{2.4.21}$$

where $\frac{8}{T(1+\exp(-\kappa h))} + o(T^{-1})$ and $\frac{4}{T(1+\exp(-\kappa h))} + o(T^{-1})$ are the two estimation biases. In both models, the discretization bias has the form

$$\nu = -\kappa + \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))} = O(h^2). \tag{2.4.22}$$

Remark 2.4.4 *From (2.4.19) the asymptotic variance for $\hat{\kappa}_2$ is*

$$AsyVar(\hat{\kappa}_2) = \frac{16(1-\exp(-\kappa h))}{Th(1+\exp(-\kappa h))^3}.$$

Remark 2.4.5 *The estimation bias is always positive in both models. If $\kappa h \in (0, 2]$,* *the discretization bias may be written as*

$$\begin{aligned}
\nu &= -\kappa + \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))} = \frac{-\kappa}{1+\exp(-\kappa h)} \sum_{i=3}^{\infty} \frac{(i-2)(-\kappa h)^{i-1}}{i!} \\
&= \frac{-\kappa}{1+\exp(-\kappa h)} \sum_{j=3,5,\cdots} \frac{(-\kappa h)^{j-1}}{(j+1)!} ((j-2)(j+1) - \kappa h(j-1)) \\
&< 0.
\end{aligned}$$

Hence, the discretization bias has the opposite sign of the estimation bias.

Remark 2.4.6 *For the unknown mean model, if* $T < 8h/(\kappa h(1+\phi) - 2(1-\phi))$, *the estimation bias is larger than the discretization bias in magnitude because this condition is equivalent to*

$$\frac{8}{T(1+\exp(-\kappa h))} > \kappa - \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))}.$$

Further

$$\begin{aligned}
\frac{8h}{\kappa h(1+\phi) - 2(1-\phi)} &= \frac{8h}{\kappa h(2-\kappa h + \frac{1}{2}\kappa^2 h^2 + O(h^3)) - 2(\kappa h - \frac{1}{2}\kappa^2 h^2 + \frac{1}{6}\kappa^3 h^3 + O(h^4))} \\
&= 8h\left(\frac{1}{6}\kappa^3 h^3 + O(h^4)\right)^{-1} = \frac{48}{\kappa^3 h^2}(1+O(h))^{-1} \\
&= \frac{48}{\kappa^3 h^2}(1+O(h)).
\end{aligned}$$

Again, in empirically relevant cases, $48/(\kappa^3 h^2)$ *is likely to take very large values thereby requiring very large values of* T *before the estimation bias is smaller than the discretization bias. For example, if* $\kappa = 0.1$ *and* $h = 1/12$, $T > 6,912,000$ *years are needed for the bias to be smaller. Hence the estimation bias is inevitably much larger than the discretization bias in magnitude for all realistic sample spans* T.

Remark 2.4.7 *It has been argued in the literature that ML should be used whenever it is available and the likelihood function should be accurately approximated when it is not available analytically; see Durham and Gallant (2002) and Aït-Sahalia (2002) for various techniques to accurately approximate the likelihood function.*

From the results in Theorems 2.4.2 and 2.4.3 we can show that the total bias of the MLE based on the exact discrete time model is bigger than that based on the Euler and the trapezoidal approximation. For example, for the estimator based on the trapezoidal approximation, considering $\nu = O(h^2)$ as $h \to 0$, when the model is the OU process with an unknown mean,

$$
\begin{aligned}
|BIAS(\hat{\kappa}_{ML})| - |BIAS(\hat{\kappa}_2)| &= \frac{5+2e^{\kappa h}+e^{2\kappa h}}{2T} - \left|\frac{8}{T(1+e^{-\kappa h})} + v\right| + o(T^{-1}) \\
&= \frac{5+2e^{\kappa h}+e^{2\kappa h}}{2T} - \frac{8}{T(1+e^{-\kappa h})} - v + o(T^{-1}) \\
&= \frac{(1-\phi)^2(1+5\phi)}{2T\phi^2(1+\phi)} - v + o(T^{-1}) \\
&> 0.
\end{aligned}
$$

Using the same method, it is easy to prove the result still holds for the OU process with an known mean. Similarly, one may show that

$$
|BIAS(\hat{\kappa}_{ML})| - |BIAS(\hat{\kappa}_1)| > 0,
$$

in both models.

Remark 2.4.8 *The two approximate estimators reduce the total bias over the exact ML and also the asymptotic variance when $\kappa > 0$. This is because*

$$
AsyVar(\hat{\kappa}_{ML}) - AsyVar(\hat{\kappa}_1) = \frac{1-\phi^2}{Th\phi^2} - \frac{1-\phi^2}{Th} > 0. \tag{2.4.23}
$$

and

$$
\begin{aligned}
AsyVar(\hat{\kappa}_{ML}) - AsyVar(\hat{\kappa}_2) &= \frac{1-\phi^2}{Th\phi^2} - \frac{16(1-\phi)}{Th(1+\phi)^3} && (2.4.24) \\
&= \frac{(1-\phi)^3}{Th\phi^2}\frac{\left(\phi^2+6\phi+1\right)}{(1+\phi)^3} > 0. && (2.4.25)
\end{aligned}
$$

In consequence, the two approximate methods are preferred to the exact ML for estimating the mean reversion parameter in the univariate setting. Of course, the *two approximate methods do NOT improve the asymptotic efficiency of the MLE. This is because the asymptotic variance of the MLE is based on large T asymptotics whereas the asymptotic variance of $\hat{\kappa}_1$ and $\hat{\kappa}_2$ is based on large n asymptotics and the two approximate estimators are inconsistent with fixed h. Nevertheless, equations (2.4.14) and (2.4.19) seem to indicate that in finite (perhaps very large finite) samples, the inconsistent estimators may lead to smaller variances than the MLE, which will be verified by simulations.*

Remark 2.4.9 *Comparing Theorem 2.4.2 and Theorem 2.4.3, it is easy to see the estimator (2.4.18) based on the trapezoidal approximation leads to a smaller discretization bias than the estimator (2.4.13) based on the Euler approximation. However, when $\kappa h > 0$ and hence $\phi = e^{-\kappa h} \in (0,1)$, the gain in the discretization error is earnt at the expense of an increase in the estimation error. For the OU process with an unknown mean,*

$$\begin{aligned}\text{estimation bias }(\hat{\kappa}_2) - \text{estimation bias }(\hat{\kappa}_1) &= \frac{8}{T(1+e^{-\kappa h})} - \frac{1+3e^{-\kappa h}}{T} + o(T^{-1}) \\ &= \frac{(1-\phi)(7+3\phi)}{T(1+\phi)} + o(T^{-1}) > 0.\end{aligned}$$

Similarly, for the OU process with a known mean,

$$\begin{aligned}\text{estimation bias }(\hat{\kappa}_2) - \text{estimation bias }(\hat{\kappa}_1) &= \frac{4}{T(1+e^{-\kappa h})} - \frac{2e^{-\kappa h}}{T} + o(T^{-1}) \\ &= \frac{(1-\phi)(4+2\phi)}{T(1+\phi)} + o(T^{-1}) > 0.\end{aligned}$$

Since the sign of the discretization bias is opposite to that of the estimation bias, and the trapezoidal rule makes the discretization bias closer to zero than the Euler approximation, we have the following result in both models.

$$|BIAS(\hat{\kappa}_2)| - |BIAS(\hat{\kappa}_1)| > 0.$$

Remark 2.4.10 *The estimator based on the Euler method leads not only to a smaller* *bias but also to a smaller variance than that based on the trapezoidal method when* $\kappa > 0$*. This is because*

$$
\begin{aligned}
AsyVar(\hat{\kappa}_2) - AsyVar(\hat{\kappa}_1) &= \frac{16(1-\phi)}{Th(1+\phi)^3} - \frac{1-\phi^2}{Th} \\
&= \frac{(1-\phi)^2(3+\phi)[4+(1+\phi)^2]}{Th(1+\phi)^3} > 0.
\end{aligned}
$$

In consequence, the Euler method is preferred to the trapezoidal method and exact ML for estimating the mean reversion parameter in the univariate setting.

2.5 Bias in General Univariate Models

2.5.1 Univariate square root model

The square root model, also known as the Cox, Ingersoll and Ross (1985, CIR hereafter) model, is of the form

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\sqrt{X(t)}dW(t). \tag{2.5.1}$$

If $2\kappa\mu/\sigma^2 > 1$, Feller (1951) showed that the process is stationary, the transitional distribution of cX_t given X_{t-1} is non-central $\chi^2_\nu(\lambda)$ with the degree of freedom $\nu = 2\kappa\mu\sigma^{-2}$ and the non-central component $\lambda = cX_{t-1}e^{-\kappa h}$, where $c = 4\kappa\sigma^{-2}(1 - e^{-\kappa h})^{-1}$. Since the non-central χ^2-density function is an infinite series involving the central χ^2 densities, the explicit expression of the MLE for $\theta = (\kappa, \mu, \sigma)$ is not attainable.

To obtain a closed-form expression for the estimator of θ, we follow Tang and Chen (2009) by using the estimator of Nowman. The Nowman discrete time representation of the square root model is

$$X_t = \phi_1 X_{t-1} + (1-\phi_1)\mu + \sigma\sqrt{X_{t-1}\frac{1-\phi_1^2}{2\kappa}}\varepsilon_t,$$

where $\phi_1 = e^{-\kappa h}$, $\varepsilon_t \sim iid\ N(0,1)$ and h is the sampling interval. Hence, Nowman's estimator of κ is

$$\hat{\kappa}_{Nowman} = -\frac{1}{h}\ln(\hat{\phi}_1),$$

where

$$\hat{\phi}_1 = \frac{n^{-2}\sum_{t=1}^{n} X_t \sum_{t=1}^{n} X_{t-1}^{-1} - n^{-1}\sum_{t=1}^{n} X_t X_{t-1}^{-1}}{n^{-2}\sum_{t=1}^{n} X_{t-1} \sum_{t=1}^{n} X_{t-1}^{-1} - 1}.$$

For the stationary square root process, Tang and Chen (2009) derived explicit expressions to approximate $E(\hat{\phi}_1 - \phi_1)$ and $Var(\hat{\phi}_1)$. Using the following relations,

$$E(\hat{\kappa}_{Nowman} - \kappa) = -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) - \frac{1}{2\phi_1^2}E(\hat{\phi}_1 - \phi_1)^2 + O(n^{-3/2})\right], \quad (2.5.2)$$

and

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}[Var(\hat{\phi}_1) + O(n^{-2})],$$

they further obtained the approximations to $E(\hat{\kappa}_{Nowman} - \kappa)$ and $Var(\hat{\kappa}_{Nowman})$. With a fixed h and $n \to \infty$ they derived the asymptotic distribution of $\sqrt{n}(\hat{\kappa}_{Nowman} - \kappa)$. The fact that the mean of the asymptotic distribution is zero implies that the Nowman method causes no discretization bias for estimating κ.

The estimator of κ based on the Euler approximation also has a closed form expression under the square root model. The Euler discrete time model is

$$X_t = \phi_2 X_{t-1} + (1-\phi_2)\mu + \sigma\sqrt{X_{t-1}h}\varepsilon_t,$$

where $\phi_2 = (1-\kappa h)$. Hence, the Euler scheme estimator of κ is

$$\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_2 - 1),$$

where

$$\hat{\phi}_2 = \frac{n^{-2}\sum_{t=1}^{n} X_t \sum_{t=1}^{n} X_{t-1}^{-1} - n^{-1}\sum_{t=1}^{n} X_t X_{t-1}^{-1}}{n^{-2}\sum_{t=1}^{n} X_{t-1} \sum_{t=1}^{n} X_{t-1}^{-1} - 1}.$$

Obviously $\hat{\phi}_2 = \hat{\phi}_1$. Hence, $\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_1 - 1)$. Considering $\phi_1 = e^{-\kappa h} = 1 -$

$\kappa h + \sum_{i=2}^{\infty}(-\kappa h)^i/i!$, the finite sample bias for $\hat{\kappa}_{Euler}$ can be expressed as

$$E(\hat{\kappa}_{Euler} - \kappa) = -\frac{1}{h}E(\hat{\phi}_1 - \phi_1) - \frac{1}{h}H, \tag{2.5.3}$$

where

$$-\frac{1}{h}H = -\frac{1}{h}\sum_{i=2}^{\infty}(-\kappa h)^i/i! = O(h), \quad \text{as} \quad h \to 0,$$

which is the discretization bias caused by discretizing the dirft function. Since the asymptotic mean of $\sqrt{n}(\hat{\phi}_1 - \phi_1)$ and hence the asymptotic mean of $\sqrt{n}(\hat{\kappa}_{Euler} - \kappa + \frac{1}{h}H)$ is zero for a fixed h and $n \to \infty$, the Euler discretization of the diffusion function introduces no discretization bias to κ under the square root model.

Furthermore, the finite sample variance for $\hat{\kappa}_{Euler}$ is

$$Var(\hat{\kappa}_{Euler}) = \frac{1}{h^2}Var(\hat{\phi}_1).$$

If $\kappa > 0$, $\phi_1 = e^{-\kappa h} < 1$. When h is fixed, we have

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}\left[Var(\hat{\phi}_1) + O(n^{-2})\right] > \frac{1}{h^2}Var(\hat{\phi}_1) = Var(\hat{\kappa}_{Euler}),$$

leading to

$$\frac{Var(\hat{\kappa}_{Euler})}{Var(\hat{\kappa}_{Nowman})} = \phi_1^2 + O(n^{-1}) < 1. \tag{2.5.4}$$

According to (2.5.4), the Euler scheme always gains over Nowman's method in terms of variance. The smaller is ϕ_1, the larger the gain.

Tang and Chen (2009) obtained a bias formula of $E(\hat{\phi}_1 - \phi_1)$ for the Nowman estimator under the square root model. Unfortunately, the expression is too complex to be used to determine the sign of the bias analytically. However, the simulation results reported in the literature (Phillips and Yu, 2009, for example) and in our own simulations reported in Section 6 suggest that $E(\hat{\kappa}_{Euler} - \kappa) > 0$. Since $H > 0$, (2.5.3) implies that

$$E(\hat{\phi}_1 - \phi_1) < 0,$$

and the estimation bias $-\frac{1}{h}E(\hat{\phi}_1-\phi_1)$ dominates the discretization bias $-\frac{1}{h}H$ in the Euler approximation. Consequently, the negative discretization bias $-\frac{1}{h}H$ reduces the total bias in the Euler method. Consequently, the bias in $\hat{\kappa}_{Nowman}$ is larger than that in $\hat{\kappa}_{Euler}$ because

$$\begin{aligned}
E(\hat{\kappa}_{Nowman}-\kappa) &= -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1-\phi_1)-\frac{1}{2\phi_1^2}E(\hat{\phi}_1-\phi_1)^2+O(n^{-3/2})\right] \\
&\geq -\frac{1}{h}\frac{1}{\phi_1}E(\hat{\phi}_1-\phi_1) \\
&\geq -\frac{1}{h}E(\hat{\phi}_1-\phi_1)-\frac{1}{h}H = E(\hat{\kappa}_{Euler}-\kappa).
\end{aligned}$$

The Milstein scheme is another popular approximation approach. For the square root model, the discrete time model obtained by the Milstein scheme is given by

$$X_t = X_{t-1}+\kappa(\mu-X_{t-1})h+\sigma\sqrt{X_{t-1}h}\varepsilon_t+\frac{1}{4}\sigma^2 h\left[\varepsilon_t^2-1\right]. \tag{2.5.5}$$

Let $a=\sigma\sqrt{X_{t-1}h}$, $b=\frac{1}{4}\sigma^2 h$, $Y_t=X_t-X_{t-1}-\kappa(\mu-X_{t-1})h+\frac{1}{4}\sigma^2 h$, then Equation (2.5.5) can be represented by

$$Y_t = a\varepsilon_t+b\varepsilon_t^2 = b\left[\left(\varepsilon_t+\frac{a}{2b}\right)^2-\frac{a^2}{4b^2}\right].$$

Since $\varepsilon_t \sim iid\, N(0,1)$, $Z=\left(\varepsilon_t+\frac{a}{2b}\right)^2$ follows a noncentral χ^2 distribution with 1 degree of freedom and noncentrality parameter $\lambda=\frac{a^2}{4b^2}$. Elerian (1998) showed that the density of Z may be expressed as

$$f(z)=\frac{1}{2}\exp\left\{-\frac{\lambda+z}{2}\right\}\left(\frac{z}{\lambda}\right)^{-1/4} I_{-1/2}\left(\sqrt{\lambda z}\right), \tag{2.5.6}$$

where

$$I_{-1/2}(x)=\sqrt{\frac{2}{x}}\sum_{i=0}^{\infty}\frac{(x/2)^{2i}}{i!\Gamma(j+0.5)}=\sqrt{\frac{1}{2\pi x}}\{\exp(x)+\exp(-x)\}.$$

This expression may be used to compute the log-likelihood function of the approximate model (2.5.5). Unfortunately, the ML estimator does not have a closed form expression and it is therefore difficult to examine the relative performance of the bias and the variance using analytic methods. The performance of the Milstein scheme is therefore compared to other methods in simulations.

2.5.2 Diffusions with linear drift

We consider the following general diffusion process with a linear drift

$$dX(t) = \kappa(\mu - X(t))dt + \sigma q(X(t);\psi)dW(t), \tag{2.5.7}$$

as a generalization to the Vasicek and the square root models, where $\sigma q(X(t);\psi)$ is a general diffusion function with parameters ψ, and $\theta = (\kappa, \mu, \sigma, \psi) \in R^d$ is the unknown parameter vector. This model include the well known Constant Elasticity of Variance (CEV) model, such as the Chan, et al (1992, CKLS) model, as a special case. In this general case, the transitional density is not analytically available.

The Nowman approximate discrete model is

$$X_t = \phi_1 X_{t-1} + (1-\phi_1)\mu + \sigma q(X_{t-1};\psi)\sqrt{\frac{1-\phi_1^2}{2\kappa}}\varepsilon_t, \tag{2.5.8}$$

The Euler approximate discrete model is

$$X_t = \phi_2 X_{t-1} + (1-\phi_2)\mu + \sigma q(X_{t-1};\psi)\sqrt{h}\varepsilon_t. \tag{2.5.9}$$

Theorem 2.5.1 *For Model (2.5.7), the MLE of κ based on the Nowman approximation is*

$$\hat{\kappa}_{Nowman} = -\frac{1}{h}\ln(\hat{\phi}_1),$$

where $\hat{\phi}_1$ is the ML estimator for ϕ_1 in (2.5.8). The MLE of κ based on the Euler approximation is

$$\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_2 - 1),$$

where $\hat{\phi}_2$ is the ML estimator for ϕ_2 in (2.5.9). Then we have

$$\hat{\phi}_2 = \hat{\phi}_1.$$

Remark 2.5.1 *The ML estimator of ϕ_1 does not have a closed-form expression. Neither does the ML estimator of ϕ_2. So numerical calculations are needed for comparisons. However, according to Theorem 2.5.1, even without a closed-form solution, we can still establish the equivalence of $\hat{\phi}_1$ and $\hat{\phi}_2$. After $\hat{\phi}_1$ and $\hat{\phi}_2$ are found numerically, one may find the estimators of κ by using the relations $\hat{\kappa}_{Nowman} = -\frac{1}{h}\ln(\hat{\phi}_1)$ and $\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_2 - 1)$.*

To compare the magnitude of the bias in $\hat{\kappa}_{Nowman}$ to that of $\hat{\kappa}_{Euler}$, no general analytic result is available. However, under some mild conditions, comparison is possible. In particular, we make the following three assumptions. Assumption 1: $\hat{\phi}_1 - \phi_1 \sim O_p(n^{-1/2})$; Assumption 2: $E(\hat{\phi}_1 - \phi_1) < 0$; Assumption 3: $-\frac{1}{h}E(\hat{\phi}_1 - \phi_1) > -\frac{1}{h}H$, i.e., the estimation bias dominates the discretization bias in Euler approximation. Under Assumption 1, we get

$$E(\hat{\kappa}_{Nowman} - \kappa) = -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) - \frac{1}{2\phi_1^2}E(\hat{\phi}_1 - \phi_1)^2 + O(n^{-3/2})\right],$$

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}[Var(\hat{\phi}_1) + O(n^{-2})],$$

$$E(\hat{\kappa}_{Euler} - \kappa) = -\frac{1}{h}E(\hat{\phi}_1 - \phi_1) - \frac{1}{h}H,$$

and

$$Var(\hat{\kappa}_{Euler}) = \frac{1}{h^2}Var(\hat{\phi}_1),$$

where $H = \sum_{i=2}^{\infty}(-\kappa h)^i/i! = O(h^2)$.

If $\kappa > 0$, $\hat{\kappa}_{Euler}$ has a smaller finite sample variance than $\hat{\kappa}_{Nowman}$ because

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}\left[Var(\hat{\phi}_1) + O(n^{-2})\right] \geq \frac{1}{h^2}Var(\hat{\phi}_1) = Var(\hat{\kappa}_{Euler}).$$

Under Assumptions 1, 2, 3, $\hat{\kappa}_{Euler}$ has a smaller bias than $\hat{\kappa}_{Nowman}$ because

$$\begin{aligned}
E(\hat{\kappa}_{Nowman} - \kappa) &= -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) - \frac{1}{2\phi_1^2}E(\hat{\phi}_1 - \phi_1)^2 + O(n^{-3/2})\right] \\
&\geq -\frac{1}{h}\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) \\
&\geq -\frac{1}{h}E(\hat{\phi}_1 - \phi_1) - \frac{1}{h}H = E(\hat{\kappa}_{Euler} - \kappa).
\end{aligned}$$

2.6 Simulation Studies

2.6.1 Linear models

To examine the performance of the proposed bias formulae and to compare the two alternative approximation scheme in multivariate diffusions, we estimate $\kappa = -A$ in the bivariate model with a known mean:

$$dX_t = AX_t dt + \Sigma dW_t, \;\; X_0 = 0, \tag{2.6.1}$$

where W_t is the standard bivariate Brownian motion whose components are independent, and

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}, \kappa = -A = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}.$$

Since A is triangular, the parameters are all identified. While keeping other parameters fixed, we let κ_{22} take various values over the interval (0,3], which covers empirically reasonable values of κ_{22} that apply for data on interest rates and volatilities. The mean reversion matrix is estimated with 10 years of monthly data. The experiment is replicated 10,000 times. Both the actual total bias and the actual standard deviation are computed across 10,000 replications. The actual total bias is split into two parts — discretization bias and estimation bias — as follows. The estimation bias is calculated as H/h and $-v$ as in (2.3.13) and (2.3.19) for the two approximate methods. The estimation bias is calculated as:

$$\text{estimation bias} = \text{actual total bias - discretization bias}$$

Figure 2.1 plots the biases of the estimate of each element in the mean reversion matrix κ, based on the Euler method, as a function of the true value of κ_{22}. Four biases are plotted, the actual total bias, the approximate total bias given by the formula in (2.3.13), the discretization bias H/h as in (2.3.13), and the estimation bias.

Several features are apparent in the figure. First, the actual total bias in all cases is large, especially when the true value of κ_{22} is small. Second, except for κ_{12} whose discretization bias is zero, the sign of the discretization bias for the other parameters is opposite to that of the estimation bias. Not surprisingly, in these cases, the actual total bias of estimator (2.3.10) is smaller than the estimation bias. The discretization bias for κ_{12} is zero because it is assumed that the true value is zero. In the bivariate set-up, however, it is possible that the sign of the discretization bias for the other parameters is the same as that of the estimation bias (for example when $\kappa_{12} = 5$ and $\kappa_{21} = -0.5$). Third, the bias in all parameters is sensitive to the true value of κ_{22}. Finally, the bias formula (2.3.13) generally works well in all cases.

Figure 2.2 plots the biases of the estimate of each element in the mean reversion matrix κ, based on the trapezoidal method, as a function of the true value of κ_{22}. Four biases are plotted, the actual total bias, the approximate total bias given by the formula in (2.3.19), the discretization bias $-\nu$ as in (2.3.19), and the estimation bias. In all cases, the discretization bias is closer to zero than that based on the Euler approximation. This suggests that the trapezoidal method indeed reduces the discretization bias. Moreover, the bias formula (2.3.19) generally works well in all cases.

The performance of the two approximation methods is compared in Figure 2.3, where the actual total bias of the estimators given by (2.3.10) and (2.3.18) is plotted. It seems that the bias of the estimator obtained from the trapezoidal approximation

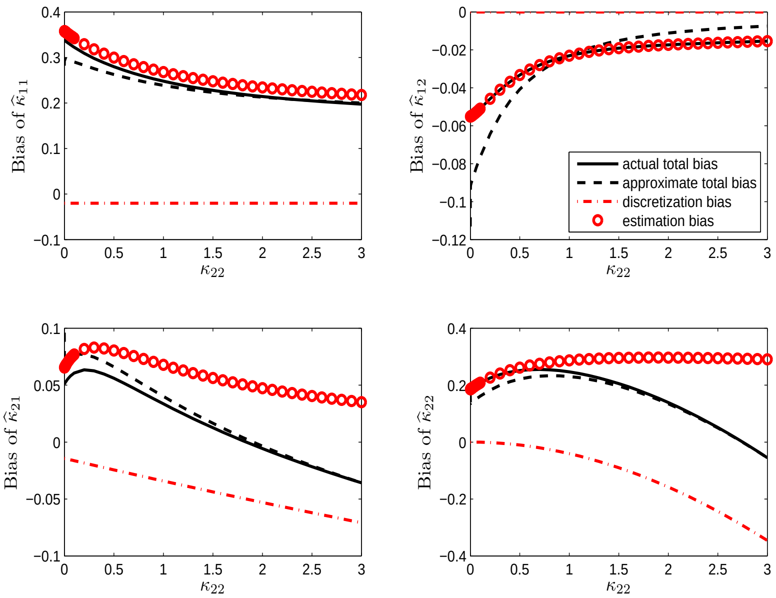

Figure 2.1: The bias of the elements in $\hat{A}$ in Model (2.6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the Euler method. The solid line is the actual total bias; the broken line is the approximate total bias according to the formula (2.3.13); the dashed line is the discretization bias H/h; the point line is the estimation bias. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

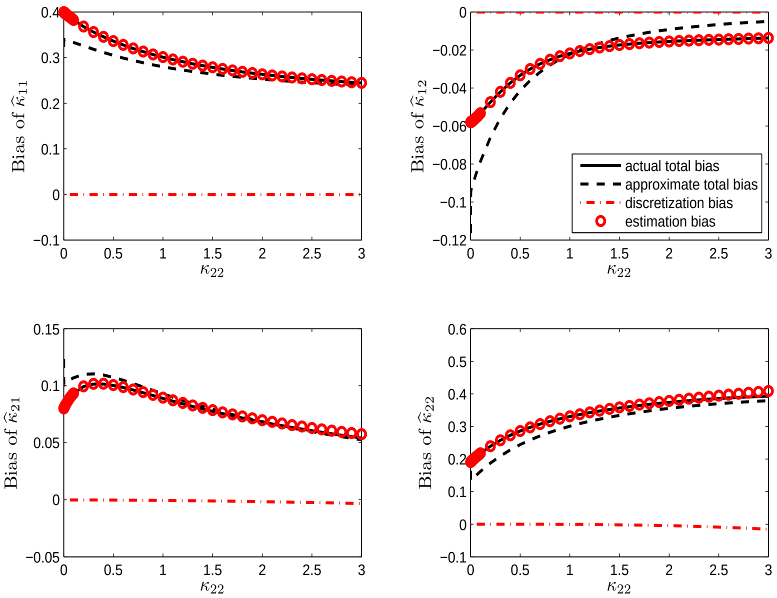

Figure 2.2: The bias of the elements in $\hat{A}$ in Model (2.6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the trapezoidal method. The solid line is the actual total bias; the broken line is the approximate bias according to the formula (2.3.13); the dashed line is the discretization bias $-v$; the point line is the estimation bias. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

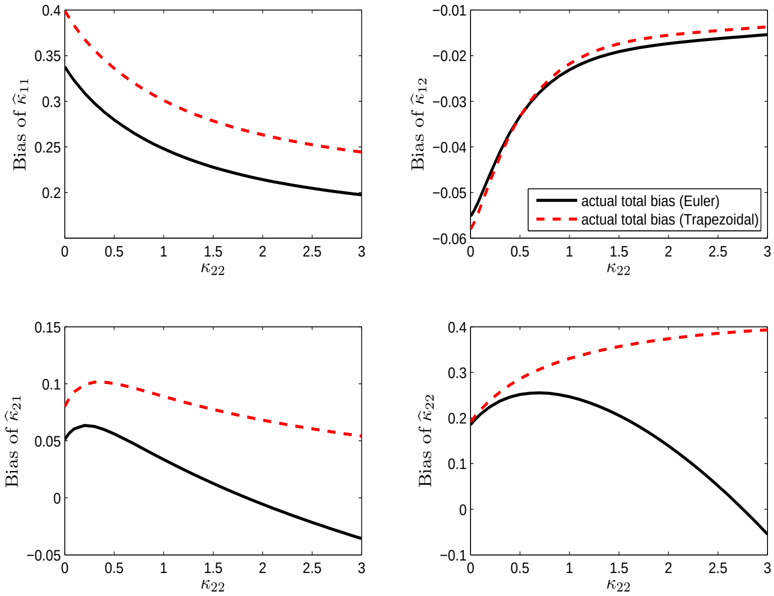

Figure 2.3: The bias of the elements in $\hat{A}$ in Model (2.6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the Euler and the trapezoidal methods, respectively. The solid line is the actual total bias for the Euler method; the broken line is the actual total bias for the trapezoidal method. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

is larger than that from the Euler approximation for all parameters except κ_{12}. For κ_{12}, the performance of the two methods are very close with the Euler method being slightly worse when κ_{22} is large.

