Limit Theory for Multivariate Linear Diffusion Estimation*

Xiaohu Wang †
Chinese University of Hong Kong

April 22, 2014

Abstract

This paper provides a new representation of the maximum likelihood (ML) estimator of the "mean reversion matrix" in the multivariate Ornstein-Uhlenbeck process. The new representation enjoys three advantages over the traditional matrix logarithmic representation. First, while the traditional matrix logarithmic representation involves an infinite polynomial series, the new representation involves only finite polynomial series whose degree is the same as the dimension of the multivariate process. This formulation avoids truncation of an infinite series and facilitates the use of the delta method and the calculation of the covariance matrix in the limit distribution. Second, the new representation has a wider domain of convergence than the traditional series. Third, the new representation readily accommodates prior knowledge about the model (e.g., the cointegrating rank) to simplify the estimating procedure, whereas it is hard to do so with the traditional series approach. The limit theory of the ML estimator based on the new representation is established for a wide range of cases, including stationary, pure unit root, or partially nonstationary processes. Special attention is given to provide an explicit expression of the asymptotic covariance matrix for low dimensional models. An empirical application is conducted on an affine term structure model, illustrating the advantages of the new representation. Although the theory is derived for linear diffusions, the method proposed in the paper is applicable to nonlinear diffusion models.

JEL Classification: C13, C32, G12.

Keywords: Multivariate diffusions; Exact maximum likelihood estimator; Principal logarithm; Limit theory; Stationarity and non-stationarity.

*My special thanks go to professor Peter Phillips and professor Jun Yu at singapore management univeristy for their valuable comments.

†Xiaohu Wang, Department of Economics, Chinese University of Hong Kong. Email: xiaohu.wang@cuhk.edu.hk

1 Introduction

Multivariate continuous time models received considerable interest in macro-econometrics over the period from 1960s to 1980s, and featured in theoretical contributions such as in Bergstrom (1966, 1984), Phillips (1972) as well as in applications such as Bergstrom and Wymer (1976) and Knight and Wymer (1978). Over the last two decades, they are once again at the forefront in the econometrics literature. The main fuel for the resurgence is the usefulness of these models in the development of modern asset pricing theory. Given the complicated interplay among economic and financial variables, it is not surprising that multivariate continuous time models, which allow for interactions among variables, are receiving 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 of interest rates (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 (LS); see, for example, Phillips (1972). In finance, many successful models allow the diffusion function to be time varying but maintain linearity for the drift function. To match the development of these 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).

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 LS 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 models is standard; see, for example, Mann and Wald (1943) for the stationary case, Phillips and Durlauf (1986) for the unit root case, and Ahn and Reinsel (1990), Park and Phillips (1988, 1989), Sims, Stock and Watson (1990) for the partially nonstationary case.

It is known that the parameter matrix in the continuous time model which measures the mean reversion speed of the continuous process in stationary case is the logarithmic transformation of the autoregressive (AR) coefficient matrix. In this paper, for notational simplicity, the matrix is referred to as the "mean reversion matrix" even if the system is not stationary. To ensure the relation to be bijective, some identification conditions must be imposed. Various identification conditions have been proposed (see, e.g., Phillips, 1973; Hansen and Sargent, 1983; McCrorie, 2003), but the simplest one which is applicable to all the cases is a condition that limits the mean reversion matrix on the principal logarithm of the AR coefficient matrix. Subsection 3.1 gives the formula of calculating matrix's principal logarithm, which is commonly used to derive an explicit relationship between the mean reversion matrix and the AR coefficient matrix. Based on this explicit relationship, the traditional expression of ML estimator of the mean reversion matrix is established.

However, this traditional expression suffers from three drawbacks. First, it is an infinite polynomial series of the estimated VAR coefficient matrix. Consequently, when the delta method is used to find the limit distribution of the estimated mean reversion matrix, another infinite polynomial series appears in the expression of the asymptotic covariance matrix. Both of these two infinite series need to be truncated in practice. Unfortunately, there is no clear guideline as to how to do the truncations. Obviously, the truncation rule should depend on the estimated value of VAR coefficient matrix. In addition, the two infinite series have different rates of convergence. Naturally, different number of terms in the truncation may be needed. Second, the domain of convergence for the two infinite series is not wide enough and excludes some practically interesting cases, including the case where the VAR coefficient matrix has a pure imaginary eigenvalue. More importantly, even if the true value of the VAR coefficient matrix is inside the domain of convergence, it is possible that the estimated value lies outside the domain of convergence. In the light of this concern, the expression with a wider domain of convergence is preferred. Third, it is not easy to impose prior knowledge of the model in the traditional expression of the ML estimator. Take a cointegrated system as an example. Although cointegrating tests can help us determine the cointegrating rank, it is not clear how to incorporate the knowledge of cointegration relationships in the traditional expression.