Figure 2.4 plots the actual standard deviations for the two approximate estimators, (2.3.10) and (2.3.18) as a function of κ_{22}. We notice that, for all the parameters, the standard deviation of the Euler method is smaller than that of the trapezoidal method. The percentage difference can be as high as 20%.

We also design an experiment to check the performance of the alternative estimators in the univariate case. Data are simulated from the univariate OU process with a known mean

$$dX(t) = -\kappa X(t)dt + \sigma dW(t), \quad X(0) = 0. \tag{2.6.2}$$

Figure 2.5 reports the bias in $\widehat{\kappa}$ obtained from the Euler method and the trapezoidal method in the OU process with a known mean. Three biases are plotted: the

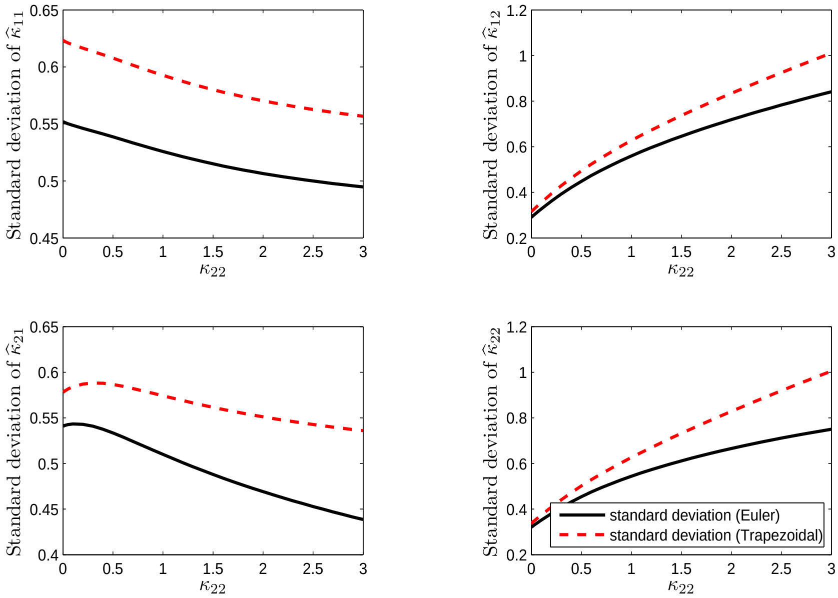

Figure 2.4: The standard deviation of the elements in $\hat{A}$ in Model (2.6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the Euler and the trapezoidal methods, respectively. The solid line is the standard deviation for the Euler method; the broken line is the standard deviation for the trapezoidal method. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

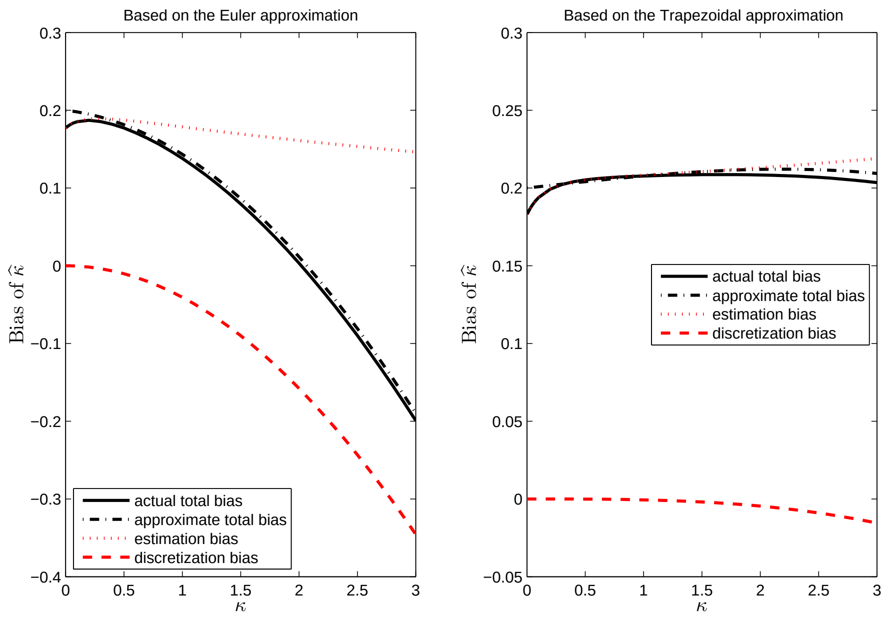

Figure 2.5: The bias of the κ estimates in the univariate model as a function of κ at the monthly frequency and $T = 10$ for the two approximate methods. The left panel is for the Euler method and the right panel is for the trapezoidal method. The solid line is the actual total bias; the dashed line is the approximate total bias; the dotted line is the estimation bias; the broken line is the discretization bias.

actual total bias, the estimation bias and the discretization bias. Figure 2.6 compares the bias in $\widehat{\kappa}$ obtained from the exact ML methods with that of the two approximate methods. Several conclusions may be drawn from these two Figures. First, our bias formula provides a good approximation to the actual total bias. Second, for the two approximate estimators, (2.4.13) and (2.4.18), the sign of the discretization bias is opposite to that of the estimation bias. Third, while the trapezoidal method leads to a smaller discretization bias than the Euler method, it has a larger estimation bias. Finally, the actual total bias for the Euler method is smaller than that of the trapezoidal method and both methods lead to a smaller total bias than the exact ML estimator (2.4.8).

Figure 2.7 reports the standard deviations for estimators (2.4.8), (2.4.13) and (2.4.18). It is easy to find that the standard deviations of estimator (2.4.13) is the smallest among those of all estimators. The standard deviations of estimator (2.4.18) are almost the same with those from the exact ML estimator (2.4.8), but

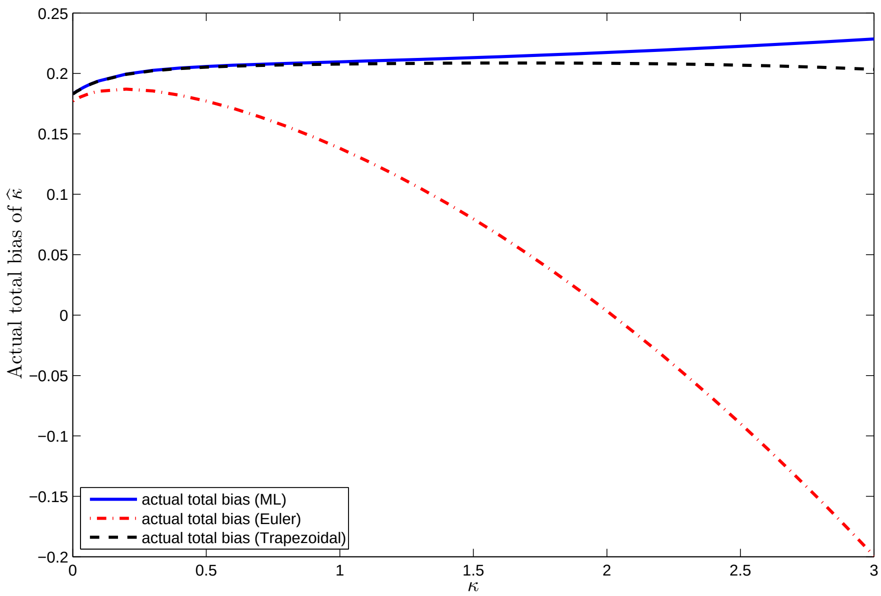

Figure 2.6: The actual total bias of the κ estimates in the univariate model as a function of κ at the monthly frequency and $T = 10$ for the two approximate methods and the exact ML. The solid line is for the exact ML; the dashed line is for the Euler method; the broken line is for the trapezoidal method.

smaller when κ is bigger than 1. Considering the sample size is 120, we can roughly say that, focusing on bias and standard deviation, the estimator (2.4.13) from the Euler approximation is better than the other estimators in comparatively small sample sizes.

2.6.2 Square root model

For the square root model, we designed an experiment to compare the performance of the various estimation methods, including the exact ML, the Euler scheme, the Nowman scheme and the Milstein scheme. In all cases we fix $h = 1/12$, $T = 120$, $\mu = 0.05$, $\sigma = 0.05$, but vary the value of κ from 0.05 to 0.5. These settings correspond to 10 years of monthly data in the estimation of κ. The experiment is replicated 10,000 times.

Table 1 reports the bias, the standard error (Std err), and the root mean square error (RMSE) of κ for all estimation methods, obtained across 10,000 replications. Several conclusions emerge from the table. First, all estimation methods suffer from

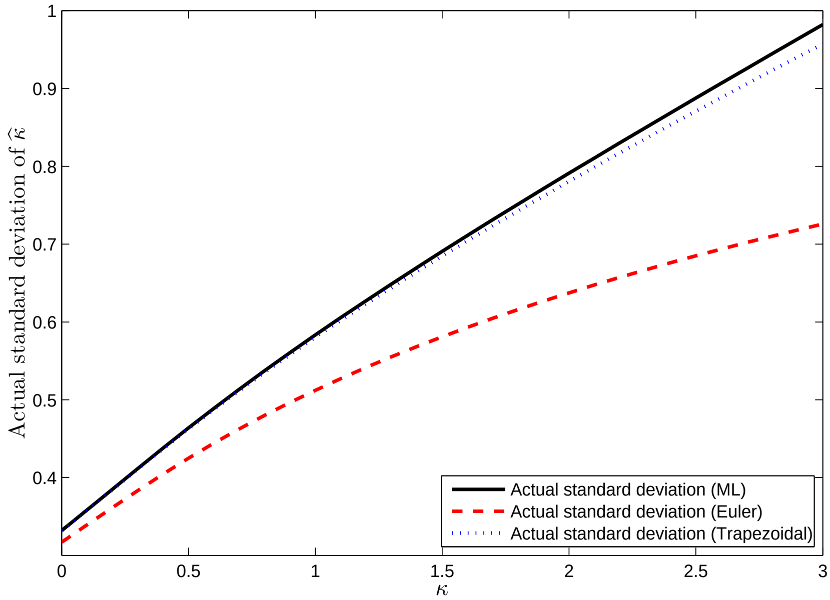

Figure 2.7: The standard deviation of the κ estimates in the univariate model as a function of κ at the monthly frequency and $T = 10$. The solid line is for the exact ML; the broken line is for the Euler method; the dotted line is for the trapezoidal method.

a serious bias problem. Second, the Euler scheme performs best both in terms of bias and variance. Third, the ratios of the standard error of $\widehat{\kappa}_{Euler}$ and that of $\widehat{\kappa}_{Norman}$ are 0.9958, 0.9917, 0.9835, 0.9592 when κ is 0.05, 0.1, 0.2, 0.5, respectively. The ratio decreases as κ increases, as predicted in (2.5.4). Finally, although the bias for the Milstein method is larger than that for the Euler method, the variances for these two methods are very close.

2.7 Conclusions

This paper provides a framework for studying the implications of different discretization schemes in estimating the mean reversion parameter in both multivariate and univariate diffusion models with a linear drift function. The approach includes the Euler method and the trapezoidal method as special cases, an asymptotic theory is developed, and finite sample bias comparisons are conducted using analytic approximations. Bias is decomposed into a discretization bias and an estimation bias.

Table 2.1: Exact and approximate ML estimation of κ from the square root model using 120 monthly observations. The experiment is replicated 10,000 times.

Method	Exact	Euler	Nowman	Milstein
		$\kappa = 0.05$		
Bias	.1156	.1126	.1152	.1132
Std err	.2251	.2205	.2249	.2206
RMSE	.2531	.2476	.2526	.2480
		$\kappa = 0.1$		
Bias	.1392	.1342	.1387	.1350
Std err	.2670	.2590	.2668	.2592
RMSE	.3011	.2917	.3007	.2922
		$\kappa = 0.2$		
Bias	.1615	.1529	.1610	.1538
Std err	.3178	.3070	.3178	.3068
RMSE	.3565	.3430	.3562	.3432
		$\kappa = 0.5$		
Bias	.1869	.1625	.1862	.1639
Std err	.4210	.3999	.4209	.3993
RMSE	.4607	.4317	.4603	.4316

It is shown that the discretization bias is of order $O(h)$ for the Euler method and $O(h^2)$ for the trapezoidal method, respectively, whereas the estimation bias is of the order of $O(T^{-1})$. Since in practical applications in finance it is very likely that h is much smaller than $1/T$, estimation bias is likely to dominate discretization bias.

Applying the multivariate theory to univariate models gives several new results. First, it is shown that in the Euler and trapezoidal methods, the sign of the discretization bias is opposite that of the estimation bias for practically realistic cases. Consequently, the bias in the two approximate method is smaller than the ML estimator based on the exact discrete time model. Second, although the trapezoidal method leads to a smaller discretization bias than the Euler method, the estimation bias is bigger. As a result, it is not clear if there is a gain in reducing the total bias by using a higher order approximation. When comparing the estimator based on the Euler method and the exact ML, we find that the asymptotic variance of the former estimator is smaller. As a result, there is clear evidence for preferring the estimator based on the Euler method to the exact ML in the univariate linear diffusion.

Simulations suggest the bias continues to be large in finite samples. It is also confirmed that for empirically relevant cases, the magnitude of the discretization bias in the two approximate methods is much smaller than that of the estimation bias. The two approximate methods lead to a smaller variance than exact ML. Most importantly for practical work, there is strong evidence that the bias formulae work well and so they can be recommended for analytical bias correction with these models.

For the univariate square root model, the Euler method is found to have smaller bias and smaller variance than the Nowman method. Discretizing the diffusion function both in the Euler method and the Nowman method causes no discretization bias on the mean reversion paramter. For the Euler method, we have derived an explicit expression for the discretization bias caused by discretizing the drift function. The simulation results suggest that the Euler method performs best in terms of both bias and variance.

The analytic and expansion results given in the paper are obtained for stationary systems. Bias analysis for nonstationary and explosive cases require different methods. For diffusion models with constant diffusion functions, it may be possible to extend recent finite sample and asymptotic expansion results for the discrete time AR(1) model (Phillips, 2010) to a continuous time setting. Such an analysis would involve a substantial extension of the present work and deserves treatment in a separate study.

Chapter 3 Limit Theory for Multivariate Linear Diffusion Estimation

3.1 Introduction

Multivariate continuous time models had received a considerable level of interest in macro-econometrics over the period from 1960s to 1980s, featured in several useful theoretical contributions such as Bergstrom (1966, 1984), Phillips (1972) and various applications such as Bergstrom and Wymer (1976) and Knight and Wymer (1978). After experiencing a quiet period among econometricians, they are once again at the forefront in academic econometric circles. The main fuel for the resurrection is the usefulness of these models in the development of the modern asset pricing theory. Given complicated interplay among economic and financial variables, not surprisingly, multivariate continuous time models, which allow for interactions among variables, have received more attention in the recent literature on asset pricing in the hope of capturing more realistic dynamic interactions. Prominent examples include stochastic volatility models for equity and exchange rate series (Duffie, Pan and Singleton, 2000) and term structure models for multiple yields (Duffie and Kan, 1996).

Continuous time models used in macroeconomics often take a linear form. Under Gaussianity, this assumption implies a diffusion model with a linear drift function and a constant diffusion function. The efficient estimation of system parameters, based on discrete observations, is achieved by the mean of maximum likelihood (ML) or least squares; see, for example, Phillips (1972). In finance, more successful models allow the diffusion function to be time varying but maintain linearity for the drift function. To match these developments in the use of more complicated multivariate continuous time models in the theoretical finance literature, various econometric techniques have been developed for estimating system parameters from discrete data. Examples include the efficient method of moments (EMM) (Gallant and Tauchen, 1996), Bayesian MCMC methods (Eraker, 2001), the empirical characteristic function method (Singleton, 2001; Knight and Yu, 2002), and in-fill ML (Pedersen, 1995; Durham and Gallant, 2002), as well as approximate ML methods based on closed-form expansions (Aït-Sahalia, 2008; Aït-Sahalia and Kimmel, 2007; Aït-Sahalia and Yu, 2006).

For multivariate continuous time models with a linear drift function, an exact discrete time vector autoregressive (VAR) model can be obtained. When the diffusion function is constant, the VAR model is Gaussian and hence can be estimated by ordinary least squares (OLS) or ML. When the diffusion function has the level effect, the VAR model becomes non-Gaussian but can be estimated by generalized least squares. The asymptotic theory for VAR estimation is standard; see, for example, Mann and Wald (1943) for the stationary case and Phillips and Durlauf (1986) for the unit root case. It is known that the mean reversion matrix in the continuous time model is the logarithmic transformation of the autoregressive (AR) coefficient matrix. Under the identification condition, this relation is bijective. It is this bijective and measurable relationship that will be used find the asymptotic theory of the estimated mean reversion matrix.

It appears that the delta method, when applied to the principal value of the logarithm of the VAR coefficient matrix, can be used to find the limit distribution of the estimated mean reversion matrix. Unfortunately, this straightforward application of the delta method leads to a covariance matrix that is practically difficult to use. The standard limit distribution is available for the estimated VAR coefficient matrix. But to utilize this distribution, the standard matrix calculus formula implies that the mean reversion matrix is expressed as an infinite polynomial of the VAR coefficient matrix. As a result, the covariance matrix involves an infinite polynomial which must be truncated in practice and hence the calculation of the asymptotic covariance is difficult to implement. This situation is in the sharp contrast to the univariate setup where the delta method is easily applied.

This paper contributes to the literature in three ways. First, under regular conditions, we derive the asymptotic distribution of the estimated mean reversion matrix whose covariance matrix is very easy to calculate. We do this by using a new result obtained in the linear algebra literature, which enables us to relate the mean reversion matrix to the VAR coefficient matrix as a polynomial function of finite order. Second, we derive the asymptotic theory for the estimated mean reversion matrix not only for the stationary case but also for the non-stationary case. Third, we provide the joint limit distribution of the estimated mean reversion matrix and its eigenvalues. The theory is established in the context of the multivariate diffusion model of an arbitrary dimension but with a linear drift and a constant diffusion. We focus on this model simply because the asymptotic theory is well developed for the exact discrete time model. However, our theory continues to work for models with a more complicated diffusion function. As long as the asymptotic theory for the exact discrete time model is known, our method is applicable.

Phillips (1972) used least squares to estimate a structural continuous time model where the mean reversion matrix depends on a set of structural parameters, established the asymptotic normality, and derived the analytical expression for the asymptotic variance based on the assumption that the derivative of the mean reversion matrix with respect to the VAR coefficient matrix is known. The setup of Phillips (1972) is simpler than the model we consider here in the sense that we estimate the full mean reversion matrix and hence the dimension of our parameter space is higher. Also, we do not assume that the derivative of the mean reversion matrix with respect to the VAR coefficient matrix is known. In the context of univariate diffusion, Ait-Sahalia (2002) developed the asymptotic theory for his approximate ML method under the long span asymptotics whereas Jeong and Park (2009) established the asymptotic theory for a wide range of estimators in the cases of stationarity and unit root with a large time span and a small sampling interval. The results obtained in our paper may be regarded as a multivariate generalization to those in the univariate diffusion although our model specification only allow a linear drift function.

The rest of the paper is organized as follows. Section 2 describes the model and introduces the estimator of the mean reversion matrix based on LS estimation of the VAR model. The consistency of the estimator is established, and some arguments on the reasons why the new estimator is preferred are addressed. Section 3 derives asymptotic properties for the stationary diffusion case. Some important special cases are discussed. The limit theory is obtained in Section 4 for the diffusion model with unit roots. In Section 5, the theory is illustrated using the daily realized volatility data on Pound, Euro and Yen exchange rates. Section 6 concludes. Proofs of the propositions and theorems are collected in the Appendix.

3.2 The Model and New Estimation Approach

We consider an m-dimensional multivariate diffusion process of the form:

$$dX(t) = (AX(t) + b)\,dt + \Sigma^{1/2} dW(t), \tag{3.2.1}$$

where $X(t) = (X_1(t), \cdots, X_m(t))'$ is an m-dimensional continuous time process, A and b are $m \times m$ and $m \times 1$ matrices, whose elements need to be estimated, $\Sigma^{1/2}$ is a matrix of the diffusion coefficients, and $W(t)$ is a m-dimensional standard Brownian motion. We assume the matrix $\Sigma = \left[\Sigma^{1/2}\right]\left[\Sigma^{1/2}\right]'$ is positive definite. This model has been used to model multiple yields in the term structure literature and the univariate version was first proposed in Vasicek (1977).

Although the process follows a continuous time stochastic differential equation system, observations are available only at discrete time points, say at T equally spaced points $\{th\}_{t=0}^{T}$, where h is the sampling interval and is taken to be fixed. In practice, h might be very small, corresponding to high-frequency data. We can also write sample size T as $T = N/h$ by letting N denote the time span of the data. In this paper, we use $X(t)$ to represent a continuous time process and X_t to represent a discrete time process. When there is no confusion, we simply write X_{th} as X_t.

The exact discrete time representation of (3.2.1) is

$$X_t = e^{Ah}X_{t-1} + \int_0^h e^{As}bds + \varepsilon_t \tag{3.2.2}$$

where the matrix exponential e^{Ah} is defined as $e^{Ah} = \sum_{j=0}^{\infty} \frac{1}{j!}(Ah)^j$, $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{mt})'$ is a Gaussian martingale difference sequence (MDS) with respect to the natural filtration with

$$E\left(\varepsilon_t \varepsilon_t^{'}\right) = \int_0^h e^{As}\Sigma e^{A's}ds := \Omega.$$

Letting $F = e^{Ah}$, $g = \int_0^h e^{As}bds$, we have the system

$$X_t = FX_{t-1} + g + \varepsilon_t, \tag{3.2.3}$$

which is a VAR model of order 1 with $\text{MDS}(0,\Omega)$ innovations.

One common method to estimate the VAR system (3.2.3) is OLS approach, which gives us an estimator equivalent to the ML estimator under constant initial conditions. Setting $Z_t = \left[X_t^{'}, 1\right]^{'}$, the least square (LS) estimator of $[F, g]$ is

$$\left[\hat{F}, \hat{g}\right] = \left[\sum_{t=1}^{n} X_t Z_{t-1}^{'}\right] \times \left[\sum_{t=1}^{n} Z_{t-1} Z_{t-1}^{'}\right]^{-1}. \tag{3.2.4}$$

If we have prior knowledge that $b = 0$ and hence $g = 0$, the LS estimator of F is:

$$\hat{F} = \left[\sum_{t=1}^{n} X_t X_{t-1}^{'}\right] \times \left[\sum_{t=1}^{n} X_{t-1} X_{t-1}^{'}\right]^{-1}. \tag{3.2.5}$$

The key issue is how to get one desired estimation of A in terms of consistency and efficiency by using the estimation of F.

In general, identification of A from the implied discrete model (3.2.2) generating discrete observations $\{X_{th}\}$ is not automatically satisfied. The necessary and sufficient condition for identifiability of A in model (3.2.1) is that the correspondence between A and $[F, g]$ be one-to-one, since (3.2.2) is effectively a reduced form for the discrete observations. Phillips (1973) studied the identifiability of (A, Σ) in (3.2.2) in terms of the identifiability of the matrix A in the matrix exponential $F = \exp(Ah)$ under possible restrictions implied by the structural functional dependence A in (3.2.1). In general, a one-to–one correspondence between A and F, requires the structural matrix A to be restricted. This is because if A satisfies $\exp\{Ah\} = F$ and some of its eigenvalues are complex, A is not uniquely identified. In fact, adding to each pair of conjugate complex eigenvalues the imaginary numbers $2ik\pi/h$ and $-2ik\pi/h$ for any integer k, leads to another matrix satisfying $\exp\{Ah\} = F$. This phenomenon is well known as aliasing in the signal processing literature. When restrictions are placed on the structural matrix A identification is possible. Phillips (1973) gave a rank condition for the case of linear homogeneous relations between the elements of a row of A. A special case is when A is triangular. Hansen and Sargent (1983) extended this result by showing that the reduced form covariance structure $G > 0$ provides extra identifying information about A, reducing the number of potential aliases.

To address this aliasing problem, we impose a principal value condition which excludes such aliases by restricting the continuous time eigenvalues to the open strip $\{\eta \in \mathbb{C}, -\pi/h < Im(\eta) < \pi/h\}$. Empirically realistic values for h are almost always small in finance (e.g. $h = 1/12$ for monthly data and $1/252$ for daily data), so that the support $(-\pi/h, \pi/h)$ implied by this condition is typically quite wide and covers empirically relevant cases.

Assumption 1: The eigenvalues in A lie in the open strip $\{\eta \in \mathbb{C}, -\pi/h < Im(\eta) < \pi/h\}$ of the complex plane.

Proposition 3.2.1 *Under Assumption 1, F has no eigenvalues on the closed negative real axis, namely,*

$$spec\{F\} \cap \mathbb{R}_0^- = \varnothing$$

where $spec\{F\}$, the spectrum of F, is the set of all the distinct eigenvalues of F.

When F has no eigenvalues on the closed negative real axis, F has a unique logarithm with eigenvalues in the open strip $\{z \in \mathbb{C}, -\pi < Im(z) < \pi\}$ of the complex plane. This is a well known result in literature of linear algebra. The unique logarithm is called the principal logarithm and is denoted by $\ln(F)$. Under Assumption 1, Ah is the principal logarithm of F, namely,

$$A = \frac{1}{h}\ln(F)$$

To get an explicit relationship between F and A that only involves a summation of finite order, we propose to use the following new result recently obtained in the linear algebra literature. It gives an explicit formula for the principal logarithm of a matrix F as a polynomial in the matrix $(I-F)$ of the finite order with simple integral formulae for the coefficients involving the coefficients of the characteristic polynomial of $I-F$.

Lemma 3.2.1 *(Cardoso (2005)) Assume $F \in \mathbb{R}^{m\times m}$. If $D = \{\tau \in \mathbb{R} |\ spec\{I-(I-F)\tau\} \cap \mathbb{R}_0^- = \varnothing\}$, then for all $\tau \in D$,*

$$\ln[I-(I-F)\tau] = f_1(\tau)I + f_2(\tau)(I-F) + \cdots + f_m(\tau)(I-F)^{m-1} \tag{3.2.6}$$

where $f_1, \cdots, f_m$ are differentiable functions in D, given by

$$f_1(\tau) = \int_0^\tau \frac{C_m S^{m-1}}{1+C_1 S+\cdots+C_m S^M} dS,$$
$$f_j(\tau) = \int_0^\tau \frac{-S^{j-2}-C_1 S^{j-1}-\cdots-C_{m-j}S^{m-j}}{1+C_1 S+\cdots+C_m S^m} dS, \text{ for } j = 2, \cdots, m-1,$$
$$f_m(\tau) = \int_0^\tau \frac{-S^{m-2}}{1+C_1 S+\cdots+C_m S^m} dS,$$

and C_j, $j = 1, \cdots, m$, are the real coefficients of the characteristic polynomial of $I-F$.

Remark 3.2.1 *Since the characteristic polynomial of $I-F$ takes the form of*

$$P(z) = \det[zI-(I-F)] = z^m + C_1 z^{m-1} + \cdots + C_{m-1}z + C_m, \qquad (3.2.7)$$

the coefficients $\{C_j\}_{j=1}^m$ have the following expressions in terms of the elementary symmetric functions

$$C_1 = (-1)\sum_{s=1}^{m}(1-\lambda_s) = -tr(I-F),$$
$$C_2 = (-1)^2 \sum_{1\le s<k\le m}(1-\lambda_s)(1-\lambda_k) = \frac{1}{2}\left\{[tr(I-F)]^2 - tr\left[(I-F)^2\right]\right\},$$
$$\cdots$$
$$C_m = (-1)^m \prod_{s=1}^{m}(1-\lambda_s) = (-1)^m \det(I-F),$$

where $\{\lambda_s\}_{s=1}^m$ are the eigenvalues of F.

By Proposition 3.2.1, $1 \in D$. Let $\tau = 1$ in (3.2.6), giving

$$Ah = \ln(F) = f_1 I + f_2 (I-F) + \cdots + f_m (I-F)^{m-1}, \qquad (3.2.8)$$

where

$$f_1 = \int_0^1 \frac{C_m S^{m-1}}{1 + C_1 S + \cdots + C_m S^m} dS, \qquad (3.2.9)$$
$$f_j = \int_0^1 \frac{-S^{j-2} - C_1 S^{j-1} - \cdots - C_{m-j} S^{m-j}}{1 + C_1 S + \cdots + C_m S^m} dS, \text{ for } j = 2, \cdots, m-1, \qquad (3.2.10)$$
$$f_m = \int_0^1 \frac{-S^{m-2}}{1 + C_1 S + \cdots + C_m S^m} dS. \qquad (3.2.11)$$

Therefore, a nature estimator of matrix A is

$$\hat{A} = \frac{1}{h}\ln\left(\hat{F}\right) = \frac{1}{h}\left\{\hat{f}_1 I + \hat{f}_2\left(I-\hat{F}\right) + \cdots + \hat{f}_m\left(I-\hat{F}\right)^{m-1}\right\}, \qquad (3.2.12)$$

where

$$
\begin{aligned}
\hat{f}_1 &= \int_0^1 \frac{\hat{C}_m S^{m-1}}{1+\hat{C}_1 S+\cdots+\hat{C}_m S^m} dS, \\
\hat{f}_j &= \int_0^1 \frac{-S^{j-2}-\hat{C}_1 S^{j-1}-\cdots-\hat{C}_{m-j} S^{m-j}}{1+\hat{C}_1 S+\cdots+\hat{C}_M S^M} dS, \text{ for } j=2,\cdots,m-1, \\
\hat{f}_m &= \int_0^1 \frac{-S^{m-2}}{1+\hat{C}_1 S+\cdots+\hat{C}_m S^m} dS,
\end{aligned}
$$

$\left\{\hat{\lambda}_s\right\}_{s=1}^m$ are the eigenvalues of $\hat{F}$, and for $j=1,\cdots,m$,

$$
\hat{C}_j = (-1)^j \sum_{1\leq s_1<s_2\ldots<s_j\leq m} \left(1-\hat{\lambda}_{s_1}\right)\cdots\left(1-\hat{\lambda}_{s_j}\right). \tag{3.2.13}
$$

When all the eigenvalues of $(I-F)$ have modulus less than unity, there is another widely used expression for the principal logarithm of F, which possesses the form of

$$
Ah = \ln F = -\sum_{j=1}^{\infty} \frac{1}{j}(I-F)^j. \tag{3.2.14}
$$

Given $\hat{F}$, the above representation leads to the estimation of A as

$$
\tilde{A} = \frac{1}{h}\ln\hat{F} = -\frac{1}{h}\sum_{j=1}^{\infty}\frac{1}{j}\left(I-\hat{F}\right)^j. \tag{3.2.15}
$$

Notice the fact that for $j=1,\cdots,\infty$,

$$
\left(I-\hat{F}\right)^j-(I-F)^j = -\sum_{s=0}^{j-1}(I-F)^s\left(\hat{F}-F\right)\left(I-\hat{F}\right)^{j-1-s}.
$$

Then, straightforward calculations allow us to show that

$$
Vec\left(\tilde{A}-A\right) = \frac{1}{h}\left\{\sum_{j=1}^{\infty}\frac{1}{j}\left[\sum_{k=0}^{j-1}(I-F)^k\otimes\left[\left(I-\hat{F}\right)^{j-1-k}\right]'\right]\right\}Vec\left(\hat{F}-F\right),
$$

where $Vec(\cdot)$ denotes raw stacking of a matrix and $\otimes$ means Kronecker product. As $\tilde{A}$ is a measurable transformation of $\hat{F}$, it seems to suggest that one can apply the standard results, such as the delta method, to obtain the asymptotic theory for $\tilde{A}$

once the asymptotic theory for $\hat{F}$ is known. And the matrix

$$\Lambda = \sum_{j=1}^{\infty} \frac{1}{j} \left[\sum_{k=0}^{j-1} (I-F)^k \otimes \left[(I-F)^{j-1-k} \right]' \right]$$

is supposed to use in the sandwich form to get the asymptotic covariance matrix of $\tilde{A}$.

While, the later mentioned estimation approach is not as applicable as $\hat{A}$ proposed in (3.2.12). First of all, an infinite sum is involved in $\tilde{A}$ as well as in Λ. Hence, the calculation of $\tilde{A}$ and its asymptotic covariance matrix requires one to truncate the infinite sum in practice. Clearly, the truncation rules for $\tilde{A}$ and Λ should be quite different. Unfortunately, there is no clear guideline as to how to truncate the infinite sum either in $\tilde{A}$ or Λ. If too few terms are used, the truncation error would be quite large, especially for the estimation of the asymptotic covariance matrix. If too many terms are used, which is necessary when the discrete model in (3.2.3) is very stationary, the computational cost might be considerably high. Moreover, once the truncation is done, $\tilde{A}$ will not be a consistent estimation any more. So is the estimation of asymptotic covariance matrix. This highlights the advantage of the proposed estimation $\hat{A}$ in (3.2.12) which only involves a summation of finite order. And, $\hat{A}$ is always a consistent estimation as we will discussed later in this section.

Secondly, the preferred representation of A in (3.2.8) (also $\hat{A}$ in (3.2.12)) has a wider circle of convergence comparing to that of representation (3.2.14) ($\tilde{A}$ in (3.2.15)), namely,

$$\left\{F : spec\{F\} \cap \mathbb{R}_0^- = \varnothing\right\} \supset \{F : \text{eigenvalues of } (I-F) \text{ have modulus less than unity}\}.$$

This is a significant advantage of representations of A in (3.2.8) and $\hat{A}$ in (3.2.12), which makes them much more applicable in practice. For example, let one eigenvalue of A is the imaginary number $i\pi/(2h)$, which satisfies the Assumption 1. Then, the corresponding eigenvalue of F is $\exp\{i\pi/2\} = i$, the imaginary unit. Consequently, $(I-F)$ has one eigenvalue whose modulus is equal to 2, which makes the representation of A in (3.2.14) undefined. However, in this case, we may still have $spec\{F\} \cap \mathbb{R}_0^- = \varnothing$. Hence, the representation of A in (3.2.8) works very well.

Moreover, even if all the eigenvalues of $(I - F)$ have modulus less than 1, some eigenvalues of $(I - \hat{F})$ may lie outside the domain corresponding to the circle of convergence - as it inevitably will for some fitted values, even though the probability of being in that domain tends to zero. For example, assume that one F's eigenvalue is 0.1. Hence, the modulus of the corresponding eigenvalue of $(I - F)$ is 0.9. Therefore, representation of A in (3.2.14) is well-defined. However, the estimated eigenvalue of $\hat{F}$ could be $0.1i$, which makes the corresponding eigenvalue of $(I - \hat{F})$ has modulus $\sqrt{1 + 0.01} > 1$. The estimation $\tilde{A}$ in (3.2.15) diverges to infinity. Not surprisingly, we still have $spec\{\hat{F}\} \cap \mathbb{R}_0^- = \varnothing$. As a result, $\hat{A}$ in (3.2.12) provide us a valid estimation.

When $spec\{\hat{F}\} \cap \mathbb{R}_0^- = \varnothing$, both $\hat{A}$ in (3.2.12) and $\tilde{A}$ in (3.2.15) fail to work. The reason for the failure of $\hat{A}$ is that some integrations in $\hat{f}_j$, for $j = 1, \cdots, m$, are infinity.

Based on the above argument, we may conclude that $\hat{A}$ works very well as long as $\tilde{A}$ works, but not vice versa.

Before leaving this section, we establish the consistency of the proposed estimation $\hat{A}$ in (3.2.12). Notice the fact that eigenvalues under ordering are continuous functions of the elements of the matrix (the ordering rule is discussed in Section 3). Hence, the eigenvalues of $\hat{F}$, $\{\lambda_s(\hat{F})\}_{s=1}^m$, converge to the eigenvalues of F, $\{\lambda_s(F)\}_{s=1}^m$, in probability, as long as $\hat{F} \xrightarrow{p} F$. Since $\hat{C}_j$, $j = 1, \cdots, m$, in representation (3.2.13) is continuous on $\{\lambda_s(\hat{F})\}_{s=1}^m$, and $\hat{f}_j$, for $j = 1, \cdots, m$, are continuous functions of $\{\hat{C}_j\}_{j=1}^m$, it is straightforward to get the consistency of $\hat{A}$. We collect these results in the following theorem.

Theorem 3.2.1 *Let $\hat{A}$ be defined in (2.3.10), Assumption 1 is hold, h is fixed and $T \to \infty$. If $\hat{F} \xrightarrow{p} F$, then*

$$\hat{A} \xrightarrow{p} A$$

3.3 Asymptotic Theory for Stationary Model

This Section develops a limit theory for $\hat{A}$ in (3.2.12) under a stationary condition.

Assumption 2: The eigenvalues in A have negative real parts.

This is one commonly used condition that ensures the discrete time model (3.2.2) corresponding to the continuous time model (3.2.1) to be covariance stationary, as all the eigenvalues of $F = \exp(Ah)$ have modulus less than 1. In this case, A is known as the mean reversion matrix. Notice that A is non-singular. The discrete model (3.2.2) can be rewritten as (see Phillips (1972))

$$\begin{aligned} X_t &= e^{Ah}X_{t-1} + A^{-1}\left[e^{Ah} - I\right]b + \varepsilon_t \\ &= FX_{t-1} + g + \varepsilon_t \end{aligned} \tag{3.3.1}$$

where ε_t are $\mathrm{MDS}(0, \Omega)$.

The ML/LS estimators of the coefficients in the discrete model (3.3.1), $\left[\hat{F}, \hat{g}\right]$ in (3.2.4), or $\hat{F}$ in (3.2.5) which is obtained under the prior knowledge of $b = 0$, are also supposed to use. Under constant initial condition, both these two kinds of estimators have the following standard limit theory (see Hannan (1970, p.329)).

Lemma 3.3.1 *When Assumption 2 is true, h is fixed and sample size T goes to infinity, we have*

(a) $\quad \hat{F} \xrightarrow{a.s} F,$

(b) $\quad \sqrt{n}\left\{Vec\left(\hat{F}\right) - Vec\left(F\right)\right\} \xrightarrow{d} N\left(0, V_F\right)$

where $Vec(F)$ denotes row stacking of F, $V_F = \Omega \otimes (V_X)^{-1}$, $V_X = Var(X_t) = \sum_{i=0}^{\infty} F^i \Omega F'^i$ and $\Omega = E\left(\varepsilon_t \varepsilon_t'\right)$.

Before reporting the limit theory of $\hat{A}$ in (3.2.12), we introduce some notations. For any matrix Ψ, $(\Psi)_{kj}$ denotes the matrix formed by deleting row k and column j from Ψ. Let $adj(\Psi)$ denote the adjoint of Ψ which is a matrix whose row k, column j element is given by $(-1)^{k+j}\left|(\Psi)_{jk}\right|$, where $\left|(\Psi)_{kj}\right|$ is the matrix's determinant.

Theorem 3.3.1 *Let Assumption 1 and 2 hold. When h is fixed and sample size T goes to infinity, we have*

$$h\sqrt{n}Vec\left(\hat{A}-A\right)\xrightarrow{d}N\left(0,\Gamma V_F\Gamma'\right),\tag{3.3.2}$$

where

$$\Gamma=\left\{\sum_{j=1}^{m}Vec\left[(I-F)^{j-1}\right]F'_jL^{-1}H-\sum_{j=2}^{m}\sum_{s=0}^{j-2}f_j\left\{(I-F)^s\otimes\left[(I-F)^{j-2-s}\right]'\right\}\right\},$$

$F'_j=\begin{bmatrix}\frac{\partial f_j}{\partial C_m} & \frac{\partial f_j}{\partial C_{m-1}} & \cdots & \frac{\partial f_j}{\partial C_1}\end{bmatrix}$*, for $j=1,\cdots,m$, with f_j taking the forms as formulae (3.2.9), (3.2.10) and (3.2.11),* $L=\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & 2 & \cdots & 2^{m-1}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & m & \cdots & m^{m-1}\end{bmatrix}$ *is a nonsingular matrix,*

$H=\begin{bmatrix}[Vec(H_1)]'\\ \vdots\\ [Vec(H_m)]'\end{bmatrix}$ *with $H_z=[adj(zI-(I-F))]'$ for $z=1,\cdots,m$, and V_F is given in Lemma 3.3.1.*

Remark 3.3.1 *In the appendix, we give a proof of*

$$\Gamma=I_{m^2}+O(h),\tag{3.3.3}$$

where I_{m^2} is a $m^2\times m^2$ identity matrix. Therefore, when h is comparatively small, Γ is nonsingular and close to I_{m^2}. Consequently, asymptotic covariance matrix $\Gamma V_F\Gamma'$ is positive definite. And we can safely claim that the formula (3.3.2) provides a non-degenerated asymptotic distribution for every elements of $h\sqrt{n}Vec\left(\hat{A}-A\right)$.

Remark 3.3.2 *The analytic expression for the asymptotic covariance in Theorem (3.3.1) involves only summations of finite order, making the implementation straightforward. From the consistency of $\hat{C}_j$, $j=1,\cdots,m$, and $\hat{F}$, we could get a consistent* *estimation of* Γ *as following*

$$\hat{\Gamma} = \left\{ \sum_{i=1}^{m} Vec\left[\left(I-\hat{F}\right)^{i-1}\right] \hat{F}_i' L^{-1}\hat{H} - \sum_{j=2}^{m}\sum_{s=0}^{j-2} \hat{f}_i \left\{ \left(I-\hat{F}\right)^{s} \otimes \left[\left(I-\hat{F}\right)^{j-2-s}\right]' \right\} \right\} \tag{3.3.4}$$

where $\hat{F}_j'$ *and* $\hat{f}_j$ *are obtained from* F_j' *and* f_j *by replacing* $\left\{C_j\right\}_{j=1}^{m}$ *with* $\left\{\hat{C}_j\right\}_{j=1}^{m}$.

Certain low dimensional models, such as $m = 1,2,3$, always attract considerable interests in practical applications. In order to facilitate the use of Theorem 3.3.1 in low dimensional cases, the following two Corollary provide a more explicit expression of the asymptotic covariance matrix, $\Gamma V_F \Gamma'$, in the cases where $m = 2,3$, respectively. The results can be derived directly by straightforward calculations based on the rules given in Theorem 3.3.1. So, the proofs are omitted. For the case $m = 1$, the long time span asymptotics have already been well studied (see, Tang and Chen (2009)).

Corollary 3.3.2 *When* $M = 2$*, the matrix* Γ *in the asymptotic covariance matrix defined in Theorem 3.3.1 takes the following form*

$$\Gamma = Vec\left[\varphi_1 I + \varphi_3 (I-F)\right]\Delta_1 + Vec\left[\varphi_2 I + \varphi_4 (I-F)\right]\Delta_2 - f_2 I_4,$$

where

$$\varphi_1 = \int_0^1 \frac{-C_2 S^2}{\left(1+C_1 S+C_2 S^2\right)^2} dS, \quad \varphi_2 = \int_0^1 \frac{S+C_1 S^2}{\left(1+C_1 S+C_2 S^2\right)^2} dS,$$

$$\varphi_3 = \int_0^1 \frac{S}{\left(1+C_1 S+C_2 S^2\right)^2} dS, \quad \varphi_4 = \int_0^1 \frac{S^2}{\left(1+C_1 S+C_2 S^2\right)^2} dS,$$

$$f_2 = \int_0^1 \frac{-1}{1+C_1 S+C_2 S^2} dS, \quad C_1 = -tr(I-F), \quad C_2 = \det(I-F),$$

$$\Delta_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}, \quad \Delta_2 = -\begin{pmatrix} 1-F_{(2,2)} & F_{(2,1)} & F_{(1,2)} & 1-F_{(1,1)} \end{pmatrix},$$

and I *and* I_4 *denote* 2×2, 4×4 *identity matrix, respectively.* $F_{(k,j)}$ *denotes the elements of* F *in row* k*, column* j*.*

Corollary 3.3.3 *When $m = 3$, the matrix Γ in the asymptotic covariance matrix defined in Theorem 3.3.1 takes the form of*

$$\Gamma = Vec\left[\xi_1 I + \xi_4 (I-F) + \xi_7 (I-F)^2\right]\Delta_3 + Vec\left[\xi_2 I + \xi_5 (I-F) + \xi_8 (I-F)^2\right]\Delta_4$$
$$+ Vec\left[\xi_3 I + \xi_6 (I-F) + \xi_9 (I-F)^2\right]\Delta_8 - f_2 I_9 - f_3\left[\sum_{s=0}^{1}\left\{(I-F)^s \otimes \left[(I-F)^{1-s}\right]'\right\}\right],$$

where

$$\xi_1 = \int_0^1 \frac{-C_3 S^3}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS, \quad \xi_2 = \int_0^1 \frac{-C_3 S^4}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS,$$

$$\xi_3 = \int_0^1 \frac{(1+C_1 S + C_2 S^2) S^2}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS, \quad \xi_4 = \int_0^1 \frac{-(C_2 + C_3 S) S^3}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS,$$

$$\xi_5 = \int_0^1 \frac{(1+C_1 S) S^2}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS, \quad \xi_6 = \int_0^1 \frac{(1+C_1 S) S^3}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS,$$

$$\xi_7 = \int_0^1 \frac{S^2}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS, \quad \xi_8 = \int_0^1 \frac{S^3}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS,$$

$$\xi_9 = \int_0^1 \frac{S^4}{(1+C_1 S + C_2 S^2 + C_3 S^3)^2} dS,$$

$$C_1 = -tr(I-F), \quad C_2 = \frac{1}{2}\left\{[tr(I-F)]^2 - tr\left[(I-F)^2\right]\right\}, \quad C_3 = -\det(I-F),$$

$$\Delta_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Delta_4 = \{-tr(I-F)\Delta_3 + \Delta_3[(I-F)\otimes I]\},$$

$$\Delta_5 = \Delta_3\left[(I-F)^2 \otimes I\right], \quad \Delta_6 = \frac{1}{2}\left\{-tr(I-F)^2\Delta_3 - 2tr(I-F)\Delta_3[(I-F)\otimes I]\right\},$$

$$\Delta_7 = \frac{1}{2}[tr(I-F)]^2\Delta_3, \quad \Delta_8 = \Delta_5 + \Delta_6 + \Delta_7,$$

$$f_2 = \int_0^1 \frac{-1 - C_1 S}{1+C_1 S + C_2 S^2 + C_3 S^3} dS, \quad f_3 = \int_0^1 \frac{-S}{1+C_1 S + C_2 S^2 + C_3 S^3} dS,$$

and I and I_9 denote the 3×3, 9×9 identity matrix, respectively.

If we are willing to assume that the mean reversion matrix A has distinct eigenvalues, the asymptotic covariance matrix representation could be much simplified, and the limit distribution of the eigenvalues could be derived.

Assumption 3: The matrix A is diagonalizable with distinct eigenvalues.

Proposition 3.3.1 *Under Assumption 1 and 3, $F = e^{Ah}$ is diagonalizable with distinct eigenvalues.*

Before reporting the limit theory when A has distinct eigenvalues, we intend to introduce a specific ordering rule for eigenvalues and a specific normalization rule of eigenvectors, in order to make eigen-decomposition of matrix unique. Firstly, we let F's eigenvalues $\{\lambda_1, \cdots, \lambda_m\}$ be ordered according to

$$Re(\lambda_1) \geq \cdots \geq Re(\lambda_m).$$

Then, any complex eigenvalues with $Re\left(\lambda_j\right) = Re\left(\lambda_{j+1}\right)$ will be ordered based on the absolute value of their imaginary parts as

$$\left|Im(\lambda_j)\right| \geq \left|Im(\lambda_{j+1})\right|.$$

Finally, for complex conjugate pairs $(\lambda_k, \lambda_{k+1})$, we order them based on the sign of the imaginary part, i.e., $Im(\lambda_m) > 0$ followed by $Im(\lambda_{m+1}) < 0$**.** This rule leads to a unique ordering of the eigenvalues. Let p_j, for $j = 1, ..., m$, are eigenvectors corresponding to eigenvalues λ_j, respectively. The normalization rule

$$p_j' p_j = 1$$

makes each corresponding eigenvector unique. As a result, F can be uniquely decomposed as

$$F = Pdiag\{\lambda_1, \cdots, \lambda_m\} Q,$$

where $P = \begin{bmatrix} p_1 & \cdots & p_m \end{bmatrix}$, $Q = P^{-1}$. The eigenvalues of A would be $\{\eta_1, \cdots, \eta_m\} = \frac{1}{h}\{\ln(\lambda_1), \cdots, \ln(\lambda_m)\}$.

Theorem 3.3.4 *Let Assumption 1, 2, and 3 hold. When h is fixed and sample size T goes to infinity, we have $h\sqrt{n}Vec\left(\hat{A} - A\right) \xrightarrow{d} N(0, \Gamma V_F \Gamma')$ as proved in Theorem*

3.3.1, but with a simplified representation of Γ *as*

$$\Gamma = \left(P \otimes Q'\right) \Lambda^{-1} \left(Q \otimes P'\right)$$

where $\Lambda = diag\{\Lambda_1, \cdots, \Lambda_M\}$*, and* Λ_k*, for* $k = 1, \cdots, m$ *is a* $m \times m$ *diagonal matrix whose* $(k,k)^{th}$ *element is equal to* $e^{\eta_k h}$*, and* $(\tau,\tau)^{th}$ *element with* $\tau \neq k$ *is equal to* $\left(e^{\eta_\tau h} - e^{\eta_k h}\right) / [(\eta_\tau - \eta_k)h]$.

Remark 3.3.3 *Notice that* $\left(e^{\eta_\tau h} - e^{\eta_k h}\right) / [(\eta_\tau - \eta_k)h] = 1 + O(h)$*, and* $e^{\eta_k h} = 1 + O(h)$*. Hence,*

$$\Gamma = \left(P \otimes Q'\right) \Lambda^{-1} \left(Q \otimes P'\right) = \left(P \otimes Q'\right) I_{m^2} \left(Q \otimes P'\right) + O(h) = I_{m^2} + O(h),$$

Therefore, Γ *is nonsingular when h is comparatively small.*

Remark 3.3.4 *Notice the fact that eigenvalues and eigenfunctions are continuous functions of elements of matrix. The following estimation, which is easy to get in practice, is consistent, as long as* $\hat{F}$ *is consistent.*

$$\hat{\Gamma} = \left(\hat{P} \otimes \hat{Q}'\right) \hat{\Lambda}^{-1} \left(\hat{Q} \otimes \hat{P}'\right),$$

where $\hat{\Lambda}$ *is obtained by replacing* η_k *with* $\hat{\eta}_k = \frac{1}{h} \ln\left(\hat{\lambda}_k\right)$ *for* $k = 1, \cdots, m$*,* $\left\{\hat{\lambda}_1, \cdots, \hat{\lambda}_m\right\}$ *are the ordered eigenvalues of* $\hat{F}$*,* $\hat{P} = \begin{bmatrix} \hat{p}_1 & \cdots & \hat{p}_m \end{bmatrix}$ *with* $\hat{p}_j$ *being the normalized eigenvector associated with the corresponding eigenvalues,* $\hat{Q} = \hat{P}^{-1}$.

Using the technics proposed by Phillips (1982), we may also derive the limit distribution of the eigenvalues and the joint limit distribution of the matrix and its eigenvalues.

Lemma 3.3.2 *Assume that the eigenvalues of F in (3.3.1) have modulus less than unity and F is diagonalizable with distinct eigenvalues. Let* $\lambda = (\lambda_1, \cdots, \lambda_m)'$ *be the ordered eigenvalues of F. When h is fixed and* $T \to \infty$*,*

(a)

$$\sqrt{T}\left(\hat{\lambda}-\lambda\right)\xrightarrow{d}N\left(0,GV_FG'\right),$$

(b)

$$\sqrt{T}\begin{pmatrix}\hat{\lambda}-\lambda\\ Vec\left(\hat{F}-F\right)\end{pmatrix}\xrightarrow{d}N\left(0,RV_FR'\right)$$

where $\hat{\lambda}=\left(\hat{\lambda}_1,\cdots,\hat{\lambda}_m\right)'$ *are the ordered eigenvalues of* $\hat{F}$*,* G *is an* $m\times m^2$ *matrix whose* j^{th} *row is* $\left(p^j\right)'\otimes p_j'$ *with* $\left(p^j\right)'$ *and* p_j *denoting the* j^{th} *row of* P^{-1} *and the* j^{th} *column of* P*, respectively, and* $R'=[G',I_{m^2}]$*.*

The same approach in the proof of Lemma 3.3.2 can be applied to obtain the limit distribution of the eigenvalues of $\hat{A}$, and the joint limit distribution of $\hat{A}$ and its eigenvalues. The results are reported in the following theorem.

Theorem 3.3.5 *Under Assumptions 1-3, let* $\eta=(\eta_1,\cdots,\eta_m)'=(\ln\lambda_1/h,\cdots,\ln\lambda_m/h)'$ *be the eigenvalues of* A*, and* $\{\lambda_1,\cdots,\lambda_m\}$ *be the ordered eigenvalues of* F*. When* h *is fixed and* $T\to\infty$*,*

(a)

$$h\sqrt{T}\left(\hat{\eta}-\eta\right)\xrightarrow{d}N\left(0,G\Pi G'\right);$$

(b)

$$h\sqrt{T}\begin{pmatrix}\hat{\eta}-\eta\\ Vec\left(\hat{A}-A\right)\end{pmatrix}\xrightarrow{d}N\left(0,R\Pi R'\right),$$

where $\hat{\eta}=(\hat{\eta}_1,\cdots,\hat{\eta}_m)'=\left(\ln\hat{\lambda}_1/h,\cdots,\ln\hat{\lambda}_m/h\right)'$*,* $\left(\hat{\lambda}_1,\cdots,\hat{\lambda}_m\right)$ *are the ordered eigenvalues of* $\hat{F}$*,* $\Pi=\Gamma V_F\Gamma'$ *defined in Theorem 3.3.4, and matrices* G *and* R *are as defined in Lemma 3.3.2.*

3.4 Asymptotic Theory for Non-Stationary Model

This section concentrates on developing limit theory for $\hat{A}$ under a non-stationary situation. We let $A=0_{m\times m}$, a zero $m\times m$ matrix. Therefore the continuous time model (3.2.1) changes to be

$$dX(t) = (b)\,dt + \Sigma^{1/2} dW(t). \tag{3.4.1}$$

The exact discrete time representation should be

$$\begin{aligned} X_t &= X_{t-1} + bh + \varepsilon_t \\ &= FX_{t-1} + g + \varepsilon_t, \end{aligned} \tag{3.4.2}$$

where $F = I$, $g = bh$, $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{mt})'$ is a Gaussian MDS $(0, \Omega = \Sigma h)$.

Setting $Z_t = \left[X_t', 1\right]'$, the LS estimator of $[F, g]$ is

$$\left[\hat{F}, \hat{g}\right] = \left[n^{-1} \sum_{t=1}^{n} X_t Z_{t-1}'\right] \times \left[n^{-1} \sum_{t=1}^{n} Z_{t-1} Z_{t-1}'\right]^{-1}. \tag{3.4.3}$$

From the functional central limit theory (FCLT), we may get

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \Rightarrow B_1(r)$$

where $r \in [0, 1]$, $B_1(r)$ is m-vector Brownian motion with covariance Σh, $\lfloor Tr \rfloor$ denotes the integer part of Tr, the symbol "$\Rightarrow$" signifies weak convergence of associated probability measures and the limit is taken as the sample size $T \uparrow \infty$. Here and elsewhere in the paper, to achieve notational economy we frequently eliminate function arguments and write, for example, B_1 in place of $B_1(r)$ and $\int_0^1 B_1$ in place of $\int_0^1 B_1(r)\,dr$.

We first discuss the limit theory of $\left[\hat{F}, \hat{g}\right]$ when $g = bh = 0$. We use the following functional introduced by Park and Phillips (1988):

$$f(B, M, \Theta) = \left(\int_0^1 dBM' + \Theta'\right) \left(\int_0^1 MM'\right)^{-1}$$

where B is vector Brownian motion, M is a process with continuous sample paths such that $\int_0^1 MM' > 0$ $a.s.$, and Θ is a (possibly random) matrix of conformable dimension. As in Theorem 3.2 of Park and Phillips (1988) we find that:

$$T\left(\hat{F}-F\right) \Rightarrow f\left(B_1, B_1^*, \triangle_{21}\right) \tag{3.4.4}$$

where $B_1^* = B_1 - \int_0^1 B_1$, $\triangle_{21} = 0_{m\times m}$ as ε_t are Gaussian MDS.

For the case in which $g = bh \neq 0$, we define $\mu_1 = g/\left(g'g\right)^{1/2} = b/\left(b'b\right)^{1/2}$ and let $U = [\mu_1, U_2]$ be an orthogonal matrix of dimension $m \times m$. We further define $\underline{B}_1 = U_2' B_1$ and $\underline{\triangle}_{21} = U_2' \triangle_{21} = 0_{m\times m}$. From Theorem 3.6 of Park and Phillips (1988) we have that

$$T\left(\hat{F}-F\right) \Rightarrow f\left(B_1, \underline{B}_1^{**}, \underline{\triangle}_{21}\right) U_2' \tag{3.4.5}$$

$$T^{3/2}\left(\hat{F}-F\right)\mu_1 \Rightarrow \left(g'g\right)^{-1/2} f\left(B_1, \underline{P}, \underline{\delta}\right) \tag{3.4.6}$$

where $\underline{B}_1^{**} = \underline{B}_1 - 4\left(\int_0^1 \underline{B}_1 - (3/2)\int_0^1 s\underline{B}_1\right) + 6r\left(\int_0^1 \underline{B}_1 - 2\int_0^1 s\underline{B}_1\right)$, $\underline{\delta} = 0_{1\times m}$, and $\underline{P} = r - 1/2 - \left(\int_0^1 s\underline{B}_1' - (1/2)\int_0^1 \underline{B}_1'\right)\left(\int_0^1 \underline{B}_1\underline{B}_1' - \int_0^1 \underline{B}_1 \int_0^1 \underline{B}_1'\right)^{-1}\left(\underline{B}_1 - \int_0^1 \underline{B}_1\right)$.

By using the reported limit theory in discrete time model, the following theorem shows the asymptotic distribution of ML/LS estimator $\hat{A}$ defined in 3.2.12.