The paper proposes a new representation of the ML estimator of the mean reversion matrix. It can overcome the aforementioned drawbacks of the traditional representation. Firstly, the new representation is a finite polynomial series of the estimated VAR coefficient matrix. Hence, the asymptotic covariance matrix also only involves finite polynomial series. Consequently, no truncation is needed to compute the estimator and the covariance. Secondly, the new representation possesses a larger domain of convergence than the traditional method. Thirdly, prior knowledge such as cointegrating rank can be utilized straightforwardly by the new representation to simplify the estimation procedure.

Based on the new representation of the ML estimator of the mean reversion matrix, the paper establishes the long time span asymptotic theory for the cases covering stationary, pure unit root and partially nonstationary models. Special attention is paid to get explicit expression of the asymptotic covariance matrix for low dimensional cases.

The theory in the paper 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. For continuous time models with a complicated diffusion function, an approximate discrete time model may be derived from the continuous time model. One way to get an approximate discrete time model was proposed by Nowman (1997). The ML estimator of the mean reversion matrix based on the approximate discrete time model of Nowman is also the principal logarithm of the estimated VAR coefficient matrix. Therefore, the new representation of the ML estimator proposed in the paper is applicable in this context.

Phillips (1972) used least squares to estimate a 3-dimensional structural continuous time model where the mean reversion matrix depends on a set of structural parameters. He also 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 what we consider here in the sense that we estimate the full mean reversion matrix. Also, we do not assume that the derivative of the mean reversion matrix with respect to the VAR coefficient matrix is known. Moreover, Phillips (1972) imposed the assumption of stationarity, while we allow the model to be potentially nonstationary. In the context of univariate diffusion, Aït-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 expanding time span and a shrinking 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 introduces the model considered in the paper and some preliminaries. Section 3 describes how to obtain the ML estimator of the mean reversion matrix in the continuous time model. Both the traditional representation and the new proposed representation are studied and compared. The properties of the new representation are also provided in this section. Based on the new representation, Section 4 derives the limit theory of the ML estimator of the mean reversion matrix. In Section 5, an empirical study based on an affine term structure model is conducted to illustrate the implementation and the advantage of the new representation. Section 6 concludes. Proofs of the propositions and theorems are collected in the Appendix.

2 The Model and Some Preliminaries

The paper considers an m-dimensional multivariate diffusion process of the form:

$$dX(t) = (AX(t) + b)\,dt + \Sigma^{1/2} dW(t), \tag{2.1}$$

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 an m-dimensional standard Brownian motion. It is assumed that the matrix $\Sigma = \left[\Sigma^{1/2}\right]\left[\Sigma^{1/2}\right]'$ is positive definite. This process has been widely used to model interest rates in the term structure literature, following the univariate version that 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 may be very small, corresponding to high-frequency data. The sample size T can also be written as $T = N/h$ by letting N denote the time span of data. In this paper, we use $X(t)$ to represent a continuous time process and X_{th} to represent a discrete time process. When there is no confusion, we simply write X_{th} as X_t.

Bergstrom (1990) provided arguments as to why it is useful for macro-economists and policy makers 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 asset pricing theory.

The exact discrete time representation of (2.1) is

$$X_t = FX_{t-1} + g + \varepsilon_t, \tag{2.2}$$

where $F = e^{Ah}$ with the matrix exponential definition $e^{Ah} = \sum_{j=0}^{\infty} \frac{1}{j!} (Ah)^j$, $g = \int_0^h e^{As} b ds$, and $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{mt})'$ is a Gaussian martingale difference sequence (MDS) with respect to the natural filtration with

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

This is just a first order VAR model with $\text{MDS}(0, \Omega)$ innovations.

To simulate data from exact discrete time model (2.2), replacing the integration representations of g and Ω by more explicit expressions is quite necessary. As it will be clear later, as the rank of A varies, the matrices F, g and Ω can be expressed explicitly in different ways. Moreover, the explicit expressions facilitate the study of the process X_t, especially when $0 < \text{rank}(A) < m$.

In the case where matrix A is nonsingular, i.e., $\text{rank}(A) = m$, the discrete model (2.2) can be rewritten as

$$X_t = e^{Ah} X_{t-1} + A^{-1}\left[e^{Ah} - I_m\right] b + \varepsilon_t, \tag{2.3}$$

where I_m denotes $m \times m$ identity matrix. Phillips (1973) deduced the relationships

$$e^{Ah} \Sigma e^{A'h} - \Sigma = A\Omega + \Omega A',$$

and

$$Vec\left(\Omega\right) = \left\{A \otimes I_m + I_m \otimes A\right\}^{-1} \left\{e^{Ah} \otimes e^{Ah} - I_m \otimes I_m\right\} Vec\left(\Sigma\right),$$

where $Vec(\cdot)$ represents the vector formed by taking the direct sum of the rows of a matrix, and $\otimes$ denotes the right hand Kronecker product.

One special case in which A is nonsingular is obtained by assuming that all the eigenvalues of A have negative real parts. This assumption is commonly used to ensure the discrete time model (2.2) to be covariance stationary, because all the eigenvalues of $F = \exp(Ah)$ would have modulus less than 1. In this case, A is known as the mean reversion matrix for the reason that the magnitude of its eigenvalues measures the reverting speed of the process X_t to the long run mean $-A^{-1}b$.

For the case where $\text{rank}(A) = 0$, i.e., $A = 0_{m\times m}$, the discrete model (2.2) is equivalent to

$$X_t = X_{t-1} + bh + \varepsilon_t, \tag{2.4}$$

with $\Omega = \Sigma h$. Therefore, we get a pure unit root process without cointegration.

An interesting case corresponds to $0 < \text{rank}(A) < m$. Let A has a reduced rank τ and can be decomposed as

$$A = \alpha\beta',$$

where both α and β are full rank matrices of dimension $m\times\tau$. Since we wish to estimate A, not α or β, there is no need to do any normalization on α or β to ensure the uniqueness of the decomposition. Let $\alpha_\perp$ and $\beta_\perp$ be the orthogonal complementary matrices of α and β. Hence, $\alpha_\perp$ and $\beta_\perp$ are full rank matrices of dimension $m \times (m - \tau)$ with the properties of $\alpha'\alpha_\perp = 0$ and $\beta'\beta_\perp = 0$. Applying Ito's lemma to model (2.1), we get

$$d\left(\alpha'_\perp X(t)\right) = \left(\alpha'_\perp b\right) dt + \alpha'_\perp \Sigma^{1/2} dW(t). \tag{2.5}$$

Obviously, $\alpha'_\perp X(t)$ is an $(m-\tau)\times 1$ dimensional unit root process with no cointegration. Together with the restriction of $0 < \text{rank}(A) < m$, another assumption is always made in the literature that all the eigenvalues of $\beta'\alpha$ have negative real parts. As a result, the $\tau \times 1$ dimensional process $\beta' X(t)$ follows

$$d\left(\beta' X(t)\right) = \left[\left(\beta'\alpha\right)\beta' X(t) + \beta' b\right] dt + \beta'\Sigma^{1/2} dW(t), \tag{2.6}$$

which is stationary. Moreover, the non-singularity of $\beta'\alpha$ leads to

$$e^{Ah} = e^{(\alpha\beta')h} = I_m + \alpha\left[e^{(\beta'\alpha)h} - I_\tau\right]\left(\beta'\alpha\right)^{-1}\beta' = I_m + \underline{\alpha}\beta',$$

where $\underline{\alpha} = \alpha\left[e^{(\beta'\alpha)h} - I_\tau\right]\left(\beta'\alpha\right)^{-1}$ is an $m \times \tau$ matrix with full rank, and I_τ is the $\tau\times\tau$ identity matrix. Consequently, the discrete time model (2.2) can be represented as

$$\Delta X_t = \underline{\alpha}\beta' X_{t-1} + g + \varepsilon_t, \tag{2.7}$$

where $\Delta X_t = X_t - X_{t-1}$. It takes the form of an error correction model (ECM). If each element of X_t is the first order unit root process, the ECM (2.7) represents a cointegrated system, and the rows of β' make up a basis for the space of cointegrating relations. Granger (1981, 1983), Granger and Weiss (1983), Engle and Granger (1987) provided the early contributions to establish the relationship between cointegrated systems and ECMs. Studies on cointegration system are extensive; see for example, Stock (1987), Johansen (1988), Phillips and Ouliaris (1990), Sims, Stock and Watson (1990), Ahn and Reinsel (1990). Most studies focus on discrete time. Phillips (1991) formulated ECMs and cointegrated systems in continuous time models. His research uncovered the fact that the long-run information could be embedded in a continuous time model and for a given discrete time model the property of cointegration should be manifest itself in the same way if economic variables are sampled equidistantly, regardless of the rate at which they sampled (see also, Stock, 1987; Franchi, 2007; McCrorie, 2009).

In this paper, the multivariate diffusion process (2.1) with reduced rank matrix $A = \alpha\beta'$ is referred to as the partially nonstationary continuous time model. This is not a nonstationary process with cointegration, because neither the continuous time model nor its corresponding discrete time model (2.7) contains more than just the cointegrated system. For example, in the model with matrix

$$A = \begin{bmatrix} -0.1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \alpha\beta' \text{ and } \beta'\alpha = -0.1 < 0,$$

the first element of X_t is stationary while the second element is nonstationary.