Theorem 3.4.1 *Assume $X(t)$ follows the model (3.4.1). When h is fixed and $T \to \infty$, we have:*

(a) when $b = 0$,

$$Th\left(\hat{A}-A\right) \xrightarrow{d} f\left(B_1, B_1^*, \triangle_{21}\right),$$

(b) when $b \neq 0$,

$$Th\left(\hat{A}-A\right) \xrightarrow{d} f\left(B_1, \underline{B}_1^{**}, \underline{\triangle}_{21}\right) U_2',$$

$$T^{3/2}h\left(\hat{A}-A\right)\mu_1 \xrightarrow{d} \left(g'g\right)^{-1/2} f\left(B_1, \underline{P}, \underline{\delta}\right),$$

where $f\left(B_1, B_1^*, \triangle_{21}\right)$, $f\left(B_1, \underline{B}_1^{**}, \underline{\triangle}_{21}\right) U_2'$ and $\left(g'g\right)^{-1/2} f\left(B_1, \underline{P}, \underline{\delta}\right)$ are defined as (3.4.4), (3.4.5) and (3.4.6), respectively, $g = bh$ and $\mu_1 = g/\left(g'g\right)^{1/2} = b/\left(b'b\right)^{1/2}$.

Remark 3.4.1 *When $b \neq 0$, $Th\left(\hat{A}-A\right) \xrightarrow{d} f\left(B_1, \underline{B}_1^{**}, \underline{\triangle}_{21}\right) U_2'$ give the asymp-*

totic theory for all linear combinations of the matrix $Th\left(\hat{A}-A\right)$*. Note that only* $Th\left(\hat{A}-A\right)\mu_1$ *is degenerate in the limit and the asymptotic theory for these vectors is given by* $T^{3/2}h\left(\hat{A}-A\right)\mu_1 \xrightarrow{d} (g'g)^{-1/2} f\left(B_1, \underline{P}, \underline{\delta}\right)$.

Remark 3.4.2 *For the case in which* $m = 1$*, the results in Theorem 3.4.1 turn out to be:*

$$T\left(\hat{A}-A\right) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r) - W(1)\int_0^1 W(r)dr}{\int_0^1 [W(r)]^2 dr - \left\{\int_0^1 W(r)dr\right\}^2}, \ \ when\ b = 0,$$

$$T^{3/2}h\left(\hat{A}-A\right) \xrightarrow{d} N\left(0, \frac{12}{b^2 h}\Sigma\right), \ \ when\ b \neq 0,$$

where $W(r)$ *is* 1*-dimensional standard Brownian motion.*

Theorem 3.4.2 *Assume* $X(t)$ *follows the model (3.4.1) and* $\hat{F}$ *is defined as in (3.4.3). Let* $\left\{\hat{\lambda}_j\right\}_{j=1}^m$ *are ordered eigenvalues of* $\hat{F}$*, and* $\left\{\hat{\eta}_j = \ln\left(\hat{\lambda}_j\right)/h\right\}_{j=1}^m$ *are corresponding eigenvalues of* $\hat{A}$*. When* h *is fixed and* $T \to \infty$*,*

(a) when $b = 0$*,*

$$T\left\{\sum_{j=1}^m \hat{\lambda}_j - m\right\} \xrightarrow{d} \Delta \cdot Vec\left[f\left(B_1, B_1^*, \triangle_{21}\right)\right],$$

$$Th\sum_{j=1}^m \hat{\eta}_j \xrightarrow{d} \Delta \cdot Vec\left[f\left(B_1, B_1^*, \triangle_{21}\right)\right],$$

(b) when $b \neq 0$*,*

$$T\left\{\sum_{j=1}^m \hat{\lambda}_j - m\right\} \xrightarrow{d} \Delta \cdot Vec\left[f\left(B_1, \underline{B}_1^{**}, \underline{\triangle}_{21}\right)U_2'\right],$$

$$Th\sum_{j=1}^m \hat{\eta}_j \xrightarrow{d} \Delta \cdot Vec\left[f\left(B_1, \underline{B}_1^{**}, \underline{\triangle}_{21}\right)U_2'\right],$$

where Δ *is a row vector of dimension* m^2 *whose* $1^{st}, [m+2]^{th}, \cdots, [(m-1)m+m]^{th}$ *elements are* 1 *and* 0 *otherwise,* $f\left(B_1, B_1^*, \triangle_{21}\right)$, $f\left(B_1, \underline{B}_1^{**}, \underline{\triangle}_{21}\right)U_2'$ *are defined as in Theorem 3.4.1.*

Remark 3.4.3 *When $b = 0$ is a apriori knowledge, the discrete model 3.4.2 changes to be a AR(1) model without drift. Hence, the LS estimator of F would be*

$$\hat{F} = \left[\sum_{t=1}^{n} X_t X_{t-1}'\right] \times \left[\sum_{t=1}^{n} X_{t-1} X_{t-1}'\right]^{-1}.$$

From Park and Phillips (1988), we have

$$T\left(\hat{F} - F\right) \xrightarrow{d} f\left(B_1, B_1, \triangle_{21}\right).$$

The approach used in this section can be applied to this simple case easily. And, similar results should be obtained.

3.5 An Empirical Illustration

To illustrate the implementation of the new theory, a multivariate Ornstein–Uhlenbeck (OU) model is conducted to describe the joint movement over time of logarithmic daily realized volatility (RV) of Pound, Euro and Yen exchange rates, all against the US dollar. The logarithmic daily RV, $X = (X_1, X_2, X_3)'$, is sampled from January 4, 1999 to August 31, 2008. The sample interval (h) is $1/252$ and the sample size is 2444. The RV data are obtained from The Oxford-Man Institute's "realized library". The logarithmic transformation is applied to RV to induce Gaussianity (Andersen, et al, 2001). The time series plot of the logarithmic daily RV data is given in Fig. 1.

The OU model can be expressed as

$$\begin{aligned} dX(t) &= \mathscr{K}\left(\theta - X(t)\right) dt + \Sigma^{1/2} dW(t) \\ &= (AX(t) + B) dt + \Sigma^{1/2} dW(t), \end{aligned}$$

where $A = -\mathscr{K}$, $B = \mathscr{K}\theta$ are 3×3 and 3×1 matrices. The exact discrete time representation is

$$X_t = FX_{t-1} + g + \varepsilon_t, \text{where } F = e^{A/252}.$$

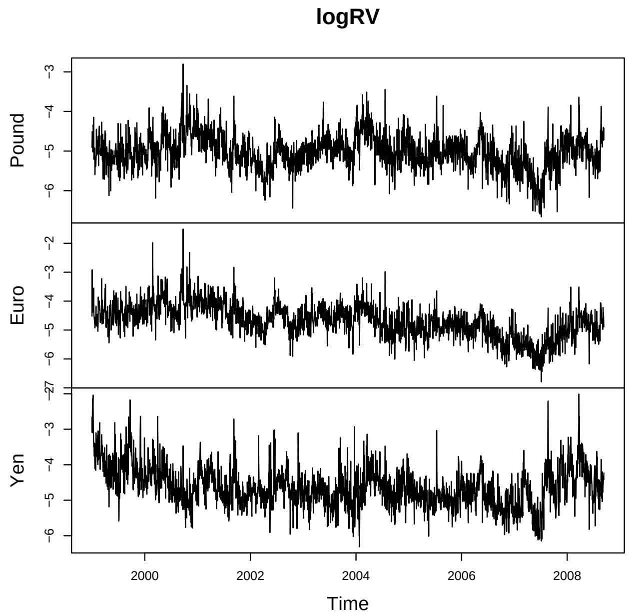

Figure 3.1: Time series plot of the logarithmic daily RV of Pound, Euro and Yen exchange rates, all against the US dollar, from January 4, 1999 to August 31, 2008.

The ML/LS estimates of F, $\hat{F}$, is given by

$$\hat{F} = \begin{pmatrix} 0.5155 & 0.0997 & -0.0186 \\ -0.1032 & 0.7847 & -0.0393 \\ -0.0669 & 0.0300 & 0.6699 \end{pmatrix}.$$

Using (3.2.12) with $\hat{F}$ calculated as in (3.2.4), the estimated $\mathscr{K}$ is

$$\hat{\mathscr{K}} = \begin{pmatrix} 163.76(10.55) & -39.06(7.87) & 6.72(5.92) \\ 41.02(11.41) & 58.02(8.17) & 14.10(6.29) \\ 27.45(12.34) & -12.29(9.03) & 100.99(6.51) \end{pmatrix}. \tag{3.5.1}$$

From Theorem 3.3.3, $h\sqrt{T}Vec\left(\hat{A}-A\right) \xrightarrow{d} N\left(0, \Gamma V_F \Gamma'\right)$. By using the explicit expression of Γ given in Theorem 3.3.3, the consistent estimator of the asymptotic covariance of $\hat{A}$, $\hat{\Gamma}\hat{V}_F'\hat{\Gamma}/\left(Th^2\right)$, is reported in Table 1. The estimated standard errors of all the elements in $\hat{\mathscr{K}}$ are reports in parenthesis in (3.5.1).

Table 1. Estimated Covariance Matrix of $Vec\left(\hat{\mathscr{K}} - \mathscr{K}\right)$

$d\mathscr{K}_{ij}$ denotes the ij^{th} element of the matrix $\hat{\mathscr{K}} - \mathscr{K}$.

	$d\mathscr{K}_{11}$	$d\mathscr{K}_{12}$	$d\mathscr{K}_{13}$	$d\mathscr{K}_{21}$	$d\mathscr{K}_{22}$	$d\mathscr{K}_{23}$	$d\mathscr{K}_{31}$	$d\mathscr{K}_{32}$	$d\mathscr{K}_{33}$
$d\mathscr{K}_{11}$	111.38	-58.48	-9.25	90.59	-50.00	-7.30	70.03	-35.48	-9.74
$d\mathscr{K}_{12}$	-58.48	61.98	-12.30	-46.12	49.01	-9.68	-35.62	37.57	-7.14
$d\mathscr{K}_{13}$	-9.25	-12.30	35.08	-7.21	-9.74	27.56	-5.56	-7.46	21.20
$d\mathscr{K}_{21}$	90.59	-46.12	-7.21	130.27	-66.45	-10.35	87.28	-44.28	-7.25
$d\mathscr{K}_{22}$	-50.00	49.01	-9.74	-66.45	66.79	-14.03	-44.49	46.91	-12.76
$d\mathscr{K}_{23}$	-7.30	-9.68	27.56	-10.35	-14.03	39.59	-6.92	-9.30	26.38
$d\mathscr{K}_{31}$	70.03	-35.62	-5.56	87.28	-44.49	-6.92	152.15	-77.21	-12.28
$d\mathscr{K}_{32}$	-35.48	37.57	-7.46	-44.28	46.91	-9.30	-77.21	81.60	-16.10
$d\mathscr{K}_{33}$	-9.74	-7.14	21.20	-7.25	-12.76	26.38	-12.28	-16.10	42.42

The ordered eigenvalues of $\hat{F}$ are $\left(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3\right) = (0.7386, 0.6705, 0.561)$, and the eigenvalues of $\hat{\mathscr{K}}$ are

$$-\hat{\eta} = -\left(\ln\left(\hat{\lambda}_1\right), \ln\left(\hat{\lambda}_2\right), \ln\left(\hat{\lambda}_3\right)\right)/h$$

$= (76.36, 100.71, 145.68)$ .Table 2 reports the estimated asymptotic covariance matrix of $-\hat{\eta}$ and Table 3 the 95% confidence intervals of $-\eta$.

The empirical results may be summarized as follows. First, in all three series, RV at period t significantly depends on RV at period $t-1$, featured by large values of t-ratio (15.5, 7.1 and 15.5). Second, all elements, with the exception two, in $\hat{\mathscr{K}}$ is statistically significant. The two insignificant elements in $\mathscr{K}$ are $\mathscr{K}_{13}$ and $\mathscr{K}_{32}$, suggesting that the RV of Pound at period t does not significantly depend on the RV of Yen at period $t-1$ and the RV of Yen at period t does not significantly depend on the RV of Euro at period $t-1$. Third, since the 95% confidence intervals of η all exclude 0, suggesting strong evidence against a unit root in the three series. Fourth, the point estimates in $\hat{\eta}$ are very large, implies strong mean reversion in all the series. In fact, using these point estimates, we can calculate the estimated half lives of a shock to volatility for the three exchange rates, which are 0.1089, 0.0826, and 0.0571 months for Pound, Euro and Yen, respectively. The estimated half lives are short.

Table 2. Estimated Covariance Matrix of $-\hat{\eta}$

	$-\hat{\eta}_1$	$-\hat{\eta}_2$	$-\hat{\eta}_3$
$-\hat{\eta}_1$	18.4334	-3.7412	-4.3002
$-\hat{\eta}_2$	-3.7412	28.0748	-4.1163
$-\hat{\eta}_3$	-4.3002	-4.1163	53.3981

Table 3. 95% Confidence Intervals of $-\eta$

	$-\eta_1$	$-\eta_2$	$-\eta_3$
Confidence Interval	(67.95,84.78)	(90.33,111.10)	(131.36,160.01)

3.6 Conclusions

This paper derives the asymptotic distribution of the ML/LS estimator of the mean reversion matrix in a multivariate diffusion model with a linear drift and a constant diffusion when only discretely sampled data are available. Both the stationary case and the unit root case are examined. The limit theory gives an analytic expression of the asymptotic covariance matrix, for which a consistent estimator is provided thereby facilitating inference about the mean reversion matrix. Our method replies on the asymptotic theory of the ML/LS estimator of the exact discrete time VAR model.

The transformation from the continuous time model to the exact discrete system involves a nonlinear matrix logarithmic mapping. The mean reversion matrix is shown to be identified under a weak condition. When identification is achieved, our method also utilizes a novel explicit relationship between the AR coefficient matrix and the mean reversion matrix. This relationship is a polynomial of an finite order, facilitating the use of the delta method and the calculation of the covariance matrix in the limit distribution.

Both in the stationary case and in the unit root case, we develop the limit theory of the ML/LS estimator of the mean reversion matrix by using the limit distribution of the estimated AR coefficient matrix only. The expression of the asymptotic covariance matrix in stationary case is a little complicated. Different situations have been discussed to get an explicit representation of the asymptotic covariance matrix. For models of low dimension, such as $m \leq 3$, using our framework, the mean reversion matrix is shown to have a straightforward expression as a continuously differentiable mapping of the AR coefficient matrix.

The new theory is illustrated in an empirical application to a multivariate OU model for the logarithmic daily realized volatility (RV) of Pound, Euro and Yen exchange rates. Using our method, we are able to obtain the estimate of the asymptotic covariance of the mean reversion matrix. The statistical inferences, conducted on these covariances, suggest that the three series are stationary and revert to their means in fast rates, and that the RV of the Pound does not depend on the lagged RV of the Yen and the RV of the Yen does not depend on the lagged RV of the Yen.

Although in the present paper we only develop the asymptotic theory for multivariate diffusion models with a linear drift and a constant diffusion, our method is generally applicable to continuous time models with a linear drift but with a more flexible diffusion function and to continuous time models which are driven by Lévy process. In this case, OLS may be applied to estimate the AR coefficient matrix of the exact discrete time system. As long as the asymptotic theory of the OLS estimator is available, our method can be applied in the same manner.

Chapter 4 Double Asymptotics for Explosive Continuous Time Models

4.1 Introduction

Continuous time models driven by the Brownian motion, i.e., diffusion processes, have found wide applications in science and social science. An important property of diffusion processes is that, under some smoothness condition on the drift function and the diffusion function, the sample path is continuous everywhere. This restriction is often found to be too strong in applications. There are different ways to introduce discontinuity into the continuous time models. For example, Poisson processes, which allow for a finite number of jumps in a finite time interval, have been used to model jumps in finance (Merton, 1976). In recent years, however, strong evidence of the presence of infinite activity jumps have been documented in finance; see, for example, Aït-Sahalia and Jacod (2011). Consequently, continuous time Lévy processes have become increasingly popular to model discontinuity in financial time series. Not surprisingly, various Lévy processes have been developed in the asset pricing literature (see, for example, Barndorff-Nielsen (1998), Madan, Carr and Chang (1998), Carr and Wu (2003)).

Independent to the development in continuous time modelling, there has been a long-standing interest in statistics for developing the asymptotic theory for explosive processes. Two of the earliest studies are White (1958) and Anderson (1959) where the asymptotic distribution of the autoregressive (AR) coefficient was derived when the root is larger than unity. Phillips and Magdalinos (2007, PM hereafter) has provided an asymptotic theory and an invariance principle for mildly explosive processes where the root is moderately deviated from unity. Magdalinos (2011) extended the result to the case where the error is serially dependent. Anu and Horvath (2007) extended the result to the case where the error is infinite. In economics, there has recently been a growing interest on using explosive processes to model asset price bubbles. Phillips et al (2011) has developed a recursive method to detect bubbles in the discrete time AR model. Phillips and Yu (2011) applied the method to analyze the bubble episodes in various markets in the U.S. and documented the bubble migration mechanism during the subprime crisis.

All the above cited studies on explosiveness focus exclusively on discrete time models. Explosive behavior can also be described using continuous time models. Let T, h, N be the sample size, the sampling interval, and the time span of the data, respectively. Obviously $T = N/h$. While the asymptotic theory in discrete time models always corresponds to the scheme of $T \to \infty$, how to develop the asymptotic theory in continuous time is less a clear cut because $T \to \infty$ is achievable from different ways. In the literature, three alternative sampling schemes have been discussed (see, for example, Jeong and Park (2011) and Zhou and Yu (2011)), namely:

$$N \to \infty, \ \ h \ \text{ is fixed;} \qquad (A1)$$

$$N \to \infty, \ \ h \to 0; \qquad (A2)$$

$$h \to 0, \ \ N \text{ is fixed.} \qquad (A3)$$

The main purpose of the present paper is to develop the double asymptotic theory under scheme (A2) for explosive continuous time models driven by Lévy processes, in which $N \to \infty$ and $h \to 0$ simultaneously. In the special case of Brownian motion driven continuous time models, three alternative double asymptotics are considered. In the first case, $N \to \infty$ and $h \to 0$ simultaneously. In the second case, a sequential asymptotic treatment is considered, i.e., $N \to \infty$ is followed by $h \to 0$. In the third case, another sequential asymptotic treatment is considered wherein, $h \to 0$

is followed by $N \to \infty$. We show that the asymptotic distributions under these three treatments are the same. Different from PM, in our double asymptotic distribution, the initial condition, either fixed or random, appears in the limiting distribution.

The paper is organized as follows. Section 2 develops the double asymptotic distribution of the persistent parameter in the explosive Ornstein-Uhlenbeck (OU) process driven by Lévy processes with $N \to \infty$ and $h \to 0$ simultaneously. Section 3 develops the sequential asymptotic distribution where $N \to \infty$ is followed by $h \to 0$ for the explosive Brownian motion driven OU process. Section 4 develops the sequential asymptotic distribution of the same process when $h \to 0$ is followed by $N \to \infty$. Section 5 concludes. Appendix collects the proof of the theoretical results.

4.2 Simultaneous Double Asymptotics for Explosive Lévy Processes

Consider the following Lévy-driven OU process:

$$dy(t) = \kappa(\mu - y(t))\,dt + \sigma dL(t),\ y(0) = y_0, \tag{4.2.1}$$

where $(L(t))_{t\geq 0}$ is a Lévy process defined on $\left(\Omega, \mathscr{F}, \{\mathscr{F}\}_{t\geq 0}, P\right)$ with $L(0) = 0$ $a.s.$ and satisfies the following three properties:

1. Independent increments: for every increasing sequence of times $t_0, \cdots, t_n$ the random variables $L(t_0), L(t_1) - L(t_0), \cdots, L(t_n) - L(t_{n-1})$ are independent;

2. Stationary increments: the law of $L(t+h) - L(t)$ is independent of t.

3. Stochastic continuity: for all $\varepsilon > 0$, $\lim_{h\to 0} P(|L(t+h) - L(t)| \geq \varepsilon) = 0$. For a given t, the probability of seeing a jump at t is zero. In other words, jumps happen at random times.

Every Lévy process has a unique modification which is càdlàg (right continuous with left limits) and which is also a Lévy process. We shall therefore assume that our Lévy process has these properties.

In the special case when $(L(t))_{t\geq 0}$ is Brownian motion, the stochastic process (4.2.1) is interpreted as an Itô equation with solution $\{y(t), t \geq 0\}$ satisfying

$$y(t) = e^{-\kappa h} y(0) + \mu \left[1 - e^{-\kappa h}\right] + \sigma \int_0^t e^{-\kappa(t-s)} dW(s),$$

where $W(t)$ is a standard Brownian motion and the integral is defined as the L_2 limit of approximating Rieman-Stieltjes sums. For the second-order driving Lévy process, $\{y(t), t \geq 0\}$ can be defined in the same way. And $\{y(t), t \geq 0\}$ can also be defined pathwise as a Rieman-Stieltjes integral, when the paths of $(L(t))_{t\geq 0}$ are almost surely of finite variation on compact intervals (Sato, 1999, Theorem 21.9).

When $y(t)$ is assumed to be observed at discrete points in time, say $t = 0, 1, 2, \ldots, T$, the exact discrete time model corresponding to (4.2.1) is

$$y_{th} = e^{-\kappa h} y_{(t-1)h} + \mu \left[1 - e^{-\kappa h}\right] + \sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)} dL(s).$$

By the properties of Lévy process, the sequence of $\left\{\sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)} dL(s)\right\}_{t=1}^{T}$ consists of independent and identically distributed (*iid*) random variables.

The characteristic function of $(L(t))_{t\geq 0}$ is $E(\exp\{isL(t)\}) = \exp\{-t\psi(s)\}$, where $\psi(\cdot): \mathbb{R} \to \mathbb{C}$ is referred to as the Lévy exponent of $(L(t))_{t\geq 0}$. For the square-integrable process $(L(t))_{t\geq 0}$, it is known that

$$i\psi'(0) = E[L(1)] = \frac{E[L(t)]}{t}, \tag{4.2.2}$$

$$\psi''(0) = Var[L(1)] = \frac{Var[L(t)]}{t}. \tag{4.2.3}$$

Therefore,

$$E\left(\sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)} dL(s)\right) = \sigma i \psi'(0) \frac{1 - e^{-\kappa h}}{\kappa},$$

and

$$Var\left(\sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)} dL(s)\right) = \sigma^2 \psi''(0) \frac{1 - e^{-2\kappa h}}{2\kappa}.$$

Let

$$g_h = \left[\mu + \frac{\sigma i \psi'(0)}{\kappa}\right]\left[1 - e^{-\kappa h}\right],$$

$$\sigma\sqrt{\psi''(0)\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_{th} = \sigma\int_{(t-1)h}^{th} e^{-\kappa(th-s)}dL(s) - \sigma i\psi'(0)\frac{1-e^{-\kappa h}}{\kappa}.$$

We rewrite the exact discrete time model corresponding to (4.2.1) as

$$y_{th} = a_h(\kappa)y_{(t-1)h} + g_h + \sigma\sqrt{\psi''(0)\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_{th},\ y_{0h} = y_0, \tag{4.2.4}$$

where $a_h(\kappa) = e^{-\kappa h}$, $\{\varepsilon_{th}\}_{t=1}^{T} \overset{iid}{\sim} (0,1)$ whose distribution depends on the specification of the Lévy measure of $L(t)$. It should be pointed out that $\{\varepsilon_{th}\}_{t=1}^{T}$ is a martingale-difference array, because in general the distribution of ε_{th} depends on the sampling interval h, although the first two moments of ε_{th} do not.

In the paper, we focus our analysis on the explosive case, $\kappa < 0$, which means that $a_h(\kappa) > 1$. The initial value, y_{0h}, may be a random variable, whose distribution is fixed and independent of the sampling interval h, or a constant. The least squares (LS) estimators of $a_h(\kappa)$ and κ are, respectively,

$$\widehat{a}_h(\kappa) = \frac{T\sum_{t=1}^{T} y_{(t-1)h}y_{th} - \left(\sum_{t=1}^{T} y_{th}\right)\left(\sum_{t=1}^{T} y_{(t-1)h}\right)}{T\sum_{t=1}^{T} y_{(t-1)h}^2 - \left(\sum_{t=1}^{T} y_{(t-1)h}\right)^2},$$

and

$$\widehat{\kappa} = -\frac{1}{h}\ln\left(\widehat{a}_h(\kappa)\right). \tag{4.2.5}$$

Letting $\lambda(h) = \sigma\sqrt{\psi''(0)\frac{1-e^{-2\kappa h}}{2\kappa}}$ which has the order $O(\sqrt{h})$, $x_{th} = y_{th}/\lambda(h)$, $x_{0h} = y_{0h}/\lambda(h)$, $\tilde{g}_h = g_h/\lambda(h)$, and dividing both sides of Model (4.2.4) by $\lambda(h)$, we get the following explosive AR(1) model

$$x_{th} = a_h(\kappa)x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th} = e^{-\kappa h}x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th},\ \ x_{0h} = y_{0h}/\lambda(h). \tag{4.2.6}$$

This model compares to Model (1) in PM,

$$x_t = \left(1 + \frac{-\kappa}{k_T}\right) x_{t-1} + \varepsilon_t,\ x_0 = o_p\left(\sqrt{k_T}\right),\ k_T \to \infty,\ \frac{k_T}{T} \to 0. \tag{4.2.7}$$

Let $k_T = 1/h$ so that Model (4.2.6) may be written as

$$x_{th} = a_h(\kappa) x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th} = e^{-\kappa/k_T} x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th},\ \ x_{0h} = y_{0h}/\lambda(h) = O_p\left(\sqrt{k_T}\right), \tag{4.2.8}$$

hence, $\widehat{a}_h(\kappa)$ and $\widehat{\kappa}$ can also be obtained from x_{th}, and

$$\widehat{a}_h(\kappa) - a_h(\kappa) = \frac{T \sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} - \left(\sum_{t=1}^{T} \varepsilon_{th}\right)\left(\sum_{t=1}^{T} x_{(t-1)h}\right)}{T \sum_{t=1}^{T} x_{(t-1)h}^2 - \left(\sum_{t=1}^{T} x_{(t-1)h}\right)^2}.$$

The double asymptotics, $h \to 0$ and $N \to \infty$, implies that

$$k_T = \frac{1}{h} \to \infty \quad \text{and} \quad \frac{k_T}{T} = \frac{1}{N} \to 0. \tag{4.2.9}$$

Model (4.2.8) is similar to Model (1) of PM but with four subtle differences. First, Model (4.2.8) includes an additional intercept term comparing with Model (1) of PM. Second, the AR coefficient in (4.2.8) is $e^{-\kappa/k_T}$ whereas it is $1 + (-\kappa/k_T)$ in PM. This difference is small since $e^{-\kappa/k_T} = 1 + (-\kappa/k_T) + O(k_T^{-2})$ and, not surprisingly, it has no impact on the limiting distribution. Third, $\{\varepsilon_{th}\}_{t=1}^{T}$ is a martingale-difference array, whereas it is assumed to be a sequence with *iid* random variables in PM. Fourth, the initial condition in (4.2.8) is $x_{0h} \sim O_p\left(\sqrt{k_T}\right)$, whereas it is assumed to be $o_p\left(\sqrt{k_T}\right)$ in PM.

In Model (1) of PM, the root, $1 + (-\kappa/k_T)$, represents moderate deviations from unity in the sense that it is in a larger neighborhood of one than the conventional local to unity root, $1 + (-\kappa/T)$. Therefore, under the double asymptotics the root in Model (4.2.8) is also moderately deviated from unity. With a different initial condition, our analysis can be regarded as an extension to PM from $x_{0h} \sim o_p\left(\sqrt{k_T}\right)$ to $x_{0h} \sim O_p\left(\sqrt{k_T}\right)$. It turns out this change of the order of magnitude in the initial condition leads to a change in the limiting distribution of the LS estimators, $\widehat{a}_h(\kappa)$ and $\widehat{\kappa}$.

Let

$$X_{Th} = \frac{1}{\sqrt{k_T}} \sum_{t=1}^{T} (a_h(\kappa))^{-(T-t)-1} \varepsilon_{th}, \tag{4.2.10}$$

$$Y_{Th} = \frac{1}{\sqrt{k_T}} \sum_{t=1}^{T} (a_h(\kappa))^{-t} \varepsilon_{th}. \tag{4.2.11}$$

We can obtain the following lemma.

Lemma 4.2.1 *Let $a_h(\kappa) = e^{-\kappa/k_T}$, $k_T = 1/h$, $T = N/h$. For some $\delta > 0$ and a constant M, assume that the martingale-difference array, $\{\varepsilon_{th}\}_{t=1}^{T} \overset{iid}{\sim} (0,1)$, satisfies $E\left(|\varepsilon_{th}|^{2+\delta}\right) < M$ for small h.*[1] *When $h \to 0$ and $N \to \infty$, we have*

(a)

$$(a_h(\kappa))^{-T} = o\left(\frac{k_T}{T}\right) = o\left(\frac{1}{N}\right);$$

(b)

$$\frac{(a_h(\kappa))^{-T}}{k_T} \sum_{t=1}^{T} \sum_{j=t}^{T} (a_h(\kappa))^{t-j-1} \varepsilon_{jh} \varepsilon_{th} \xrightarrow{L_1} 0;$$

(c)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{th} \implies N(0,1);$$

(d)

$$(X_{Th}, Y_{Th}) \implies (X, Y),$$

where X and Y are independent $N\left(0, \frac{1}{-2\kappa}\right)$ random variables.

Note that $x_{0h}/\sqrt{k_T} \xrightarrow{L_1} y_{0h}/\left(\sigma\sqrt{\psi''(0)}\right)$ and $\sqrt{k_T}\tilde{g}_h \to (\kappa\mu + \sigma i \psi'(0)) / \left(\sigma\sqrt{\psi''(0)}\right)$. Theorem 4.2.1 reports the double asymptotic distribution of $\widehat{a}_h(\kappa)$ with $h \to 0$ and $N \to \infty$ simultaneously.

[1] This condition excludes the stable process whose index parameter is less than 2, although Model (4.2.1) allows $L(t)$ to be a stable process.

Theorem 4.2.1 *Let $k_T = 1/h$, $a_h(\kappa) = e^{-\kappa h}$, $\widehat{a}_h(\kappa)$ be the LS estimator obtained from x_{th}, $\widehat{\kappa} = -(1/h)\ln(\widehat{a}_h(\kappa))$. For some $\delta > 0$ and a constant M, assume that the martingale-difference array, $\{\varepsilon_{th}\}_{t=1}^{T} \overset{iid}{\sim} (0,1)$, satisfies $E\left(|\varepsilon_{th}|^{2+\delta}\right) < M$ for small h. Under the simultaneous double asymptotics, we have*

(a)

$$\frac{(a_h(\kappa))^{-T}[a_h(\kappa)-1]}{\sqrt{k_T}}\sum_{t=1}^{T} x_{(t-1)h} \Longrightarrow \sqrt{\frac{1}{-2\kappa}}[\eta + D];$$

(b)

$$\frac{(a_h(\kappa))^{-T}}{k_T}\sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} \Longrightarrow \frac{1}{-2\kappa}\xi\,[\eta + D];$$

(c)

$$\frac{(a_h(\kappa))^{-2T}\left[(a_h(\kappa))^2-1\right]}{k_T}\sum_{t=1}^{T} x_{(t-1)h}^2 \Longrightarrow \frac{1}{-2\kappa}[\eta + D]^2;$$

(d)

$$\frac{(a_h(\kappa))^{T}}{\left[(a_h(\kappa))^2-1\right]}(\widehat{a}_h(\kappa) - a_h(\kappa)) \Longrightarrow \frac{\xi}{\eta + D};$$

(e)

$$\frac{e^{-\kappa N}}{2\kappa}(\widehat{\kappa} - \kappa) \Longrightarrow \frac{\xi}{\eta + D},$$

where $\xi = \sqrt{-2\kappa}X$, $\eta = \sqrt{-2\kappa}Y$ are independent $N(0,1)$ random variables with (X,Y) defined in Lemma 4.2.1 and $D = \sqrt{2}\,(\kappa\mu + \sigma i\psi'(0) - \kappa y_{0h}) / \left(\sigma\sqrt{-\kappa\psi''(0)}\right)$.

Remark 4.2.1 *If the long run mean is zero (i.e., $\mu = 0$), the mean of the Lévy process is zero (i.e., $i\psi'(0) = E(L(1)) = 0$), and the initial condition, y_{0h}, is also zero, we get $D = 0$,*

$$\frac{(a_h(\kappa))^{T}}{\left[(a_h(\kappa))^2-1\right]}(\widehat{a}_h(\kappa) - a_h(\kappa)) \Longrightarrow Cauchy,$$

and

$$\frac{e^{-\kappa N}}{2\kappa}(\widehat{\kappa} - \kappa) \Longrightarrow Cauchy.$$

Remark 4.2.2 *For the discrete time explosive AR(1) model without intercept, Anderson (1959) showed that the limiting distribution is dependent on the distribution* *of the errors and no invariance principle applies. Only under the assumption that the error distribution is Gaussian, was he able to show that the limiting distribution is Cauchy. However, the results in Lemma 4.2.1 and Theorem 4.2.1 suggest that although the invariance principle does not cover the discrete time explosive model, it covers the continuous time explosive model under the simultaneous double asymptotics.*

Remark 4.2.3 *Let $\widehat{g}_h$ be the LS estimator of g_h in Model (4.2.4), the t statistic is*

$$t = \frac{\hat{a}_h(\kappa) - a_h(\kappa)}{\hat{\delta}_{\hat{a}_h(\kappa)}} = \frac{[\hat{a}_h(\kappa) - a_h(\kappa)] \left\{ T \sum_{t=1}^{T} y_{(t-1)h}^2 - \left(\sum_{t=1}^{T} y_{(t-1)h} \right)^2 \right\}^{1/2}}{\left\{ \sum_{t=1}^{T} \left[y_{th} - \hat{a}_h(\kappa) y_{(t-1)h} - \widehat{g}_h \right]^2 \right\}^{1/2}}.$$

It can be identically expressed as

$$t = \frac{T \sum_{t=1}^{T} x_{(t-1)h} \varepsilon_{th} - \left(\sum_{t=1}^{T} \varepsilon_{th} \right) \left(\sum_{t=1}^{T} x_{(t-1)h} \right)}{\left\{ \sum_{t=1}^{T} \left[x_{th} - \hat{a}_h(\kappa) x_{(t-1)h} - \widehat{\widetilde{g}}_h \right]^2 \right\}^{1/2} \left\{ T \sum_{t=1}^{T} x_{(t-1)h}^2 - \left(\sum_{t=1}^{T} x_{(t-1)h} \right)^2 \right\}^{1/2}}$$

where $\widehat{\widetilde{g}}_h$ is LS estimator of $\widetilde{g}_h$ in Model (4.2.8). By using the Law of Large Number of the martingale-difference array (see e.g. Hall and Heyde, 1980, Theorem 2.23), it can be shown that

$$\frac{1}{T} \sum_{t=1}^{T} \left[x_{th} - \hat{a}_h(\kappa) x_{(t-1)h} - \widehat{\widetilde{g}}_h \right]^2 \xrightarrow{p} 1$$

Based on the results in Lemma 4.2.1 and Theorem 4.2.1, we may show that

$$t \Longrightarrow \xi \sim N(0,1). \tag{4.2.12}$$

Remark 4.2.4 *Using Lemma 4.2.1 and Theorem 4.2.1, we can obtain the following* *results under the simultaneous double asymptotics,*

$$\frac{1}{N}\sum_{t=1}^{T}\left[y_{th}-\hat{a}_h(\kappa)\,y_{(t-1)h}-\hat{g}_h\right]^2 \xrightarrow{p} \sigma^2\psi''(0),$$

$$\hat{g}_h/h \xrightarrow{p} \kappa\mu+\sigma i\psi'(0),$$

which in turn give a consistent estimator of D.

In the special case of Brownian motion driven OU processes with a known mean (without the loss of generality, it is assumed to be zero),

$$dy(t) = -\kappa y(t)dt+\sigma dW(t),\ y(0)=y_0, \tag{4.2.13}$$

the exact discrete time model (4.2.4) becomes

$$y_{th} = a_h(\kappa)y_{(t-1)h}+\sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_{th},\ y_{0h}=y_0, \tag{4.2.14}$$

where $a_h(\kappa)=e^{-\kappa h}$, $\{\varepsilon_{th}\}_{t=1}^{T} \overset{iid}{\sim} N(0,1)$. The LS estimators of $a_h(\kappa)$ and κ are, respectively,

$$\widehat{a}_h(\kappa)=\frac{\sum_{t=1}^{T} y_{(t-1)h}y_{th}}{\sum_{t=1}^{T} y_{(t-1)h}^2},$$

and

$$\widehat{\kappa} = -\frac{1}{h}\ln\left(\widehat{a}_h(\kappa)\right). \tag{4.2.15}$$

Let $k_T=1/h$, rewrite the Model (4.2.6) as

$$x_{th}=a_h(\kappa)x_{(t-1)h}+\varepsilon_{th}=e^{-\kappa/k_T}x_{(t-1)h}+\varepsilon_{th},\ \ x_{0h}=y_{0h}/\lambda(h)=O_p\left(\sqrt{k_T}\right), \tag{4.2.16}$$

The limit theory for this traditional OU process is reported in Corollary 4.2.2.

Corollary 4.2.2 *Let* $k_T=1/h$, $a_h(\kappa)=e^{-\kappa h}$*,* $\widehat{a}_h(\kappa)$ *be the LS estimator from (4.2.16),* $\widehat{\kappa}=-(1/h)\ln(\widehat{a}_h(\kappa))$*. Under the simultaneous double asymptotics, we have*

(a)

$$\frac{(a_h(\kappa))^T}{\left[(a_h(\kappa))^2-1\right]}\left(\widehat{a}_h(\kappa)-a_h(\kappa)\right)\Longrightarrow\frac{\xi}{\eta+d}; \tag{4.2.17}$$

(b)

$$\frac{e^{-\kappa N}}{2\kappa}\left(\widehat{\kappa}-\kappa\right)\Longrightarrow\frac{\xi}{\eta+d}, \tag{4.2.18}$$

where ξ, η *are independent* $N(0,1)$ *random variables, and* $d=y_{0h}\sqrt{-2\kappa}/\sigma$.

Remark 4.2.5 *To facilitate the comparison of our results with those of PM, we may rewrite the limit theory in (4.2.17) as*

$$\frac{(a_h(\kappa))^T k_T}{-2\kappa}\left(\widehat{a}_h(\kappa)-a_h(\kappa)\right)\Longrightarrow\frac{X}{Y+y_0/\sigma}, \tag{4.2.19}$$

where X, Y *are defined in Lemma 4.2.1. When* $y_0=0$*, the limiting distribution is Cauchy and the same as in PM. Since the finite sample distribution always depends on the initial value in continuous time models, we expect that the double asymptotic distribution in (4.2.19) is a better approximation than the Cauchy distribution when* y_0 *is different from* 0.

4.3 Sequential Asymptotics: $N\to\infty$ Followed by $h\to 0$

Focusing on the explosive OU process driven by the Brownian motion, we now study two alternative sequential limit theory in the rest of the paper. This section develops the sequential asymptotic distribution where $N\to\infty$ is followed by $h\to 0$. In Section 4 we will develop the sequential asymptotic distribution where $h\to 0$ is followed by $N\to\infty$.

When h is fixed, the discrete time model (4.2.16) is an explosive AR(1) model with Gaussian errors. Letting $N \to \infty$, Anderson (1959) showed that

$$\frac{(a_h(\kappa))^T\left[\widehat{a}_h(\kappa)-a_h(\kappa)\right]}{(a_h(\kappa))^2-1} \Longrightarrow \frac{Y_a}{Z_a+a_h(\kappa)x_0} \overset{d}{=} \frac{N\left(0,1/\left[1-(a_h(\kappa))^{-2}\right]\right)}{N\left(0,1/\left[1-(a_h(\kappa))^{-2}\right]\right)+a_h(\kappa)x_0},$$

where Y_a and Z_a are independent. The proof was done under the condition that x_{0h} is a constant, but it still holds when $x_{0h} \sim O_p(1)$. It is straightforward to show that

$$\frac{Y_a}{Z_a+a_h(\kappa)x_{0h}} \overset{d}{=} \frac{N(0,1)}{N(0,1)+x_{0h}\sqrt{(a_h(\kappa))^2-1}} \overset{d}{=} \frac{\xi}{\eta+d},$$

because $d = y_{0h}\sqrt{-2\kappa}/\sigma$, and

$$x_{0h} = \frac{y_{0h}}{\lambda(h)} = \frac{y_{0h}}{\sigma}\sqrt{\frac{-2\kappa}{(a_h(\kappa))^2-1}}.$$

Letting $h \to 0$, the sequential limiting distribution is

$$\lim_{h\to 0}\lim_{N\to\infty}\frac{(a_h(\kappa))^T\left[\widehat{a}_h(\kappa)-a_h(\kappa)\right]}{(a_h(\kappa))^2-1} = \lim_{h\to 0}\frac{Y_a}{Z_a+a_h(\kappa)x_0} \overset{d}{=} \frac{\xi}{\eta+d},$$

which is the same as the double asymptotic distribution derived in Section 2. We now collect these results together in the following theorem.

Theorem 4.3.1 *Let $a_h(\kappa) = e^{-\kappa h}$, $\widehat{a}_h(\kappa)$ be the LS estimator obtained from x_{th} in model (4.2.16), $\widehat{\kappa} = -(1/h)\ln(\widehat{a}_h(\kappa))$, we have*

(a)

$$\lim_{h\to 0}\lim_{N\to\infty}\frac{(a_h(\kappa))^T\left[\widehat{a}_h(\kappa)-a_h(\kappa)\right]}{(a_h(\kappa))^2-1} \overset{d}{=} \frac{\xi}{\eta+d};$$

(b)

$$\lim_{h\to 0}\lim_{N\to\infty}\frac{\exp\{-\kappa N\}}{2\kappa}(\widehat{\kappa}-\kappa) \overset{d}{=} \frac{\xi}{\eta+d},$$

where ξ, η are independent $N(0,1)$ random variables and $d = y_0\sqrt{-2\kappa}/\sigma$.

Remark 4.3.1 *Although the sequential asymptotic theory developed here is the same as that developed in Section 2 for the explosive Lévy process, there is a* *clear advantage of deriving the asymptotic theory under the assumption of $N \to \infty$ and $h \to 0$ simultaneously, that is, the invariance principle can be established.*

4.4 Sequential Asymptotics: $h \to 0$ Followed by $N \to \infty$

Perron (1991) derived a sequential limiting distribution for the LS estimator of the persistent parameter κ in the explosive OU process driven by the Brownian motion; see Corollary 1 (ii) on Page 217 and the corresponding proof on Page 234 in Perron (1991). The sequential asymptotics first requires $h \to 0$ and then $N \to \infty$. To our surprise, however, his sequential limiting distribution is different from the limiting distributions that we obtained in Section 2 and Section 3. It is important to find the reasons that cause this discrepancy. In this section we investigate the double asymptotic theory under the sequential limits where $h \to 0$ is followed by $N \to \infty$.

The continuous time OU process considered in Perron is given in (4.2.13) where the initial condition is assumed to be constant, $y_0 = b$. First, by letting time interval h goes to zero with fixed time span N, Perron developed the in-fill asymptotics for $\widehat{a}_h(\kappa)$,

$$T\left(\widehat{a}_h(\kappa) - a_h(\kappa)\right) \Longrightarrow \frac{A(\gamma,c)}{B(\gamma,c)}, \tag{4.4.1}$$

where

$$A(\gamma,c) = \gamma \int_0^1 \exp\{cr\}\, dW(r) + \int_0^1 J_c(r)\, dW(r), \tag{4.4.2}$$

$$B(\gamma,c) = \gamma^2 \frac{\exp\{2c\}-1}{2c} + 2\gamma \int_0^1 \exp\{cr\} J_c(r)\, dr + \int_0^1 J_c(r)^2\, dr, \tag{4.4.3}$$

and $J_c(r) = \int_0^r \exp\{c(r-s)\}\, dW(s)$ is generated by the stochastic differential equation

$$dJ_c(r) = cJ_c(r)\, dr + dW(r),$$

with the initial condition $J_c(0) = 0$, $c = -\kappa N$, $\gamma = b/\left(\sigma\sqrt{N}\right)$.

To derive the sequential limiting distribution, he then let $N \to \infty$, namely, $c \to$

$+\infty$, and showed that (see (vi) and (viii) of Lemma A.2 in his paper)

$$(2c)^{3/2}e^{-2c}\int_0^1 \exp\{cr\}J_c(r)\,dr \Longrightarrow N(0,1),$$

and

$$(2c)^{1/2}e^{-c}\int_0^1 \exp\{cr\}\,dW(r) \Longrightarrow N(0,1).$$

Then he argued, without a proof, that these two limiting distributions are identical (call it η). Based on this argument and the two results in Phillips (1987), Perron obtained the limiting distributions of $A(\gamma,c)$ and $B(\gamma,c)$, and the sequential limiting distribution for $\widehat{a}_h(\kappa)$ and $\widehat{\kappa}$,

$$\lim_{N\to\infty}\lim_{h\to 0}\frac{e^{-\kappa N}\left(\widehat{a}_h(\kappa)-a_h(\kappa)\right)}{-2\kappa h}=\lim_{c\to\infty}\frac{(2c)\,e^{-c}A(\gamma,c)}{(2c)^2e^{-2c}B(\gamma,c)}=\frac{d\eta+\xi\eta}{[d+\eta]^2}, \quad (4.4.4)$$

$$\lim_{N\to\infty}\lim_{h\to 0}\frac{e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)}{2\kappa}=\frac{d\eta+\xi\eta}{[d+\eta]^2}, \quad (4.4.5)$$

where ξ and η are independent $N(0,1)$ variates, $d=y_0\sqrt{-2\kappa}/\sigma=b\sqrt{-2\kappa}/\sigma$.

The limiting distribution in (4.4.4) (or (4.4.5)) is different from that in (4.2.17) (or (4.2.18)) unless $y_0=b=0$ where the two limiting distributions become the Cauchy distribution. In this section we will show that the limiting distributions of $(2c)^{3/2}e^{-2c}\int_0^1\exp\{cr\}J_c(r)\,dr$ and $(2c)^{1/2}e^{-c}\int_0^1\exp\{cr\}\,dW(r)$ are not identical and hence, his sequential limiting distribution is not correct. Instead, the two limiting distributions are independent. The correct sequential limiting distribution turns out to be identical to the simultaneous double asymptotic distribution developed in Section 2.

Let us start the investigation from the joint moment generating function (MGF) of $A(\gamma,c)$ and $B(\gamma,c)$ given by Perron. Firstly, we derive the limiting joint MGF in Theorem 4.4.1, from which we obtain the sequential limiting distribution. Secondly, in Theorem 4.4.2, we give the correct limiting distributions of $(2c)^{3/2}e^{-2c}\int_0^1\exp\{cr\}J_c(r)\,dr$ and $(2c)^{1/2}e^{-c}\int_0^1\exp\{cr\}\,dW(r)$ and show that they are actually independent.

Theorem 4.4.1 *Let $d = b\sqrt{-2\kappa}/\sigma$, $\gamma = b/\left(\sigma\sqrt{N}\right)$. When $N \to +\infty$, $c = -\kappa N \to +\infty$, we have*

(a) The limiting joint MGF of $(2c)e^{-c}A(\gamma,c)$ and $(2c)^2 e^{-2c}B(\gamma,c)$ is

$$\begin{aligned}\lim_{c\to+\infty} M(\widetilde{v},\widetilde{u}) &= \lim_{c\to+\infty} E\left[\exp\left(\widetilde{v}(2c)e^{-c}A(\gamma,c)+\widetilde{u}(2c)^2e^{-2c}B(\gamma,c)\right)\right] \\ &= \frac{1}{\{1-2\widetilde{u}-\widetilde{v}^2\}^{1/2}}\exp\left\{\frac{d^2\left[2\widetilde{u}+\widetilde{v}^2\right]}{2(1-2\widetilde{u}-\widetilde{v}^2)}\right\}.\end{aligned}$$

(b) Letting ξ and η be independent $N(0,1)$ random variables, then

$$\left((2c)e^{-c}A(\gamma,c),(2c)^2e^{-2c}B(\gamma,c)\right) \Longrightarrow \left(\xi\,[d+\eta],[d+\eta]^2\right).$$

(c)

$$\lim_{N\to\infty}\lim_{h\to 0}\frac{e^{-\kappa N}\left(\widehat{a}_h(\kappa)-a_h(\kappa)\right)}{-2\kappa h} = \lim_{c\to\infty}\frac{(2c)e^{-c}A(\gamma,c)}{(2c)^2e^{-2c}B(\gamma,c)} = \frac{\xi\,[d+\eta]}{[d+\eta]^2} = \frac{\xi}{d+\eta},$$

$$\lim_{N\to\infty}\lim_{h\to 0}\frac{e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)}{2\kappa} = \frac{\xi}{d+\eta}.$$

Remark 4.4.1 *The new sequential limiting distribution wherein $h \to 0$ is followed by $N \to \infty$ is the same as the simultaneous double asymptotic distribution derived in Section 2 and the sequential limiting distribution wherein $N \to \infty$ is followed by $h \to 0$ derived in Section 3.*

Remark 4.4.2 *Anderson (1959) proved that, when the error term in the explosive AR(1) model is independent over time and the initial condition is a constant, the limit distribution for the LS estimator should be a ratio of two independent random variables. Our new sequential limiting distributions reported in Theorem 4.4.1 and Theorem 4.3.1, and the double asymptotic distribution are consistent with this result. However, the asymptotic distribution developed in Perron (1991) is at odds with Anderson's result.*

Remark 4.4.3 *It is easy to show that the joint MGF of* $d\eta+\xi\eta$ *and* $[d+\eta]^2$ *is*

$$\frac{1}{\{1-2\widetilde{u}-\widetilde{v}^2\}^{1/2}}\exp\left\{\frac{d^2\left[2\widetilde{u}-2\widetilde{u}\widetilde{v}^2+4\widetilde{u}\widetilde{v}+\widetilde{v}^2\right]}{2\left(1-2\widetilde{u}-\widetilde{v}^2\right)}\right\},$$

which is different from the limiting joint MGF of $(2c)\,e^{-c}A(\gamma,c)$ *and* $(2c)^2\,e^{-2c}B(\gamma,c)$. *This supports the conclusion that the limiting distribution developed in Corollary 1 in Perron (1991) is not correct.*

Theorem 4.4.2 *Let* $J_c(r)=\int_0^r\exp\{c(r-s)\}\,dW(s)$, ξ *and* η *be independent* $N(0,1)$ *variates,* $N\to+\infty$, *and* $c=-\kappa N\to+\infty$. *Then*

(a)

$$(2c)^2\,e^{-2c}\int_0^1 J_c(r)^2\,dr\Longrightarrow\eta^2; \tag{4.4.6}$$

(b)

$$(2c)\,e^{-c}\int_0^1 J_c(r)\,dW(r)\Longrightarrow\xi\eta; \tag{4.4.7}$$

(c)

$$\left\{(2c)^{3/2}\,e^{-2c}\int_0^1\exp\{cr\}\,J_c(r)\,dr\right\}^2\Longrightarrow\eta^2; \tag{4.4.8}$$

(d)

$$(2c)^{1/2}\,e^{-c}\int_0^1\exp\{cr\}\,dW(r)\Longrightarrow\xi. \tag{4.4.9}$$

Comparing results in Theorem 4.4.2 to those in (ii), (iv), (vi) and (viii) of Lemma A.2 in Perron, we notice that we disagree on the limit of $(2c)^{1/2}\,e^{-c}\int_0^1\exp\{cr\}\,dW(r)$. While Perron argued the limit is the same as that of $(2c)^{3/2}\,e^{-2c}\int_0^1\exp\{cr\}\,J_c(r)\,dr$, they should be independent. This difference is the source of the discrepancy in the sequential limits. Interestingly, the sequential limiting distribution holds true even when y_0 is a random variable. This is reported in the following Corollary. Therefore, the simultaneous and sequential double asymptotic distributions are identical to each other no matter what the initial condition is, fixed or random.

Corollary 4.4.3 *Let y_0 be any random variable whose distribution is fixed and independent of sampling interval h or be a constant. Then*

$$\lim_{N\to\infty}\lim_{h\to0}\frac{e^{-\kappa N}\left(\widehat{a}_h(\kappa)-a_h(\kappa)\right)}{-2\kappa h}\overset{d}{=}\lim_{c\to\infty}\frac{(2c)\,e^{-c}A(\gamma,c)}{(2c)^2\,e^{-2c}B(\gamma,c)}\overset{d}{=}\frac{\xi\,[d+\eta]}{[d+\eta]^2}=\frac{\xi}{d+\eta},$$

$$\lim_{N\to\infty}\lim_{h\to0}\frac{e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)}{2\kappa}=\frac{\xi}{d+\eta}.$$

To understand the difference between the two limiting distributions, (4.4.5) and (4.2.18), in Table 1, we report the 0.5%, 1%, 2.5%, 5%, 10%, 90%, 95%, 97.5%, 99%, 99.5% percentiles of the two distributions for five different initial conditions. Several conclusions can be made. First, in all cases, the difference between two distributions are very substantial. For example when $d\sim N(0,1)$, the 99% confidence interval for Perron's distribution is more than 40 times wider than that for the distribution derived in the present paper. Second, the difference in the left tail is more substantial. Third, Perron's distribution is skewed to the left.

To further appreciate the difference between the two limiting distributions, we apply them to a real dataset – the real quarterly U.S. housing market data between 1996:Q1 to 2007:Q1 used in Shiller (2005). The Gaussian OU model is fitted to the logarithmic housing market data with $h=1/4$, $N=11$, $T=44$, and $y_0=4.68$ which is the logarithmic housing price in 1996:Q1. For this initial value, the 0.5% percentile of the two distributions, (4.4.5) and (4.2.18), is -2.8002 and -0.6602, while the 99.5% percentile of the two distributions is 0.2735 and 0.6602, respectively. The estimated value of κ is -0.016. The 99% confidence interval of κ is $(-0.093,-0.009)$ based on Perron's distribution, indicating that there is evidence of explosiveness as it excludes $\kappa=0$. On the other hand, the 99% confidence interval of κ is $(-0.034,0.002)$ based on the correct distribution, indicating that there is no evidence of explosiveness.

d	Limit Theory	0.5%	1%	2.5%	5%	10%	90%	95%	97.5%	99%	99.5%
$N(0,1)$	Correct	-45.0	-22.5	-9.0	-4.5	-2.2	2.2	4.5	9.0	22.5	45.0
	Perron	-3583	-895	-143	-35.4	-8.6	1.0	2.6	8.8	51.9	205
$N(0,4)$	Correct	-28.5	-14.2	-5.7	-2.8	-1.4	1.4	2.8	5.7	14.2	28.5
	Perron	-2280	-570	-91.0	-22.6	-5.5	0.5	1.1	3.3	18.0	70.3
$N(0,1/4)$	Correct	-56.9	-28.5	-11.4	-5.6	-2.8	2.8	5.6	11.4	28.5	56.9
	Perron	-2470	-617	-99.0	-24.9	-6.4	1.7	4.7	16.0	94.3	374
$N(5,1)$	Correct	-0.75	-0.62	-0.47	-0.37	-0.28	0.28	0.37	0.47	0.62	0.75
	Perron	-3.82	-2.36	-1.30	-0.82	-0.49	0.16	0.20	0.23	0.26	0.29
$N(5,0)$	Correct	-0.60	-0.53	-0.43	-0.35	-0.27	0.27	0.35	0.43	0.53	0.60
	Perron	-2.25	-1.66	-1.07	-0.73	-0.46	0.16	0.19	0.22	0.24	0.26

Table 4.1: This table reports various percentiles of the two limiting distributions, (4.4.5) and (4.2.18), for five initial conditions. The last initial condition is simply a constant 5.

4.5 Conclusions

This paper develops the double asymptotic limit theory for the persistent parameter in the explosive continuous time models driven by Lévy processes with a large number of time span (N) and a small number of sampling interval (h). It is shown that the invariance principle applies to the explosive continuous time models under the simultaneous double asymptotics. This finding differs from the well known result for the explosive discrete time model where the limiting distribution is dependent on the error distribution and no invariance principle applies (Anderson, 1959). In the special case of the explosive OU process driven by the Brownian motion but with a known mean, both the simultaneous limits and the two alternative sequential limits have been considered. The three limiting distributions are identical and the expression works for both the fixed and the random initial condition. These results complement the asymptotic theory for stationary continuous time models developed in Tang and Chen (2009). However, the new asymptotic theory is different from the sequential limit theory derived in Perron (1991). We have identified the source of the discrepancy. While Perron argued the limit of $(2c)^{1/2}e^{-c}\int_0^1 \exp\{cr\}\,dW(r)$ is the same as that of $(2c)^{3/2}e^{-2c}\int_0^1 \exp\{cr\}J_c(r)\,dr$, the two limits should actually be independently distributed. The empirical application to the U.S. house market data highlights the difference between the two distributions.

Chapter 5 Summary of Conclusions

The Chapter 2 provides a framework for studying the implications of different discretization schemes in estimating the mean reversion parameter in both multivariate and univariate diffusion models with a linear drift function. The approach includes the Euler method and the trapezoidal method as special cases, an asymptotic theory is developed, and finite sample bias comparisons are conducted using analytic approximations. Bias is decomposed into a discretization bias and an estimation bias. It is shown that the discretization bias is of order $O(h)$ for the Euler method and $O(h^2)$ for the trapezoidal method, respectively, whereas the estimation bias is of the order of $O(T^{-1})$. Since in practical applications in finance it is very likely that h is much smaller than $1/T$, estimation bias is likely to dominate discretization bias.

Applying the multivariate theory to univariate models gives several new results. First, it is shown that in the Euler and trapezoidal methods, the sign of the discretization bias is opposite that of the estimation bias for practically realistic cases. Consequently, the bias in the two approximate method is smaller than the ML estimator based on the exact discrete time model. Second, although the trapezoidal method leads to a smaller discretization bias than the Euler method, the estimation bias is bigger. As a result, it is not clear if there is a gain in reducing the total bias by using a higher order approximation. When comparing the estimator based on the Euler method and the exact ML, we find that the asymptotic variance of the former estimator is smaller. As a result, there is clear evidence for preferring the estimator based on the Euler method to the exact ML in the univariate linear diffusion.

Simulations suggest the bias continues to be large in finite samples. It is also confirmed that for empirically relevant cases, the magnitude of the discretization bias in the two approximate methods is much smaller than that of the estimation bias. The two approximate methods lead to a smaller variance than exact ML. Most importantly for practical work, there is strong evidence that the bias formulae work well and so they can be recommended for analytical bias correction with these models.

For the univariate square root model, the Euler method is found to have smaller bias and smaller variance than the Nowman method. Discretizing the diffusion function both in the Euler method and the Nowman method causes no discretization bias on the mean reversion paramter. For the Euler method, we have derived an explicit expression for the discretization bias caused by discretizing the drift function. The simulation results suggest that the Euler method performs best in terms of both bias and variance.

The analytic and expansion results given in the paper are obtained for stationary systems. Bias analysis for nonstationary and explosive cases require different methods. For diffusion models with constant diffusion functions, it may be possible to extend recent finite sample and asymptotic expansion results for the discrete time AR(1) model (Phillips, 2010) to a continuous time setting. Such an analysis would involve a substantial extension of the present work and deserves treatment in a separate study.

The Chapter 3 derives the asymptotic distribution of the ML/LS estimator of the mean reversion matrix in a multivariate diffusion model with a linear drift and a constant diffusion when only discretely sampled data are available. Both the stationary case and the non-stationary case are examined. The limit theory gives an analytic expression of the asymptotic covariance matrix, for which a consistent estimator is provided thereby facilitating inference about the mean reversion matrix. Our method replies on the asymptotic theory of the ML/LS estimator of the exact discrete time VAR model.

The transformation from the continuous time model to the exact discrete system involves a nonlinear matrix logarithmic mapping. The mean reversion matrix is shown to be identified under a weak condition. When identification is achieved, our method also utilizes a novel explicit relationship between the AR coefficient matrix and the mean reversion matrix. This relationship is a polynomial of an finite order, facilitating the use of the delta method and the calculation of the covariance matrix in the limit distribution.

Both in the stationary case and in the unit root case, we develop the limit theory of the ML/LS estimator of the mean reversion matrix by using the limit distribution of the estimated AR coefficient matrix only. The expression of the asymptotic covariance matrix in stationary case is a little complicated. Different situations have been discussed to get an explicit representation of the asymptotic covariance matrix. For models of low dimension, such as $m \leq 3$, using our framework, the mean reversion matrix is shown to have a straightforward expression as a continuously differentiable mapping of the AR coefficient matrix.

The new theory is illustrated in an empirical application to a multivariate OU model for the logarithmic daily realized volatility (RV) of Pound, Euro and Yen exchange rates. Using our method, we are able to obtain the estimate of the asymptotic covariance of the mean reversion matrix. The statistical inferences, conducted on these covariances, suggest that the three series are stationary and revert to their means in fast rates, and that the RV of the Pound does not depend on the lagged RV of the Yen and the RV of the Yen does not depend on the lagged RV of the Yen.

Although in chapter 3 we only develop the asymptotic theory for multivariate diffusion models with a linear drift and a constant diffusion, our method is generally applicable to continuous time models with a linear drift but with a more flexible diffusion function and to continuous time models which are driven by Lévy process. In this case, OLS may be applied to estimate the AR coefficient matrix of the exact discrete time system. As long as the asymptotic theory of the OLS estimator is available, our method can be applied in the same manner.

The Chapter 4 develops the double asymptotic limit theory for the persistent parameter in the explosive continuous time models driven by Lévy processes with a large number of time span (N) and a small number of sampling interval (h). It is shown that the invariance principle applies to the explosive continuous time models under the simultaneous double asymptotics. This finding differs from the well known result for the explosive discrete time model where the limiting distribution is dependent on the error distribution and no invariance principle applies (Anderson, 1959). In the special case of the explosive OU process driven by the Brownian motion but with a known mean, both the simultaneous limits and the two alternative sequential limits have been considered. The three limiting distributions are identical and the expression works for both the fixed and the random initial condition. These results complement the asymptotic theory for stationary continuous time models developed in Tang and Chen (2009). However, the new asymptotic theory is different from the sequential limit theory derived in Perron (1991). We have identified the source of the discrepancy. While Perron argued the limit of $(2c)^{1/2} e^{-c} \int_0^1 \exp\{cr\} dW(r)$ is the same as that of $(2c)^{3/2} e^{-2c} \int_0^1 \exp\{cr\} J_c(r) dr$, the two limits should actually be independently distributed. The empirical application to the U.S. house market data highlights the difference between the two distributions.

References

Aït-Sahalia, Y., 1999, Transition Densities for Interest Rate and Other Non-linear Diffusions. *Journal of Finance*, 54, 1361–1395.

Aït-Sahalia, Y., 2002, Maximum Likelihood Estimation of Discretely Sampled Diffusion: A Closed-form Approximation Approach. *Econometrica*, 70, 223–262.

Aït-Sahalia, Y., 2008, Closed-Form Likelihood Expansions for Multivariate Diffusions. *Annals of Statistics*, 36, 906-937.

Aït-Sahalia, Y. and J. Jacod, 2011, Testing Whether Jumps Have Finite or Infinite Activity. *Annals of Statistics*, 39, 1689-1719.

Aït-Sahalia, Y., and Kimmel R., 2007, Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Financial Economics*, 83, 413–452.

Aït-Sahalia, Y., and Yu, J., 2006, Saddlepoint Approximations for Continuous-Time Markov Processes. *Journal of Econometrics*, 134, 507-551.

Anderson, T. W., 1959, On asymptotic distribution of estimates of parameters of stochastic difference equations. *Ann. Math. Statist*, 30, 676–687.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001, The distribution of realized stock return volatility.*Journal of Financial Economics*, 61, 43–76.

Aue, A. and L. Horvath, 2007, A limit theorem for mildly explosive autoregression with stable errors. *Econometric Theory*, 23, 201-220.

Bao, Y. and Ullah, A., 2009, On Skewness and Kurtosis of Econometric Estimators, *Econometrics Journal*, 12, 232-247.

Barndorff-Nielsen, O. E., 1998, Processes of Normal Inverse Gaussian Type. *Finance and Stochastics*, 2, 41–68.

Bergstrom, A. R., 1966, Nonrecursive Models as Discrete Approximations to Systems of Stochastic Differential Equations. *Econometrica*, 34, 173–182.

Bergstrom, A. R., 1984, Continuous Time Stochastic Models and Issues of Aggregation over Time. *Handbook of Econometrics*, 1146-1211

Bergstrom, A. R., 1990, *Continuous Time Econometric Modelling*. Oxford University Press, Oxford.

Bergstrom, A. R., and C. R. Wymer, 1976, A Model of Disequilibrium Neoclassical Growth and its Application to the United Kingdom, in *Statistical Inference in Continuous Time Economic Models*, ed. by A. R. Bergstrom. Amsterdam: North-Holland, 1976.

Black, F., and Scholes, M., 1973, The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81, 654-673.

Cardoso, J.R., 2005, An Explicit Formula for the Matrix Logarithm. *South African Optometrist*, 64, 80-83.

Carr, P. and L. Wu, 2003, The Finite Moment Log Stable Process and Option Pricing. *Journal of Finance*, 58, 753–778.

Chan, K., Karolyi, F, Longstaff, F., and A. Sanders, 1992, An empirical comparison of alternative models of short term interest rates. *Journal of Finance* 47, 1209–1227.

Cox, J., Ingersoll, J., and Ross, S., 1985, A Theory of the Term Structure of Interest Rates. *Econometrica*, 53, 385–407.

Durham, G., and Gallant, A. R., 2002, Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes. *Journal of Business and Economic Statistics*, 20, 297–316.

Duffie, D., and Kan, R., 1996, A Yield-factor Model of Interest Rate. *Mathematical Finance*, 6, 379-406.

Duffie, D., J. Pan, and K. J. Singleton, 2000, Transform Analysis and Asset Pricing for Affine Jump-diffusions. *Econometrica*, 68, 1343-1376.

Elerian, O. (1998), A Note on the Existence of a Closed-form Conditional Transition Density for the Milstein Scheme, Economics discussion paper 1998-W18, Nuffield College, Oxford.

Eraker, B., 2001, MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business and Economic Statistics*, 19, 177-191

Florens-Zmirou, D., 1989, Approximate Discrete-time Schemes for Statistics of Diffusion Processes. *Statistics*, 20, 547-557.

Feller, W., 1951, Two Singular Diffusion Problems. *Annals of Mathematics*, 54, 173-182.

Fuller, W. A., 1976, *Introduction to Statistical Time Series*. 20, New York: Wiley.

Gallant, A. R., and G. Tauchen, 1996, Which Moments to Match? *Econometric Theory*, 12, 657–681.

Hall, P. and C. C. Heyde, 1980, *Martingale Limit Theory and Its Application*. Academic Press.

Hannan, E. J., 1970, *Multiple Time Series*. 20, New York: Wiley.

Hansen, L. P., and Sargent, T. J., 1983, The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities. *Econometrica*, 51, 377-388.

Jeong, M., and J.Y. Park, 2009, Asymptotics for the Maximum Likelihood Estimators of Diffusion Models. Working Paper, Texas A&M University.

Kallenberg, O., 2002, Foundations of Modern Probability. Springer, Berlin.

Knight, M. D., and C. R. Wymer, 1978, A Macroeconomic Model of the United Kingdom. *IMF Staff Papers*, 25, 742-778.

Knight, J. L., and J. Yu, 2002, The Empirical Characteristic Function in Time Series Estimation. *Econometric Theory*, 18, 691-721.

Lo, A. W., 1988, Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data. *Econometric Theory*, 4, 231-247.

Madan, D., Carr, P., and E. Chang, 1999, The Variance Gamma Processes and Option Pricing. *European Finance Review,* 2, 79–105.

Magdalinos, T., 2011, Mildly Explosive Autoregression under Weak and Strong Dependence. *Journal of Econometrics*, forthcoming.

Magnus, J. R., and H. Neudecker, 2007, Matrix Differential Calculus with Applications in Statistics and Econometrics. New York: John Wiley.

Mann, H.B., and A. Wald, 1943, On the Statistical Treatment of Linear Stochastic Difference Equations. *Econometrica*, 11, 173-220.

Merton, R., 1976, Option pricing when underlying stock returns are discontinuous. *Journal Financial Economics*, 3, 125–144.

Milstein, G. N., 1978, A Method of Second-Order Accuracy Integration of Stochastic Differential Equations. *Theory of Probability and its Applications*, 23, 396-401.

Nicholls, D. F., and Pope, A. L., 1988, Bias in the Estimation of Multivariate Autoregressions. *Australian Journal of Statistics*, 30, 296-309.

Nowman, K. B., 1997, Gaussian Estimation of Single-factor Continuous Time Models of the Term Structure of Interest Rates. *Journal of Finance*, 52, 1695-1703.

Park, J. Y., and P.C.B. Phillips, 1988, Statistical Inference in Regressions with Integrated Processes: Part 1. *Econometric Theory*, 4, 468-497.

Park, J. Y., and P.C.B. Phillips, 1989, Statistical Inference in Regressions with Integrated Processes: Part 2. *Econometric Theory*, 5, 95-131.

Pedersen, A., 1995, A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observation. *Scandinavian Journal of Statistics*, 22, 55-71.

Perron, P., 1991, A continuous time approximation to the unstable first order autoregressive processes: The case without an intercept. *Econometrica*, 59, 211–236.

Phillips, P. C. B., 1972, The Structural Estimation of Stochastic Differential Equation Systems. *Econometrica*, 40, 1021-1041.

Phillips, P. C. B., 1973, The Problem of Identification in Finite Parameter Continuous Time Models. *Journal of Econometrics*, 1, 351-362.

Phillips, P. C. B., 1982, A Simple Proof of the Latent Root Sensitivity. *Economics Letters*, 9, 57-59.

Phillips, P. C. B., 2010, Folklore theorems, implicit maps, and new unit root limit theory, Working paper, Yale University.

Phillips, P. C. B., and S. Durlauf, 1986, Multiple Time Series Regression with Integrated Processes. *Review of Economic Studies*, 53, 473-495.

Phillips, P.C.B., and T. Magdalinos, 2007, Limit theory for moderate deviation from unity. *Journal of Econometrics*, 136, 115–130.

Phillips, P.C.B., Wu, Y. and J. Yu, 2011, Explosive Behavior in the 1990s Nasdaq: When did Exuberance Escalate Asset Values? *International Economic Review*, 52, 201-226.

Phillips, P. C. B., and Yu, J., 2005a, Jackknifing Bond Option Prices. *Review of Financial Studies*, 18, 707-742.

Phillips, P. C. B., and Yu, J., 2005b, Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics. *Statistical Science*, 20, 338-343.

Phillips, P. C. B., and Yu, J., 2009a, Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Handbook of Financial Time Series*, 707-742.

Phillips, P. C. B., and Yu, J., 2009b, Simulation-based Estimation of Contingent-claims Prices. *Review of Financial Studies*, 22, 3669-3705.

Phillips, P.C.B., and J. Yu, 2011, Dating the Timeline of Financial Bubbles During the Subprime Crisis. *Quantitative Economics*, 2, 455-491.

Piazzesi, M., 2009, Affine Term Structure Models, in *Handbook for Financial Econometrics*, edt by Aït-Sahalia, Y. and L. Hansen, North-Holland.

Sargan, J. D., 1974, Some Discrete Approximations to Continuous Time Stochastic Models. *Journal of the Royal Statistical Society, Series B*, 36, 74-90.

Sato, K., 1999, Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press.

Shiller, R. J., 2005, *Irrational Exuberance*, Princeton University Press, 2nd edition.

Singleton, K. J., 2001, Estimation of Affine Asset Pricing Models using the Empirical Characteristic Function. *Journal of Econometrics*, 102, 111–141.

Sundaresan, S.M., 2000, Continuous-Time Models in Finance: A review and an Assessment. *Journal of Finance*, 55, 1569-1622.

Tang, C. Y., and Chen, S. X., 2009, Parameter Estimation and Bias correction for Diffusion processes. *Journal of Econometrics*, 149, 65-81.

Vasicek, O., 1977, An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5, 177–186.

Ullah, A., Wang, Y., and Yu, J., 2010, Bias in the Mean Reversion Estimator in the Continuous Time Gaussian and Levy Processes. Working Paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University.

White, J., 1958, The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *Ann. Math. Statist*, 29, 1188-1197.

Yamamoto, T. and Kunitomo, N., 1984, Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models. *Annals of the Institute of Statistical Mathematics*, 36, 419-430.

Yu, J., 2009, Bias in the Estimation of Mean Reversion Parameter in Continuous Time Models. Working Paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University.

Zhou, Q. and J. Yu, 2011, Asymptotic Distributions of the Least Squares Estimator for Diffusion Processes. Working Paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University.

Appendix

A Proofs in Chapter 2

Proof of Lemma 2.3.3. Let $C = F^{'}$ and then

$$\sum_{t=0}^{\infty} F^{'k} = (I - F^{'})^{-1} = (1 - C),$$

$$\sum_{k=0}^{\infty} F^{'k} tr(F^{k+1}) = \sum_{k=0}^{\infty} F^{'k} \sum_{\lambda \in spec(F)} \lambda^{k+1} = \sum_{\lambda \in spec(F)} [\lambda \sum_{k=0}^{\infty} \lambda^k F^{'k}]$$
$$= \sum_{\lambda \in spec(C)} [\lambda \sum_{k=0}^{\infty} \lambda^k C^k] = \sum_{\lambda \in spec(C)} \lambda (I - \lambda C)^{-1},$$

where $Spec(C)$ denotes the set of eigenvalues of C. Thus,

$$\sum_{k=0}^{\infty} F^{'2k+1} = \sum_{k=0}^{\infty} C^{2k+1} = C(I - C^2)^{-1},$$

$$\Gamma(0) = Var(x_t) = \sum_{i=0}^{\infty} F^i \cdot G \cdot F^{'i} = D,$$

$$B_n = BIAS(\hat{F}) = E(\hat{F}) - F = -\frac{b}{n} + O(n^{-\frac{3}{2}}).$$

■

Proof of Lemma 2.3.1. By Lemma 2.3.1, for fixed h, as $n \to \infty$, $\hat{F} \xrightarrow{p} F$. Hence,

$$\hat{A} - A = \frac{1}{h}[\hat{F} - F] + \frac{1}{h} H \xrightarrow{p} \frac{1}{h} H.$$

From Equations (2.3.8), $\frac{1}{h}H = \frac{1}{h}[F - I - Ah] = O(h)$ as $h \to 0$, proving the first part.

(b) According to Lemma 2.3.1, fixed h, as $n \to \infty$,

$$\sqrt{n}\{Vec(\hat{F}) - Vec(F)\} \xrightarrow{d} N(0, (\Gamma(0))^{-1} \otimes G),$$

$$\begin{aligned}\sqrt{n}hVec[\hat{A} - \frac{1}{h}(F - I)] &= \sqrt{n}Vec[\hat{A}h - (F - I)] \\ &= \sqrt{n}Vec[\hat{F} - F] \xrightarrow{d} N(0, (\Gamma(0))^{-1} \otimes G),\end{aligned}$$

giving the second part. ■

Proof of Lemma 2.3.2. According to formulae (2.3.8), (2.3.9) and Lemma 2.3.3,

$$\begin{aligned}E(\hat{A} - A) &= \frac{1}{h}E(\hat{F} - F) + \frac{1}{h}H = \frac{1}{h}E(\frac{-b}{n} + O(n^{-3/2})) + \frac{1}{h}H \\ &= -\frac{b}{T} + \frac{1}{h}H + o(T^{-1}).\end{aligned}$$

■

Proof of Lemma 2.3.3. (a) From formulae (2.3.17),

$$\begin{aligned}\hat{A} - A &= \frac{2}{h}(\hat{F} - I)(\hat{F} + I)^{-1} - \frac{2}{h}(F - I)(F + I)^{-1} - \nu \\ &= \frac{2}{h}(\hat{F} + I - 2I)(\hat{F} + I)^{-1} - \frac{2}{h}(F - I)(F + I)^{-1} - \nu \\ &= \frac{2}{h}[I - 2(\hat{F} + I)^{-1}] - \frac{2}{h}[I - 2(F + I)^{-1}] - \nu \\ &= -\frac{4}{h}[(\hat{F} + I)^{-1} - (F + I)^{-1}] - \nu \\ &= \frac{4}{h}(I + F)^{-1}(\hat{F} - F)(I + \hat{F})^{-1} - \nu.\end{aligned}$$

As h is fixed, according Lemma 2.3.1, as $n \to \infty$, $\hat{F} \xrightarrow{p} F$, the first part of above equation goes to zero. And from formula (2.3.17),

$$\hat{A} - A \xrightarrow{p} -\nu = \frac{2}{h}(F - I)(F + I)^{-1} - A.$$

(b)

$$Vec(\hat{A}-A+\nu) = Vec[\hat{A}-\frac{2}{h}(F-I)(F+I)^{-1}] = \frac{4}{h}Vec[(I+F)^{-1}(\hat{F}-F)(I+\hat{F})^{-1}]$$
$$= \frac{4}{h}\{(\hat{F}'+I)^{-1}\otimes(F+I)^{-1}\}Vec(\hat{F}-F).$$

Again when h is fixed, according to Lemma 2.3.1, as $n\to\infty$, $\sqrt{n}(\hat{F}-F)\xrightarrow{d}N(0,\Gamma(0)^{-1}\otimes G)$, and we get

$$h\sqrt{n}Vec[\hat{A}-\frac{2}{h}(F-I)(F+I)^{-1}]\xrightarrow{d}N(0,\Psi),$$

where

$$\Psi = 16\Upsilon[\Gamma(0)^{-1}\otimes G]\Upsilon', \quad \Upsilon = (F'+I)^{-1}\otimes(F+I)^{-1}$$

■

Proof of Lemma 2.3.4. From the proof of theorem 2.3.3, we have

$$E[\hat{A}]-A = -\frac{4}{h}E[(\hat{F}+I)^{-1}-(F+I)^{-1}]-\nu$$
$$= -\frac{4}{h}E[(\hat{F}+I)^{-1}]+\frac{4}{h}(F+I)^{-1}-\nu.$$

For the first term, we note that

$$(\hat{F}+I)^{-1} = (I+F+\hat{F}-F)^{-1} = [(I+F)(I+(I+F)^{-1}(\hat{F}-F))]^{-1}$$
$$= [I+(I+F)^{-1}(\hat{F}-F)]^{-1}(I+F)^{-1},$$

and

$$[I+(I+F)^{-1}(\hat{F}-F)]^{-1} = \sum_{i=0}^{\infty}(-1)^i[(I+F)^{-1}(\hat{F}-F)]^i$$
$$= I-(I+F)^{-1}(\hat{F}-F)+[(I+F)^{-1}(\hat{F}-F)]^2$$
$$+\sum_{i=3}^{\infty}(-1)^i[(I+F)^{-1}(\hat{F}-F)]^i.$$

By Lemma 2.3.1, we have

$$\sqrt{n}[Vec(\hat{F}) - Vec(F)] \xrightarrow{d} N(0, \Gamma(0)^{-1} \otimes G),$$

and so,

$$\hat{F}_{ij} - F_{ij} = O_P(n^{-\frac{1}{2}}).$$

Then,

$$[(I+F)^{-1}(\hat{F}-F)]^3 = O_p(n^{-\frac{3}{2}}) \text{ and } [(I+F)^{-1}(\hat{F}-F)]^i = o_p(n^{-\frac{3}{2}}),\ i \geq 3,$$

$$[I+(I+F)^{-1}(\hat{F}-F)]^{-1} = I-(I+F)^{-1}(\hat{F}-F)+[(I+F)^{-1}(\hat{F}-F)]^2+O_p((n^{-\frac{3}{2}})),$$

and

$$\begin{aligned}
E[\hat{A}-A] &= -\frac{4}{h}E\{[I+(I+F)^{-1}(\hat{F}-F)^{-1}]\}(I+F)^{-1} + \frac{4}{h}(F+I)^{-1} + O(h^2) \\
&= \frac{4}{h}E\{(I+F)^{-1}(\hat{F}-F)(I+F)^{-1}\} - \frac{4}{h}E\{[(I+F)^{-1}(\hat{F}-F)]^2(I+F)^{-1}\} \\
&+ \frac{1}{h}O(n^{-\frac{3}{2}}) - \nu.
\end{aligned}$$

Now let $\hat{g} = [(I+F)^{-1}(\hat{F}-F)]$, so that

$$\sqrt{n}\cdot Vec[\hat{g}] = \sqrt{n}\cdot Vec[(I+F)^{-1}(\hat{F}-F)] = [I_M \otimes (I+F)^{-1}]\sqrt{n}Vec(\hat{F}-F) \xrightarrow{d} N(0,\Delta),$$

where $\Delta = [I_M \otimes (I+F)^{-1}] \cdot \Gamma(0)^{-1} \otimes G \cdot [I_M \otimes (I+F)^{-1}]'$. As a result,

$$Var(\sqrt{n}\cdot Vec(\hat{g})) = \Delta + o(1) \rightarrow Var[Vec(\hat{g})] = \frac{\Delta}{n} + o(n^{-1}),$$

and

$$\begin{aligned}
E[Vec(\hat{g})\cdot Vec(\hat{g})^T] &= Var[Vec(\hat{g})] + E[Vec(\hat{g})]\cdot E[Vec(\hat{g})]^T \\
&= \frac{\Delta}{n} + E[Vec(\hat{g})]\cdot E[Vec(\hat{g})]^T + o(n^{-1}).
\end{aligned}$$

From Lemma 2.3.3,

$$B_n = E(\hat{F}) - F = -\frac{b}{n} + O(n^{-\frac{3}{2}}).$$

When the exact discrete model involves an unknown $B(\theta)$ we have

$$b = G[(I-C)^{-1} + C(I-C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1},$$

and when we have a prior knowledge that $B(\theta) = 0$ in (2.2.2), we have

$$b = G[C(I-C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1}.$$

Then

$$\begin{aligned}
E[Vec(\hat{g})] &= E[(I_M \otimes (I+F)^{-1})Vec(\hat{F}-F)] \\
&= [I_M \otimes (I+F)^{-1}]E[Vec(\hat{F}-F)] \\
&= [I_M \otimes (I+F)^{-1}]Vec[E(\hat{F}-F)] \\
&= [I_M \otimes (I+F)^{-1}]Vec[-\frac{b}{n} + O(n^{-\frac{3}{2}})] = O(n^{-1}) \\
&\rightarrow E[Vec(\hat{g})Vec(\hat{g})^T] = \frac{\Delta}{n} + o(n^{-1}).
\end{aligned}$$

Here we assume $\hat{W} = [(I+F)^{-1}(\hat{F}-F)]^2 = \hat{g}\hat{g}$ and $\hat{W}_{ij} = \sum_{s=1}^{M} \hat{g}_{is}\hat{g}_{sj}$. It is easy to find that $\hat{g}_{is}$ is the $(M(s-1)+i)$th element of $Vec(\hat{g})$, and $\hat{g}_{is}\hat{g}_{sj}$ is the $(M(s-1)+i, M(j-1)+s)^{th}$ element of $Vec(\hat{g})Vec(\hat{g})$. Defining e_i to be the column vector of dimension M^2 whose i^{th} element is 1 and other elements are 0, we have

$$\begin{aligned}
E[\hat{g}_{is}\hat{g}_{sj}] &= e'_{M(s-1)+i}E[Vec(\hat{g})Vec(\hat{g})]'] e_{M(j-1)+s} \\
&= \frac{1}{n} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s} + o(n^{-1}),
\end{aligned}$$

$$
\begin{aligned}
E[\hat{W}_{ij}] &= \sum_{s=1}^{M} E[\hat{g}_{is}\hat{g}_{sj}] \\
&= \sum_{s=1}^{M} \frac{1}{n} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s} + o(n^{-1}).
\end{aligned}
$$

Next, define the matrix P with (i, j) element

$$
P_{ij} = \frac{1}{n} \sum_{s=1}^{M} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s}.
$$

Then

$$
E\{[(I+F)^{-1}(\hat{F}-F)]^2\} = E(\hat{W}) = P + o(n^{-1}).
$$

Again, using Lemma 2.3.3, the formula for the estimation bias is

$$
\begin{aligned}
E[\hat{A}-A] &= \frac{4}{h}E\{(I+F)^{-1}(\hat{F}-F)(I+F)^{-1}\} - \frac{4}{h}E\{[(I+F)^{-1}(\hat{F}-F)^2](I+F)^{-1}\} \\
&+ \frac{1}{h}O(n^{-\frac{3}{2}}) - \nu \\
&= \frac{4}{h}(I+F)^{-1}[-\frac{b}{n} + O(n^{-\frac{3}{2}})](I+F)^{-1} \\
&- \frac{4}{h} \cdot W \cdot (I+F)^{-1} + \frac{1}{h}o(n^{-1}) + \frac{1}{h}O(n^{-\frac{3}{2}}) - \nu \\
&= -\frac{4}{T}(I+F)^{-1} \cdot b \cdot (I+F)^{-1} - \frac{4}{h} \cdot W \cdot (I+F)^{-1} - \nu + o(T^{-1}).
\end{aligned}
$$

■

Proof of Lemma 2.4.1. Using (2.4.4) and (2.4.5) in (2.4.3), we have

$$\begin{aligned}
\sum_{t=1}^{n} \frac{1}{h}(X_t - X_{t-1})V_t' &= \frac{1}{2h}\sum_{t=1}^{n} X_t X_{t-1}' - \frac{1}{2h}\sum_{t=1}^{n} X_{t-1}X_{t-1}' \\
&+ \frac{1}{2h}\sum_{t=1}^{n} X_t X_{t-1}' \left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1} \sum_{t=1}^{n} X_{t-1}X_t' - \frac{1}{2h}\sum_{t=1}^{n} X_{t-1}X_t' \\
&= \frac{1}{2h}[\left(\sum_{t=1}^{n} X_t X_{t-1}'\right)\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1} - I \\
&+ \left(\sum_{t=1}^{n} X_t X_{t-1}'\right)\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n} X_{t-1}X_t'\right)\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1} \\
&- \left(\sum_{t=1}^{n} X_{t-1}X_t'\right)\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1}]\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right) \\
&= \frac{1}{2h}[\hat{F} - I + \hat{F}\left(\sum_{t=1}^{n} X_{t-1}X_t'\right)\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1} \\
&- \left(\sum_{t=1}^{n} X_{t-1}X_t'\right)\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1}]\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right) \\
&= \frac{1}{2h}(\hat{F} - I)\left[I + \sum_{t=1}^{n} X_{t-1}X_t'\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)^{-1}\right]\sum_{t=1}^{n} X_{t-1}X_{t-1}' \\
&= \frac{1}{2h}(\hat{F} - I)\left[\sum_{t=1}^{n} X_{t-1}X_{t-1}' + \sum_{t=1}^{n} X_{t-1}X_t'\right] \\
&= \frac{1}{2h}(\hat{F} - I)\left(\sum_{t=1}^{n} X_{t-1}X_{t-1}'\right)(\hat{F}' + I).
\end{aligned}$$

By the same method, it is easy to obtain

$$\left[\sum_{t=1}^{n} \frac{1}{2}(X_t + X_{t-1})V_t'\right]^{-1} = \left[\frac{1}{4}(\hat{F} + I)(\sum_{t=1}^{n} X_{t-1}X_{t-1}')(\hat{F}' + I)\right]^{-1}$$

Using the above two formulae in (2.4.3), the two stage least squares estimator is

$$\hat{A} = \frac{2}{h}(\hat{F} - I)(\hat{F} + I)^{-1}.$$

■

Proof of Lemma 2.5.1. The Nowman approximate discrete time model yields the following transition function

$$f(X_iX_{(i-1)}) = \frac{[(1-e^{-2\kappa h})/2\kappa]^{-1/2}}{\sqrt{2\pi}\sigma g(X_{i-1};\psi)} \exp\left\{-\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]^2}{2\sigma^2g^2(X_{i-1};\psi)(1-e^{-2\kappa h})/2\kappa}\right\},$$

and the following log-likelihood function

$$\begin{aligned}\ell(\theta) = &-\frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n}\ln[g(X_{i-1};\psi)] - \frac{n}{2}\ln(\frac{1-e^{-2\kappa h}}{2\kappa}) \\ &-\sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]^2}{2\sigma^2g^2(X_{i-1};\psi)(1-e^{-2\kappa h})/2\kappa}.\end{aligned}$$

The first order conditions are

$$\frac{\partial\ell(\theta)}{\partial\mu} = 0 \;\Rightarrow\; \sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]}{g^2(X_{i-1};\psi)} = 0, \quad (.0.1)$$

$$\frac{\partial\ell(\theta)}{\partial\sigma^2} = 0 \;\Rightarrow\; \sigma^2\left(\frac{1-e^{-2\kappa h}}{2\kappa}\right) - \frac{1}{n}\sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)} = 0, \quad (.0.2)$$

$$\begin{aligned}\frac{\partial\ell(\theta)}{\partial\kappa} = 0 \;\Rightarrow\; 0 = &-\frac{n}{2}\left[\frac{2he^{-2\kappa h}}{1-e^{-2\kappa h}} - \frac{1}{\kappa}\right] \\ &- he^{-\kappa h}\sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu](X_{i-1}-\mu)}{\sigma^2g^2(X_{i-1};\psi)(1-e^{-2\kappa h})/2\kappa} \\ &- \sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]^2}{2\sigma^2g^2(X_{i-1};\psi)}\left[\frac{2(1-e^{-2\kappa h})-4\kappa he^{-2\kappa h}}{(1-e^{-2\kappa h})^2}\right]. \quad (.0.3)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\ell(\theta)}{\partial\psi_j} = 0 \;\Rightarrow\; 0 = &\,\sigma^2\frac{1-e^{-2\kappa h}}{2\kappa}\sum_{i=1}^{n}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)} \\ &- \sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)} \quad (.0.4)\end{aligned}$$

Taking Equation (.0.2) into (.0.3), the first term and the third term cancel and we obtain

$$\sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu](X_{i-1}-\mu)}{g^2(X_{i-1};\psi)} = 0. \quad (.0.5)$$

Taking Equation (.0.2) into (.0.4), we have

$$
\begin{aligned}
0 = & \frac{1}{n}\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)}\sum_{i=1}^{n}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)} \\
& -\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)}.
\end{aligned}
\tag{.0.6}
$$

Equations (.0.1), (.0.5) and (.0.6) yield the ML estimators, $\hat{\phi}_1$, $\hat{\mu}$ and $\hat{\psi}$ and Equation (.0.2) gives the ML estimator, $\hat{\sigma}^2$.

The Euler approximate discrete model yields the following log-likelihood function,

$$
\ell(\theta) = -\frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n}\ln[g(X_{i-1};\psi)] - \sum_{i=1}^{n}\frac{[X_i-\phi_2 X_{i-1}-(1-\phi_2)\mu]^2}{2\sigma^2 h g^2(X_{i-1};\psi)}.
$$

It is easy to obtain the first order conditions, three of which are identical to those in (.0.1), (.0.5) and (.0.6). Hence,

$$
\hat{\phi}_2 = \hat{\phi}_1.
$$

■

B Proofs in Chapter 3

Proof of Proposition 3.2.1.

The eigenvalues of $F = e^{Ah}$ are $e^{\eta_j h}, j = 1, \cdots m$, where $\{\eta_j\}_{j=1}^m$ are the eigenvalues of A. For any $j = 1, \cdots m$, if η_j is a real number, $e^{\eta_j h} > 0$. If η_j is a complex number, say $\eta_j = \alpha + i\beta$, then

$$e^{\eta_j h} = e^{(\alpha+i\beta)h} = e^{\alpha h}\left[\cos(\beta h) + i\sin(\beta h)\right].$$

Since $\beta h \in (-\pi, \pi)$ by Assumption 1, $e^{\eta_j h}$ cannot be a negative real number or zero for any $j = 1, \cdots, m$. Hence,

$$spec\{F\} \cap \mathbb{R}_0^- = \varnothing.$$

■

Proof of Theorem 3.3.1.

Based on the formulae (3.2.8) and (3.2.12), straightforward calculation allow us to show

$$\begin{aligned} h\left(\hat{A} - A\right) &= \left(\hat{f}_1 - f_1\right) I + \sum_{j=2}^m \left(\hat{f}_j - f_j\right)\left(I - \hat{F}\right)^{j-1} + \sum_{j=2}^m f_j \left\{\left(I - \hat{F}\right)^{j-1} - (I - F)^{j-1}\right\} \\ &= \sum_{j=1}^m \left(\hat{f}_j - f_j\right)\left(I - \hat{F}\right)^{j-1} - \sum_{j=2}^m f_j \left\{\sum_{s=0}^{j-2} (I - F)^s \left(\hat{F} - F\right)\left(I - \hat{F}\right)^{j-2-s}\right\}, \end{aligned}$$

which leads to

$$\begin{aligned} hVec\left(\hat{A} - A\right) = &\sum_{j=1}^m \left\{Vec\left[\left(I - \hat{F}\right)^{j-1}\right]\left(\hat{f}_j - f_j\right)\right\} \\ &- \sum_{j=2}^m \sum_{s=0}^{j-2} f_j \left\{(I - F)^s \otimes \left[\left(I - \hat{F}\right)'\right]^{j-2-s}\right\} Vec\left(\hat{F} - F\right). \end{aligned}$$

Notice the fact that $\{f_j\}_{j=1}^m$ defined as in (3.2.9), (3.2.10) and (3.2.11), are differentiable functions on $\{C_j\}_{j=1}^m$, and $\{C_j\}_{j=1}^m$ are continuous functions of the elements of F. Let $F_j = \begin{bmatrix} \frac{\partial f_j}{\partial C_m} & \frac{\partial f_j}{\partial C_{m-1}} & \cdots & \frac{\partial f_j}{\partial C_1} \end{bmatrix}'$, for $j = 1,2,\cdots,m$. We could get that

$$\hat{f}_j - f_j = F'_j \begin{bmatrix} \hat{C}_M - C_M \\ \vdots \\ \hat{C}_1 - C_1 \end{bmatrix} + o_p\left(Vec\left(\hat{F} - F\right)\right), \quad \text{for } j = 1,2,\cdots,m.$$

Let $|\cdot|$ denotes the determinant of matrix, $\hat{\psi}_z = zI - \left(I - \hat{F}\right)$, $\psi_z = zI - (I - F)$ are matrix polynomials with $z \in R$. Then, we have

$$\begin{aligned}
&\begin{bmatrix} 1 & z & z^2 & \cdots & z^{m-1} \end{bmatrix} \begin{bmatrix} \hat{C}_m - C_m \\ \vdots \\ \hat{C}_1 - C_1 \end{bmatrix} \\
&= \det\left[zI - \left(I - \hat{F}\right)\right] - \det\left[zI - (I - F)\right] = |\hat{\psi}_z| - |\psi_z| \\
&= \left.\frac{\partial |\hat{\psi}_z|}{\partial \left[Vec\left(\hat{\psi}_z\right)\right]'}\right|_{\hat{\psi}_z = \psi_z} Vec\left(\hat{\psi}_z - \psi_z\right) + o_p\left(Vec\left(\hat{\psi}_z - \psi_z\right)\right) \\
&= \left.\frac{\partial |\hat{\psi}_z|}{\partial \left[Vec\left(\hat{\psi}_z\right)\right]'}\right|_{\hat{\psi}_z = \psi_z} Vec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right) \\
&= \left[Vec\left(H_z\right)\right]' Vec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right)
\end{aligned}$$

where $H_z = \left[adj\left(\psi_z\right)\right]'$. The first equation comes from the representation of characteristic polynomial in (3.2.7). The third equation can be obtained by simply using the first order Taylor expansion. The last equation is a standard result on matrix derivatives.

Let

$$L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & m & \cdots & m^{m-1} \end{bmatrix},$$

whose k row, $k = 1,2,\cdots,m$, is equivalent to the raw vector $\begin{bmatrix} 1 & z & z^2 & \cdots & z^{m-1} \end{bmatrix}$ with

$z = k$. And L is a nonsingular matrix as $\det(L) = \prod_{1 \le j < s \le m} (s-j) \neq 0$. Let $H = \begin{bmatrix} Vec(H_1) & \cdots & Vec(H_m) \end{bmatrix}'$. Therefore, we could get

$$L \begin{bmatrix} \hat{C}_m - C_m \\ \vdots \\ \hat{C}_1 - C_1 \end{bmatrix} = HVec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right).$$

Together with the above representation for $\hat{f}_j - f_j$, we could get that

$$\hat{f}_j - f_j = F_j' L^{-1} HVec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right).$$

And, finally, by using the result in Lemma 3.3.1, we have

$$\begin{aligned}
&h\sqrt{n}Vec\left(\hat{A} - A\right) \\
&= \left(\sum_{j=1}^{m} \left\{ Vec\left[\left(I - \hat{F}\right)^{j-1}\right] F_j' L^{-1} H \right\} - \sum_{j=2}^{m} \sum_{s=0}^{j-2} f_j \left\{ (I-F)^s \otimes \left[\left(I - \hat{F}\right)'\right]^{j-2-s} \right\} \right) \\
&\times \sqrt{n}Vec\left(\hat{F} - F\right) + o_p\left(\sqrt{n}Vec\left(\hat{F} - F\right)\right) \\
&= \tilde{\Gamma}\sqrt{n}Vec\left(\hat{F} - F\right) + o_p(1) \xrightarrow{d} \Gamma N(0, V_F) \stackrel{d}{=} N\left(0, \Gamma V_F \Gamma'\right).
\end{aligned}$$

■

Proof of the formula 3.3.3. From Lemma 3.3.1, $\hat{F} \xrightarrow{a.s} F$. As eigenvalues are continuous functions of the elements of the matrix, we get the consistency of the eigenvalues

$$\hat{\lambda}_j\left(\hat{F}\right) \xrightarrow{a.s} \lambda_j(F), \quad \text{for } j = 1, 2, \cdots, m.$$

When Assumption 1 is true, $spec\{F\} \cap R_0^- = \varnothing$. Hence, when the sample size T is large enough, we could get

$$spec\left\{\hat{F}\right\} \cap R_0^- = \varnothing$$

Therefore, based on Lemma 3.2.1, $\hat{A}$ represented in (3.2.12) is the principle logarithm of $\hat{F}$. We could rewrite the relationship between $\hat{A}$ and $\hat{F}$ as $\hat{F} = \exp\{\hat{A}h\} = \sum_{j=0}^{\infty} (\hat{A}h)^j / j!$. As a result,

$$\hat{F} - F = (\hat{A} - A)\, h + \sum_{j=2}^{\infty} \frac{(\hat{A}h)^j - (Ah)^j}{j!} = (\hat{A} - A)\, h + \sum_{j=2}^{\infty} \frac{(h)^j}{j!} \left\{ \sum_{s=0}^{j-1} A^s (\hat{A} - A) (\hat{A})^{j-1-s} \right\},$$

which leads to

$$Vec(\hat{F} - F) = \underbrace{\left\{ I_{m^2} + \sum_{j=2}^{\infty} \sum_{s=0}^{j-1} \frac{(h)^{j-1}}{j!} \left[A^s \otimes (\hat{A}')^{j-1-s} \right] \right\}}_{\widetilde{E}} hVec(\hat{A} - A)$$

From $\hat{F} \xrightarrow{a.s} F$, it is easy to get $\hat{A} \xrightarrow{a.s} A$. Hence,

$$\widetilde{E} \xrightarrow{a.s} I_{m^2} + \underbrace{\sum_{j=2}^{\infty} \sum_{s=0}^{j-1} \frac{(h)^{j-1}}{j!} \left[A^s \otimes (A')^{j-1-s} \right]}_{E_2} \equiv E.$$

Let $\|\cdot\|$ denote the *Frobenius* norm of a matrix. Then, the second part satisfies

$$\begin{aligned} \|E_2\| &= \left\| \sum_{j=2}^{\infty} \sum_{s=0}^{j-1} \frac{(h)^{j-1}}{j!} \left[A^s \otimes (A')^{j-1-s} \right] \right\| \leq \sum_{j=2}^{\infty} \sum_{s=0}^{j-1} \frac{(h)^{j-1}}{j!} \|A^s\| \left\| (A')^{j-1-s} \right\| \\ &\leq \sum_{j=2}^{\infty} \sum_{s=0}^{j-1} \frac{(h)^{j-1}}{j!} \|A\|^s \|A\|^{j-1-s} = \sum_{j=2}^{\infty} \sum_{s=0}^{j-1} \frac{(h)^{j-1}}{j!} \|A\|^{j-1} = \sum_{j=2}^{\infty} \frac{(h)^{j-1}}{(j-1)!} \|A\|^{j-1} \\ &= \exp\{\|A\| h\} - 1 = O(h) \end{aligned}$$

Hence

$$\widetilde{E} \xrightarrow{a.s} E = I_{m^2} + O(h)$$

which also means that the matrix $\hat{E}$ is nonsingular almost surely when T is large and h is small. Therefore,

$$h\sqrt{n}Vec(\hat{A} - A) = \widetilde{E}^{-1} \sqrt{n} Vec(\hat{F} - F) \xrightarrow{d} E^{-1} N(0, V_F).$$

Together with the conclusion in Theorem 3.3.1, we get $\Gamma = E^{-1} = I_{M^2} + O(h)$. ■

Proof of Proposition 3.3.1. Let $\eta_1 = \mu_1 + i\nu_1$ and $\eta_2 = \mu_2 + i\nu_2$ be any two distinct eigenvalues of A, where μ_j, ν_j, for $j = 1,2$, are real numbers and $i = \sqrt{-1}$. If $e^{\eta_1 h} = e^{\eta_2 h}$, we have

$$\begin{aligned} 1 = e^{\eta_1 h}/e^{\eta_2 h} &= \exp\{(\mu_1 - \mu_2)h + i(\nu_1 - \nu_2)h\} \\ &= \exp\{(\mu_1 - \mu_2)h\}[\cos(\nu_1 - \nu_2)h + i\sin(\nu_1 - \nu_2)h]. \end{aligned}$$

This implies that $\mu_1 - \mu_2 = 0$ and $(\nu_1 - \nu_2)h = 2k\pi$, where k is any integral number. Under Assumption 1, ν_1, $\nu_2 \in (-2\pi/h, 2\pi/h)$, and hence $(\nu_1 - \nu_2)h \in (-2\pi, 2\pi)$. As a result, $k = 0$ is the only possible solution to ensure $e^{\eta_1 h} = e^{\eta_2 h}$. In this case, we have $\mu_1 - \mu_2 = 0$ and $(\nu_1 - \nu_2)h = 0$ and hence $\eta_1 = \eta_2$. This contradicts to the assumption that η_1 and η_2 are distinct. In general, all the eigenvalues of F are distinct. ■

Proof of Theorem 3.3.4. Under Assumption 3, A has the Jordan decomposition form as

$$A = Pdiag\{\eta_1, \cdots, \eta_m\}Q = PVQ \quad (.0.7)$$

Therefore, the coefficient matrix E mentioned in the proof of the formula 3.3.3 can be rewritten as

$$\begin{aligned} E &= I_{M^2} + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[A^s \otimes (A')^{j-1-s}\right] \\ &= I_{M^2} + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[(PV^sQ) \otimes \left(Q'V^{j-1-s}P'\right)\right] \\ &= I_{M^2} + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[\left(P \otimes Q'\right)\left(V^s \otimes V^{j-1-s}\right)\left(Q \otimes P'\right)\right] \\ &= \left(P \otimes Q'\right)\left\{I_{M^2} + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[\left(V^s \otimes V^{j-1-s}\right)\right]\right\}\left(Q \otimes P'\right) \\ &= \left(P \otimes Q'\right) diag\{\Lambda_1, \cdots, \Lambda_m\}\left(Q \otimes P'\right), \end{aligned}$$

where, for $k = 1, \cdots, m$,

$$\Lambda_k = diag\left(\left\{1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^s\eta_\tau^{j-1-s}\right\}_{\tau=1}^{m}\right).$$

When $k = \tau$, it is easy to get

$$1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^s\eta_\tau^{j-1-s} = 1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^{j-1} = 1 + \sum_{j=2}^{\infty}\frac{(h)^{j-1}}{(j-1)!}\eta_k^{j-1} = e^{\eta_k h}.$$

When $k \neq \tau$, as all the eigenvalues are distinct, we assume $|\eta_k| < |\eta_\tau|$ (the same result is easy to get when $|\eta_k| > |\eta_\tau|$). Then

$$\begin{aligned}
&1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^s\eta_\tau^{j-1-s} = 1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left(\frac{\eta_k}{\eta_\tau}\right)^s\eta_\tau^{j-1} \\
&= 1 + \sum_{j=2}^{\infty}\frac{(h)^{j-1}\eta_\tau^{j-1}}{j!}\frac{1-(\eta_k/\eta_\tau)^j}{1-(\eta_k/\eta_\tau)} \\
&= 1 + \sum_{j=2}^{\infty}\frac{(h)^{j-1}}{j!}\frac{\eta_\tau^j - \eta_k^j}{\eta_\tau - \eta_k} = 1 + \frac{1}{(\eta_\tau - \eta_k)h}\sum_{j=2}^{\infty}\frac{(h)^j}{j!}\left(\eta_\tau^j - \eta_k^j\right) \\
&= 1 + \frac{1}{(\eta_\tau - \eta_k)h}\{(\exp\{\eta_\tau h\} - 1 - \eta_\tau h) - (\exp\{\eta_k h\} - 1 - \eta_k h)\} \\
&= \frac{e^{\eta_\tau h} - e^{\eta_k h}}{(\eta_\tau - \eta_k)h}.
\end{aligned}$$

■

Proof of Lemma 3.3.2. Let $(\lambda_1, \cdots, \lambda_m)'$ be the ordered set of eigenvalues, p_j be the corresponding eigenvectors after normalization, and $P = \begin{pmatrix} p_1 & \cdots & p_m \end{pmatrix}$. Hence,

$$F = P\text{diag}(\lambda_1, \cdots, \lambda_m)\, p^{-1} = P\Lambda P^{-1}.$$

Since $\hat{F} \overset{a.s}{\to} F$, $\hat{F}$ is diagonalizable almost surely and could be expressed as following when T is large,

$$\hat{F} = \hat{P}\text{diag}\left(\hat{\lambda}_1, \cdots, \hat{\lambda}_m\right)\hat{P}^{-1} = \hat{P}\hat{\Lambda}\hat{P}^{-1},$$

where $\hat{p}_j$ is the normalized eigenvector corresponding to the eigenvalue $\hat{\lambda}_j$. As a result,

$$
\begin{aligned}
\hat{F}-F=dF&=\hat{P}\hat{\Lambda}\hat{P}^{-1}-P\Lambda P^{-1}\\
&=\left(\hat{P}-P\right)\hat{\Lambda}\hat{P}^{-1}+P\left(\hat{\Lambda}-\Lambda\right)\hat{P}^{-1}+P\Lambda\left(\hat{P}^{-1}-P^{-1}\right)\\
&=\left(\hat{P}-P\right)\hat{\Lambda}\hat{P}^{-1}+P\left(\hat{\Lambda}-\Lambda\right)\hat{P}^{-1}-P\Lambda P^{-1}\left(\hat{P}-P\right)\hat{P}^{-1}\\
&=(dP)\hat{\Lambda}\hat{P}^{-1}+P(d\Lambda)\hat{P}^{-1}-P\Lambda P^{-1}(dP)\hat{P}^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
P^{-1}(dF)\hat{P}&=P^{-1}(dP)\hat{\Lambda}+d\Lambda-\Lambda P^{-1}(dP)\\
&=d\Lambda+P^{-1}(dP)\left(\hat{\Lambda}-\Lambda\right)+P^{-1}(dP)\Lambda-\Lambda P^{-1}(dP)
\end{aligned}
$$

Note that the diagonal elements of $P^{-1}(dP)\Lambda$ and $\Lambda P^{-1}(dP)$ are identical (c.f., Phillips, 1982) and we get

$$
\left(P^{-1}(dF)\hat{P}\right)_{(j,j)}=\left(\hat{\lambda}_j-\lambda_j\right)+\left(P^{-1}(dP)\right)_{(j,j)}\left(\hat{\lambda}_j-\lambda_j\right).
$$

Let $\left(p^j\right)'$ and $\hat{p}_j$ denote the j^{th} row of P^{-1} and the j^{th} column of $\hat{P}$, respectively, we have

$$
\left\{1+\left[P^{-1}(dP)\right]_{(j,j)}\right\}\left(\hat{\lambda}_j-\lambda_j\right)=\left(p^j\right)'(dF)\hat{p}_j=\left[\left(p^j\right)'\otimes\hat{p}_j'\right]Vec(dF).
$$

As $\sqrt{n}Vec\left(\hat{F}-F\right)\xrightarrow{d}N(0,V_F)$, and $\hat{P}=\begin{pmatrix}\hat{p}_1 & \cdots & \hat{p}_m\end{pmatrix}\xrightarrow{a.s}p=\begin{pmatrix}p_1 & \cdots & p_m\end{pmatrix}$, we get

$$
\sqrt{n}\left(\hat{\lambda}_j-\lambda_j\right)\xrightarrow{d}\left[\left(p^j\right)'\otimes p_j'\right]N(0,V_F).
$$

Let $G=\left[\left(p^j\right)'\otimes p_j'\right]_{m\times m^2}$. We obtain the limit distribution of the eigenvalues of $\hat{F}$

as

$$\sqrt{n}\left(\hat{\lambda}-\lambda\right)=\sqrt{n}\begin{pmatrix}\hat{\lambda}_1-\lambda_1\\ \vdots\\ \hat{\lambda}_m-\lambda_m\end{pmatrix}=\sqrt{n}GVec\left(dF\right)+o_p(1)\xrightarrow{d}N\left(0,GV_FG'\right).$$

The joint asymptotic distribution of $\left(\hat{\lambda}-\lambda\right)$ and $Vec\left(\hat{F}-F\right)$ is:

$$\sqrt{n}\begin{pmatrix}\hat{\lambda}-\lambda\\ Vec\left(\hat{F}-F\right)\end{pmatrix}=\begin{pmatrix}G\\ I_{m^2}\end{pmatrix}\sqrt{n}Vec\left(\hat{F}-F\right)+o_p(1)\xrightarrow{d}N\left(0,RV_FR'\right).$$

where $R=\begin{pmatrix}G\\ I_{m^2}\end{pmatrix}_{\left(m^2+m\right)\times m^2}$. ■

Proof of Theorem 3.4.1. We only give the proof of (a), as the proof of (b) can be obtained in a similar way. First, we give the proof for the case in which $m>1$. As $A=0$ and $F=I$, straightforward calculation can give the results that $C_j=0$, for $j=1,\cdots,m$, and $f_1=0$, $f_s=-1/\left(s-1\right)$, for $s=2,\cdots,m$. we could also get

$$Th\left(\hat{A}-A\right)=T\hat{A}=T\ln\left(\hat{F}\right)=T\hat{f}_1I+\hat{f}_2T\left(I-\hat{F}\right)+\cdots+\hat{f}_mT\left(I-\hat{F}\right)^{m-1}.$$

The convergence result in (3.4.4) signifies that

$$T\left(I-\hat{F}\right)^j\xrightarrow{p}0\ \text{ for } j>1.$$

From the consistency of $\hat{f}_j$, $j=1,\cdots,m$, the following expression is obtained

$$Th\left(\hat{A}-A\right)=T\hat{f}_1I+\hat{f}_2T\left(I-\hat{F}\right)+o_p(1).$$

Note that

$$T\hat{f}_1=\int_0^1\frac{T\hat{C}_mS^{m-1}}{1+\hat{C}_1S+\cdots+\hat{C}_mS^m}dS,$$

with

$$T\hat{C}_m = T(-1)^m \det\left(I-\hat{F}\right) = n(-1)^m \sum_{j=1}^{m!} \zeta_j.$$

where ζ_j, for any j, is a multiplication of elements in matrix $\left(I-\hat{F}\right)$ with the number of m. Because that $T\left(I-\hat{F}\right) = O_p(1)$ and $m>1$, it it easy to get $T\zeta_j \xrightarrow{p} 0$. As a result, $T\hat{C}_m \xrightarrow{p} 0$. Based on the consistency of $\hat{C}_j$, $j=1,\cdots,m$, we have $T\hat{f}_1 \xrightarrow{p} 0$.Consequently,

$$Th\left(\hat{A}-A\right) = \hat{f}_2 T\left(I-\hat{F}\right) + o_p(1) \xrightarrow{d} -f_2 \cdot f\left(B_1, B_1^*, \triangle_{21}\right) = f\left(B_1, B_1^*, \triangle_{21}\right)$$

For the case $m=1$, we have

$$Th\left(\hat{A}-A\right) = T\hat{A} = T\hat{f}_1,$$

and

$$T\hat{f}_1 = T\hat{C}_1 \int_0^1 \frac{1}{1+\hat{C}_1 S} dS, \text{ with } \hat{C}_1 = (-1)\left(1-\hat{F}\right) \xrightarrow{p} 0.$$

Again, as $T\left(\hat{F}-1\right) \xrightarrow{d} f\left(B_1, B_1^*, \triangle_{21}\right)$, we have

$$Th\left(\hat{A}-A\right) = T\hat{C}_1 \int_0^1 \frac{1}{1+\hat{C}_1 S} dS = T\left(\hat{F}-1\right) + o_p(1) \xrightarrow{d} f\left(B_1, B_1^*, \triangle_{21}\right)$$

■

Proof of Theorem 3.4.2. When $b=0$, we have $T\left(\hat{F}-I\right) \xrightarrow{d} f\left(B_1, B_1^*, \triangle_{21}\right)$. Then, it is easy to get

$$\begin{aligned} T\left\{\sum_{j=1}^{m} \hat{\lambda}_j - m\right\} &= T\sum_{j=1}^{m}\left(\hat{\lambda}_j - 1\right) = T \times tr\left(\hat{F}-I\right) \\ &= T\Delta Vec\left(\hat{F}-I\right) \xrightarrow{d} \Delta \cdot Vec\left[f\left(B_1, B_1^*, \triangle_{21}\right)\right], \end{aligned}$$

where Δ is a row vector of dimension m^2, whose $1^{st}, [m+2]^{th}, \cdots, [(m-1)m+m]^{th}$ elements are 1 and 0 otherwise. The other parts of the theorem can be easily proved by using the same method. ■

C Proofs in Chapter 4

Proof of Proposition 4.2.1. (a), (b) are similar to those in PM and are omitted.

(c) Notice that $\{\varepsilon_{th}\}_{t=1}^{T}$ is a martingale-difference array. For any fixed h and N, $T = \frac{N}{h}$, $\{\varepsilon_{th}\}_{t=1}^{T} \overset{iid}{\sim} (0,1)$. Therefore,

$$\sum_{t=1}^{T} E\left(\left(\frac{\varepsilon_{th}}{\sqrt{T}}\right)^2 \middle| \mathscr{F}_{T,t-1}\right) = \sum_{t=1}^{T} \frac{1}{T} = 1,$$

and

$$\max_{1\le t\le T} E\left(\left(\frac{\varepsilon_{th}}{\sqrt{T}}\right)^2 \middle| \mathscr{F}_{T,t-1}\right) = \frac{1}{T} \to 0.$$

The Lindeberg condition holds since, when $h \to 0$, $N \to \infty$, for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{t=1}^{T} E\left(\left(\frac{\varepsilon_{th}}{\sqrt{T}}\right)^2 1\left\{\left|\frac{\varepsilon_{th}}{\sqrt{T}}\right| > \varepsilon\right\}\right) \\
= \;& E\left(\varepsilon_{1h}^2 1\left\{\varepsilon_{1h}^2 > T\varepsilon^2\right\}\right) \le \left(E\left(|\varepsilon_{1h}|^{2+\delta}\right)\right)^{\frac{2}{2+\delta}} \left(P\left\{\varepsilon_{1h}^2 > T\varepsilon^2\right\}\right)^{\frac{\delta}{2+\delta}} \\
\le \;& \left(E\left(|\varepsilon_{1h}|^{2+\delta}\right)\right)^{\frac{2}{2+\delta}} \left(\frac{E\left(\varepsilon_{1h}^2\right)}{T\varepsilon^2}\right)^{\frac{\delta}{2+\delta}} \le M^{\frac{2}{2+\delta}} \left(\frac{1}{T\varepsilon^2}\right)^{\frac{\delta}{2+\delta}} \to 0,
\end{aligned}$$

by the assumption that for some $\delta > 0$, $E\left(|\varepsilon_{1h}|^{2+\delta}\right)$ is uniformly bounded about h when h is small. Using the Central Limit Theory for the martingale-difference array (see e.g. Hall and Heyde, 1980, Theorem 3.5), we get $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{th} \Longrightarrow N(0,1)$.

(d) Denote $a_h(\kappa) = a_h$ when there is no confusion. By the Cramér-Wold device (e.g. Kallenberg, 2002, Corollary 5.5), it is sufficient to show that

$$aX_{Th} + bY_{Th} \Longrightarrow aX + bY \quad \text{for all } a, b \in \mathbb{R}, \tag{.0.8}$$

where X and Y are independent $N(0, 1/(-2\kappa))$ random variables. If Z is an $N\left(0, \left(a^2+b^2\right)/(-2\kappa)\right)$ random variable, $aX + bY \overset{d}{=} Z$, so $aX_{Th} + bY_{Th} \Longrightarrow Z$ is sufficient for (.0.8). We can write $aX_{Th}+bY_{Th}=\sum_{t=1}^{T}\zeta_{Tt}$, where

$$\{\zeta_{Tt}\}_{t=1}^{T}=\left\{\left(a\,[a_h]^{-(T-t)-1}+b\,[a_h]^{-t}\right)\frac{\varepsilon_{th}}{\sqrt{k_T}}\right\}_{t=1}^{T}$$

is a martingale-difference array, as $\{\varepsilon_{th}\}_{t=1}^{T}\overset{iid}{\sim}(0,1)$ for any fixed h, N and $T=N/h$. Hence,

$$\begin{aligned}\sum_{t=1}^{T}E\left(\left(\zeta_{Tt}\right)^2\middle|\mathscr{F}_{T,t-1}\right) &= \frac{1}{k_T}\sum_{t=1}^{T}\left(a\,[a_h]^{-(T-t)-1}+b\,[a_h]^{-t}\right)^2\\ &= \frac{1}{k_T}\left\{\sum_{t=1}^{T}a^2\,[a_h]^{-2(T-t)-2}+\sum_{t=1}^{T}b^2\,[a_h]^{-2t}+2Tab\,[a_h]^{-T-1}\right\}\\ &= \left\{a^2+b^2\right\}\frac{1}{k_T}\sum_{t=1}^{T}[a_h]^{-2t}+O\left(\frac{T\,[a_h]^{-T}}{k_T}\right)\\ &= \frac{a^2+b^2}{-2\kappa}+o(1)\;\to\;\frac{a^2+b^2}{-2\kappa},\qquad \text{as } h\to 0,\ N\to\infty.\end{aligned}$$

And as $a_h=a_h(\kappa)=\exp\{-kh\}>1$,

$$\begin{aligned}\max_{1\le t\le T}E\left(\left(\zeta_{Tt}\right)^2\middle|\mathscr{F}_{T,t-1}\right) &= \max_{1\le t\le T}\left(a\,[a_h]^{-(T-t)-1}+b\,[a_h]^{-t}\right)^2\frac{1}{k_T}\\ &\le \max_{1\le t\le T}2\left(a^2\,[a_h]^{-2(T-t)-2}+b^2\,[a_h]^{-2t}\right)\frac{1}{k_T}\\ &\le 2\left(a^2+b^2\right)\frac{1}{k_T}\to 0\qquad \text{as } h\to 0,\ N\to\infty.\end{aligned}$$

The Lindeberg condition holds since, when $h \to 0$, $N \to \infty$, for any $\varepsilon > 0$,

$$
\begin{aligned}
& \sum_{t=1}^{T} E\left((\zeta_{Tt})^2 1\{|\zeta_{Tt}| > \varepsilon\} \right) \\
= & \sum_{t=1}^{T} \frac{\left(a(a_h)^{-(T-t)-1} + b(a_h)^{-t} \right)^2}{k_T} E\left((\varepsilon_{th})^2 1\left\{ \frac{\left| a(a_h)^{-(T-t)-1} + b(a_h)^{-t} \right| |\varepsilon_{th}|}{\sqrt{k_T}} > \varepsilon \right\} \right) \\
\leq & \left\{ \frac{a^2+b^2}{-2\kappa} + o(1) \right\} \max_{1 \leq t \leq T} E\left((\varepsilon_{th})^2 1\left\{ \left(a(a_h)^{-(T-t)-1} + b(a_h)^{-t} \right)^2 (\varepsilon_{th})^2 > k_T \varepsilon^2 \right\} \right) \\
\leq & \left\{ \frac{a^2+b^2}{-2\kappa} + o(1) \right\} \max_{1 \leq t \leq T} E\left((\varepsilon_{th})^2 1\left\{ 2\left(a^2+b^2\right)(\varepsilon_{th})^2 > k_T \varepsilon^2 \right\} \right) \\
= & \left\{ \frac{a^2+b^2}{-2\kappa} + o(1) \right\} E\left((\varepsilon_{1h})^2 1\left\{ 2\left(a^2+b^2\right)(\varepsilon_{1h})^2 > k_T \varepsilon^2 \right\} \right) \\
\leq & \left\{ \frac{a^2+b^2}{-2\kappa} + o(1) \right\} \left\{ E\left(|\varepsilon_{1h}|^{2+\delta} \right) \right\}^{\frac{2}{2+\delta}} \left(P\left\{ 2\left(a^2+b^2\right)(\varepsilon_{1h})^2 > k_T \varepsilon^2 \right\} \right)^{\frac{\delta}{2+\delta}} \\
\leq & \left\{ \frac{a^2+b^2}{-2\kappa} + o(1) \right\} \left\{ E\left(|\varepsilon_{1h}|^{2+\delta} \right) \right\}^{\frac{2}{2+\delta}} \left\{ \frac{2\left(a^2+b^2\right)}{k_T \varepsilon^2} \right\}^{\frac{\delta}{2+\delta}} \\
= & \; o(1),
\end{aligned}
$$

by the assumption that for some $\delta > 0$, $E\left(|\varepsilon_{1h}|^{2+\delta} \right)$ is uniformly bounded about h when h is small. Using the same Central Limit Theory for the martingale-difference array, we get $aX_{Th} + bY_{Th} \Longrightarrow Z$, for all $a, b \in \mathbb{R}$, establishing (.0.8).

■

Proof of Proposition 4.2.1. (a) Denote $a_h(\kappa) = a_h$ when there is no confusion. From Model (4.2.8) we get

$$
x_{th} = a_h x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th} = \tilde{g}_h \frac{1 - a_h^t}{1 - a_h} + \sum_{j=1}^{t} a_h^{t-j} \varepsilon_{jh} + a_h^t x_{0h}.
$$

Hence,

$$\begin{aligned}
\frac{a_h^{-T}}{\sqrt{k_T}} x_{Th} &= \frac{a_h^{-T}}{\sqrt{k_T}} \left(\tilde{g}_h \frac{1-a_h^T}{1-a_h} + \sum_{j=1}^{T} a_h^{T-j} \varepsilon_{jh} + a_h^T x_{0h} \right) \\
&= \sqrt{k_T} \tilde{g}_h \frac{a_h^{-T} - 1}{(1-a_h)\, k_T} + Y_{Th} + \frac{x_{0h}}{\sqrt{k_T}} \\
&\Longrightarrow \frac{\kappa\mu + \sigma i \psi'(0)}{-\kappa\sigma\sqrt{\psi''(0)}} + Y + \frac{y_{0h}}{\sigma\sqrt{\psi''(0)}} \\
&= \sqrt{\frac{1}{-2\kappa}} \left(\sqrt{-2\kappa} Y + \sqrt{-2\kappa} \frac{\kappa\mu + \sigma i \psi'(0) - \kappa y_{0h}}{-\kappa\sigma\sqrt{\psi''(0)}} \right) = \sqrt{\frac{1}{-2\kappa}} [\eta + D].
\end{aligned}$$

From Lemma 4.2.1 (c), η is $N(0,1)$ variate.

Notice that $x_{th} - x_{(t-1)h} = (a_h - 1)\, x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th}$. Therefore,

$$x_{Th} - x_{0h} = (a_h - 1) \sum_{t=1}^{T} x_{(t-1)h} + T \tilde{g}_h + \sum_{t=1}^{T} \varepsilon_{th},$$

and

$$\begin{aligned}
\frac{a_h^{-T}(a_h - 1)}{\sqrt{k_T}} \sum_{t=1}^{T} x_{(t-1)h} &= \frac{a_h^{-T}}{\sqrt{k_T}} x_{Th} - a_h^{-T} \frac{x_{0h}}{\sqrt{k_T}} - \frac{a_h^{-T} T}{k_T} \sqrt{k_T} \tilde{g}_h - \frac{a_h^{-T}\sqrt{T}}{\sqrt{k_T}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{th} \\
&= \frac{a_h^{-T}}{\sqrt{k_T}} x_{Th} + o_p(1) \to \sqrt{\frac{1}{-2\kappa}} [\eta + D],
\end{aligned}$$

by Lemma 4.2.1 (a), (c), $a_h^{-T} = o(k_T/T)$ and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{th} \sim O_p(1)$.

(b) From Model (4.2.8) we get

$$x_{th} = a_h x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th} = \tilde{g}_h \frac{1-a_h^t}{1-a_h} + \sum_{j=1}^{t} a_h^{t-j} \varepsilon_{jh} + a_h^t x_{0h},$$

and

$$
\begin{aligned}
&\sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} \\
= \quad & \tilde{g}_h \sum_{t=1}^{T} \frac{1-a_h^{t-1}}{1-a_h}\varepsilon_{th} + \sum_{t=1}^{T}\left[\sum_{j=1}^{t-1} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} + x_{0h}\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th} \\
= \quad & \tilde{g}_h \sum_{t=1}^{T} \frac{1-a_h^{t-1}}{1-a_h}\varepsilon_{th} + \sum_{t=1}^{T}\left[\sum_{j=1}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} - \sum_{t=1}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} + x_{0h}\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th} \\
= \quad & \tilde{g}_h \sum_{t=1}^{T} \frac{1-a_h^{t-1}}{1-a_h}\varepsilon_{th} + \left[\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\right]\left[\sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh}\right] - \sum_{t=1}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} + x_{0h}\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th} \\
= \quad & \frac{\tilde{g}_h}{1-a_h}\sum_{t=1}^{T}\varepsilon_{th} + \left[\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\right]\left[\frac{\tilde{g}_h}{a_h-1} + \sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh} + x_{0h}\right] - \sum_{t=1}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th}.
\end{aligned}
$$

From Lemma 4.2.1 (b), (c), we have $\frac{a_h^{-T}}{k_T}\sum_{t=1}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} \xrightarrow{L_1} 0$, and $\frac{a_h^{-T}\tilde{g}_h}{(1-a_h)k_T}\sum_{t=1}^{T}\varepsilon_{th} = \frac{\sqrt{N}\exp\{\kappa N\}}{(1-a_h)k_T}\tilde{g}_h\sqrt{k_T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th}\right) = o_p(1)$ as $h \to 0$, $N \to \infty$. Hence,

$$
\begin{aligned}
&\frac{a_h^{-T}}{k_T}\sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} \\
= \quad & \left[\frac{1}{\sqrt{k_T}}\sum_{t=1}^{T} a_h^{-(T-t)-1}\varepsilon_{th}\right]\left[\frac{\tilde{g}_h}{(a_h-1)\sqrt{k_T}} + \frac{1}{\sqrt{k_T}}\sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh} + \frac{x_{0h}}{\sqrt{k_T}}\right] + o_p(1) \\
= \quad & X_{Th}\left[\frac{\tilde{g}_h}{(a_h-1)\sqrt{k_T}} + Y_{Th} + \frac{x_{0h}}{\sqrt{k_T}}\right] + o_p(1) \\
\Longrightarrow \quad & X\left[\frac{\kappa\mu + \sigma i\psi'(0)}{-\kappa\sigma\sqrt{\psi''(0)}} + Y + \frac{y_{0h}}{\sigma\sqrt{\psi''(0)}}\right] \\
= \quad & \frac{1}{-2\kappa}\left[\sqrt{-2\kappa}X\right]\left[\sqrt{-2\kappa}Y + \sqrt{-2\kappa}\frac{\kappa\mu + \sigma i\psi'(0) - \kappa y_{0h}}{-\kappa\sigma\sqrt{\psi''(0)}}\right] = \frac{1}{-2\kappa}\xi\left[\eta + D\right],
\end{aligned}
$$

where, by Lemma 4.2.1 (c), ξ and η are independent $N(0,1)$ variates.

(c) Since $x_{th} = a_h x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th}$, we get

$$
x_{th}^2 = a_h^2 x_{(t-1)h}^2 + 2\tilde{g}_h a_h x_{(t-1)h} + 2a_h x_{(t-1)h}\varepsilon_{th} + \tilde{g}_h^2 + \varepsilon_{th}^2 + 2\tilde{g}_h\varepsilon_{th},
$$

and

$$x_{th}^2 - x_{(t-1)h}^2 = \left(a_h^2 - 1\right) x_{(t-1)h}^2 + 2\tilde{g}_h a_h x_{(t-1)h} + 2a_h x_{(t-1)h}\varepsilon_{th} + \tilde{g}_h^2 + \varepsilon_{th}^2 + 2\tilde{g}_h\varepsilon_{th}.$$

Hence,

$$\begin{aligned} & \left(a_h^2 - 1\right) \sum_{t=1}^{T} x_{(t-1)h}^2 \\ = \;& \left(x_{Th}^2 - x_{0h}^2\right) - 2\tilde{g}_h a_h \sum_{t=1}^{T} x_{(t-1)h} - 2a_h \sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} - T\tilde{g}_h^2 - \sum_{t=1}^{T} \varepsilon_{th}^2 - 2\tilde{g}_h \sum_{t=1}^{T} \varepsilon_{th}. \end{aligned}$$

Notice that $a_h^{-2T}\frac{x_{0h}^2}{k_T} \to 0$, $\frac{a_h^{-2T}}{k_T}T\tilde{g}_h^2 \to 0$. The proof of Part (a) and (b) suggests

$$2\tilde{g}_h a_h \frac{a_h^{-2T}}{k_T} \sum_{t=1}^{T} x_{(t-1)h} = 2\tilde{g}_h \sqrt{k_T} \frac{a_h^{-T+1}}{[a_h - 1] k_T} \left(\frac{a_h^{-T}[a_h - 1]}{\sqrt{k_T}} \sum_{t=1}^{T} x_{(t-1)h} \right) \to 0,$$

$$2a_h \frac{a_h^{-2T}}{k_T} \sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} = 2a_h^{-T+1} \left(\frac{a_h^{-T}}{k_T} \sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} \right) \to 0.$$

And, as $E\left|\frac{a_h^{-2T}}{k_T}\sum_{t=1}^{T}\varepsilon_{th}^2\right| = \frac{a_h^{-2T}}{k_T}\sum_{t=1}^{T}E\left(\varepsilon_{th}^2\right) = \frac{a_h^{-2T}T}{k_T} \to 0$, we get $\frac{a_h^{-2T}}{k_T}\sum_{t=1}^{T}\varepsilon_{th}^2 \xrightarrow{L_1} 0$. By Lemma 4.2.1 (c),

$$2\tilde{g}_h \frac{a_h^{-2T}}{k_T} \sum_{t=1}^{T} \varepsilon_{th} = 2\tilde{g}_h \frac{a_h^{-2T}T}{k_T} \left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{th} \right) \to 0.$$

Therefore, from the proof of Part (a), we have

$$\left(a_h^2 - 1\right) \frac{a_h^{-2T}}{k_T} \sum_{t=1}^{T} x_{(t-1)h}^2 = \left(\frac{a_h^{-T}}{\sqrt{k_T}} x_{Th} \right)^2 + o_p(1) \Longrightarrow \frac{1}{-2\kappa} (\eta + D)^2.$$

(d) This is an immediate consequence of (a), (b) and (c).

(e) Since $\kappa = -(1/h)\ln(a_h(\kappa))$ and $\widehat{\kappa} = -(1/h)\ln(\widehat{a}_h(\kappa))$, by the mean value theorem,

$$-h(\widehat{\kappa} - \kappa) = \ln(\widehat{a}_h(\kappa)) - \ln(a_h(\kappa)) = \frac{1}{\widetilde{a}_h(\kappa)} (\widehat{a}_h(\kappa) - a_h(\kappa)) \qquad (.0.9)$$

for some $\widetilde{a}_h(\kappa)$ whose value is between $\widehat{a}_h(\kappa)$ and $a_h(\kappa)$. The *Delta* method is not directly applicable since $a_h(\kappa)$ is a not constant but a real sequence that goes to 1 as $h \to 0$. However, if we can show $\widetilde{a}_h(\kappa) \xrightarrow{p} 1$, we can obtain the limiting distribution for $\widehat{\kappa}$. For any $\varepsilon > 0$, when h is small enough, $|a_h(\kappa) - 1| < \varepsilon/2$, and

$$\begin{aligned}
\Pr\{|\widetilde{a}_h(\kappa) - 1| > \varepsilon\} &= \Pr\{|\widetilde{a}_h(\kappa) - a_h(\kappa) + a_h(\kappa) - 1| > \varepsilon\} \\
&\leq \Pr\{|\widetilde{a}_h(\kappa) - a_h(\kappa)| + |a_h(\kappa) - 1| > \varepsilon\} \\
&\leq \Pr\{|\widehat{a}_h(\kappa) - a_h(\kappa)| + |a_h(\kappa) - 1| > \varepsilon\} \\
&= \Pr\{|\widehat{a}_h(\kappa) - a_h(\kappa)| > \varepsilon - |a_h(\kappa) - 1|\} \\
&\leq \Pr\{|\widehat{a}_h(\kappa) - a_h(\kappa)| > \varepsilon/2\} \\
&\to 0, \quad \text{as } h \to 0, \text{ and } N \to \infty,
\end{aligned}$$

where the first inequality is the triangular inequality, the second comes from the fact that $\widetilde{a}_h(\kappa)$ is between $\widehat{a}_h(\kappa)$ and $a_h(\kappa)$, and the final result based on the fact that $\widehat{a}_h(\kappa) - a_h(\kappa) \xrightarrow{p} 0$. Hence, $\widetilde{a}_h(\kappa) \xrightarrow{p} 1$ and

$$\frac{e^{-\kappa N}}{2\kappa}(\widehat{\kappa} - \kappa) = \frac{1 - (a_h(\kappa))^2}{2\kappa h} \frac{1}{\widetilde{a}_h(\kappa)} \left[\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2 - 1} (\widehat{a}_h(\kappa) - a_h(\kappa)) \right] \Longrightarrow \frac{\xi}{\eta + D}. \tag{.0.10}$$

■

Proof of Proposition 4.4.1. (a) Letting $\gamma = b/\left(\sigma\sqrt{N}\right)$, Perron (1991) derived the joint MGF of $A(\gamma, c)$ and $B(\gamma, c)$ as

$$\begin{aligned}
& M(v,u) \\
=\; & E\left[\exp\left(vA(\gamma,c) + uB(\gamma,c)\right)\right] \\
=\; & \Psi_c(v,u) \exp\left\{ -\left(\frac{\gamma^2}{2}\right)(v + c - \lambda)\left[1 - \exp(v + c + \lambda)\Psi_c^2(v,u)\right] \right\} \\
=\; & \underbrace{\Psi_c(v,u)}_{I} \underbrace{\exp\left\{ -\left(\frac{\gamma^2}{2}\right)(v + c - \lambda) \right\}}_{II} \underbrace{\exp\left\{ \left(\frac{\gamma^2}{2}\right)(v + c - \lambda)\exp(v + c + \lambda)\Psi_c^2(v,u) \right\}}_{III},
\end{aligned}$$

where

$$\lambda = \left(c^2 + 2cv - 2u\right)^{1/2},$$

$$\Psi_c(v,u) = \left[\frac{2\lambda \exp\{-(v+c)\}}{(\lambda + (v+c))\exp\{-\lambda\} + (\lambda - (v+c))\exp\{\lambda\}}\right]^{1/2}.$$

Let $v = \widetilde{v}(2c)e^{-c}$ and $u = \widetilde{u}(2c)^2 e^{-2c}$. The joint MGF of $(2c)\, e^{-c} A(\gamma, c)$ and $(2c)^2\, e^{-2c} B(\gamma, c)$ is

$$M(\widetilde{v}, \widetilde{u}) = E\left[\exp\left(\widetilde{v}(2c)e^{-c} A(\gamma, c) + \widetilde{u}(2c)^2 e^{-2c} B(\gamma, c)\right)\right].$$

We get

$$\begin{aligned}
\lambda &= \left\{c^2 + (2c)^2 e^{-c}\widetilde{v} - 2(2c)^2 e^{-2c}\widetilde{u}\right\}^{1/2} \\
&= \left\{\left[c + (2c)e^{-c}\widetilde{v} - 2(2c)e^{-2c}\widetilde{u} - (2c)\, e^{-2c}\widetilde{v}^2\right]^2 + O\left(e^{-3c}\right)\right\}^{1/2} \\
&= c + (2c)e^{-c}\widetilde{v} - 2(2c)e^{-2c}\widetilde{u} - (2c)\, e^{-2c}\widetilde{v}^2 + O\left(e^{-3c}\right),
\end{aligned}$$

$$\lambda + (v+c) = 2c + 2(2c)e^{-c}\widetilde{v} - 2(2c)e^{-2c}\widetilde{u} - (2c)e^{-2c}\widetilde{v}^2 + O\left(e^{-3c}\right),$$

$$\lambda - (v+c) = -2\,(2c)\, e^{-2c}\widetilde{u} - (2c)\, e^{-2c}\widetilde{v}^2 + O\left(e^{-3c}\right),$$

$$e^{-\lambda} = e^{-c} - (2c)e^{-2c}\widetilde{v} + O\left(e^{-3c}\right),$$

and

$$(\lambda - (v+c))\, e^{\lambda} = -\,(2c)\, e^{-c}\left[2\widetilde{u} + \widetilde{v}^2\right] + O\left(e^{-2c}\right).$$

The denominator of $\Psi_c^2(v,u)$ is

$$(\lambda + (v+c))\, e^{-\lambda} + (\lambda - (v+c))\, e^{\lambda} = (2c)\, e^{-c}\left[1 - 2\widetilde{u} - \widetilde{v}^2\right] + O\left(e^{-2c}\right).$$

The numerator of $\Psi_c^2(v,u)$ is

$$2\lambda \exp\{-(v+c)\} = 2\lambda \exp\left\{-(2c)e^{-c}\widetilde{v} - c\right\} = (2c)\, e^{-c} + O\left(e^{-2c}\right).$$

Hence,

$$I = \Psi_c(v,u) = \left\{ \frac{(2c)\,e^{-c} + O\left(e^{-2c}\right)}{(2c)\,e^{-c}\left[1 - 2\widetilde{u} - \widetilde{v}^2\right] + O\left(e^{-2c}\right)} \right\}^{1/2} \to \left\{ \frac{1}{1 - 2\widetilde{u} - \widetilde{v}^2} \right\}^{1/2}.$$

It is easy to show that $II \to 1$ because

$$-\left(\frac{\gamma^2}{2}\right)(v + c - \lambda) = \left(\frac{-b^2\kappa}{2\sigma^2 c}\right)\left[-2\,(2c)\,e^{-2c}\widetilde{u} - (2c)\,e^{-2c}\widetilde{v}^2 + O\left(e^{-3c}\right)\right] \to 0.$$

Since

$$\exp\{\lambda + v + c\} = e^{2c}\exp\left\{2(2c)e^{-c}\widetilde{v} - 2(2c)e^{-2c}\widetilde{u} - (2c)e^{-2c}\widetilde{v}^2 + O\left(e^{-3c}\right)\right\},$$

letting $d = b\sqrt{-2\kappa}/\sigma$, we get $(2c)\,\gamma^2 = d^2$ and

$$\begin{aligned} & \left(\frac{\gamma^2}{2}\right)(v + c - \lambda)\exp\{\lambda + v + c\} \\ = \;& \frac{d^2}{2}\left[2\widetilde{u} + \widetilde{v}^2 + O\left(e^{-c}\right)\right]\exp\left\{2(2c)e^{-c}\widetilde{v} - 2(2c)e^{-2c}\widetilde{u} - (2c)e^{-2c}\widetilde{v}^2 + O\left(e^{-3c}\right)\right\} \\ \to \;& \frac{d^2}{2}\left[2\widetilde{u} + \widetilde{v}^2\right]. \end{aligned}$$

Therefore,

$$III \to \exp\left\{ \frac{d^2\left[2\widetilde{u} + \widetilde{v}^2\right]}{2\left[1 - 2\widetilde{u} - \widetilde{v}^2\right]} \right\}.$$

The limiting behavior of I, II and III gives rise to the limiting joint MGF of $(2c)\,e^{-c}A(\gamma,c)$ and $(2c)^2\,e^{-2c}B(\gamma,c)$.

(b) Since ξ and η are independent $N(0,1)$ random variables and $d = b\sqrt{-2\kappa}/\sigma$

is a constant, we have

$$
\begin{aligned}
M(\widetilde{v},\widetilde{u}) &= E\left\{\exp\left(\xi\,[d+\eta]\,\widetilde{v}+[d+\eta]^2\,\widetilde{u}\right)\right\} \\
&= E\left\{E\left[\exp\left(\xi\,[d+\eta]\,\widetilde{v}+[d+\eta]^2\,\widetilde{u}\right)\middle|\mathcal{F}_{\xi}\right]\right\} \\
&= E\left\{\exp\left([d+\eta]^2\,\widetilde{u}\right)\exp\left(\frac{[d+\eta]^2\,\widetilde{v}^2}{2}\right)\right\} \\
&= \frac{1}{\{1-2\widetilde{u}-\widetilde{v}^2\}^{1/2}}\exp\left\{\frac{d^2\left[2\widetilde{u}+\widetilde{v}^2\right]}{2\left[1-2\widetilde{u}-\widetilde{v}^2\right]}\right\}.
\end{aligned}
$$

This is the joint MGF of $\xi\,[d+\eta]$ and $[d+\eta]^2$ and is equivalent to the result in (a).

(c) This is an immediate consequence of (b). ■

Proof of Proposition 4.4.2. (a) and (b) are the classical results from Phillips (1987) and also identical to (ii) and (iv) of Lemma A.2 in Perron (1991).

(c) Note that $2c\int_0^1\exp\{cr\}\,J_c(r)\,dr = e^c J_c(1)-\int_0^1\exp\{cr\}\,dW(r)$ and hence,

$$
\begin{aligned}
&\left\{(2c)^{3/2}\,e^{-2c}\int_0^1\exp\{cr\}\,J_c(r)\,dr\right\}^2 \\
&= (2c)\,e^{-4c}\left\{e^c J_c(1)-\int_0^1\exp\{cr\}\,dW(r)\right\}^2 \\
&= (2c)\,e^{-2c}J_c(1)^2+e^{-2c}\left[(2c)^{1/2}\,e^{-c}\int_0^1\exp\{cr\}\,dW(r)\right]^2 \\
&\quad -2e^{-c}\left[(2c)^{1/2}\,e^{-c}J_c(1)\right]\left[(2c)^{1/2}\,e^{-c}\int_0^1\exp\{cr\}\,dW(r)\right].
\end{aligned}
$$

By stochastic differentiation of $\{\int_0^r\exp\{-cs\}\,dW(s)\}^2$, we deduce the following useful relationship, as pointed out in Phillips (1987),

$$
\{J_c(1)\}^2 = 1+2c\int_0^1 J_c(r)^2\,dr+2\int_0^1 J_c(r)\,dW(r).
$$

From (a) and (b), we get

$$
\begin{aligned}
&\left\{(2c)^{1/2}\,e^{-c}J_c(1)\right\}^2 \\
&= (2c)\,e^{-2c}+(2c)^2\,e^{-2c}\int_0^1 J_c(r)^2\,dr+2\,(2c)\,e^{-2c}\int_0^1 J_c(r)\,dW(r)\Longrightarrow\eta^2.
\end{aligned}
$$

As $\int_0^1 \exp\{cr\}\, dW(r) \sim N\left(0, \frac{\exp\{2c\}-1}{2c}\right)$, $(2c)^{1/2} e^{-c} \int_0^1 \exp\{cr\}\, dW(r)$ is $O_p(1)$, we have

$$\left\{(2c)^{3/2} e^{-2c} \int_0^1 \exp\{cr\} J_c(r)\, dr\right\}^2 = \left\{(2c)^{1/2} e^{-c} J_c(1)\right\}^2 + o_p(1) \Longrightarrow \eta^2.$$

(d) Based on the results in (a), (b), (c), and $2c\gamma^2 = -2\kappa b^2/\sigma^2 = d^2$, we get

$$\begin{aligned}
& \frac{(2c)\, e^{-c} A(\gamma, c)}{(2c)^2 e^{-2c} B(\gamma, c)} \\
= & \frac{(2c)^{1/2} \gamma \left[(2c)^{1/2} e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right] + (2c)\, e^{-c} \int_0^1 J_c(r)\, dW(r)}{\gamma^2 (2c) [1 - e^{-2c}] + 2\gamma (2c)^{1/2} \left[(2c)^{3/2} e^{-2c} \int_0^1 \exp\{cr\} J_c(r)\, dr\right] + (2c)^2 e^{-2c} \int_0^1 J_c(r)^2\, dr} \\
= & \frac{d \left[(2c)^{1/2} e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right] + [\xi\eta + o_p(1)]}{[d^2 + o(1)] + 2d [\eta^2 + o_p(1)]^{1/2} + [\eta^2 + o_p(1)]} \\
= & \frac{d \left[(2c)^{1/2} e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right] + [\xi\eta + o_p(1)]}{[d + \eta]^2 + o_p(1)}.
\end{aligned}$$

From Theorem 4.4.1 we have

$$\left((2c)\, e^{-c} A(\gamma, c), (2c)^2 e^{-2c} B(\gamma, c)\right) \Longrightarrow \left(\xi\, [d + \eta], [d + \eta]^2\right).$$

Therefore, $(2c)^{1/2} e^{-c} \int_0^1 \exp\{cr\}\, dW(r) \Longrightarrow \xi$. ∎