By using an approach similar to Phillips (1973), we provide the following explicit expressions of g and Ω when A has a reduced rank:

$$g = \left\{ \underline{\alpha} \left(\beta'\alpha\right)^{-1} \beta' - \alpha \left(\beta'\alpha\right)^{-1} \beta' h + I_m h \right\} b = \vartheta\left(h\right) b, \tag{2.8}$$

and

$$\Omega = \alpha\Xi\alpha' - \underline{\alpha}\Upsilon\alpha' - \alpha\Upsilon\underline{\alpha}' + \alpha\Upsilon\alpha' h + \vartheta\left(h\right)\Sigma + \Sigma\left[\vartheta\left(h\right)\right]' - \Sigma h, \tag{2.9}$$

where $\Upsilon = \left(\beta'\alpha\right)^{-1} \beta'\Sigma\beta \left(\alpha'\beta\right)^{-1}$ and Ξ satisfies

$$Vec\left(\Xi\right) = \left\{ \left(\beta'\alpha\right) \otimes I_\tau + I_\tau \otimes \left(\beta'\alpha\right) \right\}^{-1} \left\{ e^{(\beta'\alpha)h} \otimes e^{(\beta'\alpha)h} - I_\tau \otimes I_\tau \right\} Vec\left(\Upsilon\right).$$

One equality deserves special attention (see the Appendix for the proof [1]), that is,

$$\alpha \left(\beta'\alpha\right)^{-1} \beta' + \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp = I_m. \tag{2.10}$$

Based on (2.10), the process X_t can be decomposed into an ergodic part and a Brownian motion,

$$dX\left(t\right) = \alpha \left(\beta'\alpha\right)^{-1} d\left[\beta' X\left(t\right)\right] + \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} d\left[\alpha'_\perp X\left(t\right)\right] \tag{2.11}$$

and

$$X_t = \alpha e^{(\beta'\alpha)h} \left(\beta'\alpha\right)^{-1} \left[\beta' X_{t-1}\right] + \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \left[\alpha'_\perp X_{t-1}\right] + g + \varepsilon_t, \tag{2.12}$$

[1] The author learned this proof from Peter Phillips in his class "Advanced Research Topics in Time Series Econometrics" at Singapore Management University in 2011.

where $\beta' X_{t-1}$ is stationary and $\alpha'_{\perp} X_{t-1}$ is a pure unit root process with no cointegration. Similar results can be found in Comte (1999), and Kessler and Anders (2004). As if will become clear later, this decomposition is essential to the development of an asymptotic theory for the estimation of A when it comes to the partially nonstationary continuous time model.

3 Estimating The Mean Reversion Matrix

Despite the fact that expressions of g and Ω varies according to different rank conditions of A, a common form of the discrete time model corresponding to the diffusion process (2.1) is given by

$$X_t = F X_{t-1} + g + \varepsilon_t,$$

where ε_t follows a Gaussian MDS$(0, \Omega)$.

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

$$\left[\hat{F}, \hat{g}\right] = \left[\sum_{t=1}^{T} X_t Z'_{t-1}\right] \times \left[\sum_{t=1}^{T} Z_{t-1} Z'_{t-1}\right]^{-1}. \tag{3.1}$$

If we have a prior knowledge that $b = 0$ and hence $g = 0$, the LS estimator changes to be

$$\hat{F} = \left[\sum_{t=1}^{T} X_t X'_{t-1}\right] \times \left[\sum_{t=1}^{T} X_{t-1} X'_{t-1}\right]^{-1}. \tag{3.2}$$

For the model considered in the paper, the simplest LS estimation is equivalent to the maximum likelihood (ML) estimation and the generalized least square (GLS) estimation (see, e.g., Zellner and Theil, 1962).

The well-known aliasing problem gives rise to the first difficulty in estimating A from $\hat{F}$. The basic idea of the aliasing problem is that the pair (A, Σ) is unidentifiable in (F, Ω), because the correspondence between them is not bijective. In the literature, many researchers tend to place additional restrictions on models under different settings to achieve identification. For the stationary diffusion model in which A has distinct characteristic roots, Phillips (1973) showed that (A, Σ) is identifiable if and only if the matrix A is identifiable in $F = \exp\{Ah\}$. However, many different matrices share the same exponential F. In particular, if some of the eigenvalues of A are complex, then by adding to each pair of conjugate complex eigenvalues the imaginary numbers $2ik\pi/h$ and $-2ik\pi/h$ for any integer k, another matrix A_1 satisfying $\exp\{A_1 h\} = F$ is obtained. To get a unique solution for A, 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 Ω provides extra identifying information about A, reducing the number of potential aliases. McCrorie

(2003), Kessler and Rahbek (2004) studied the identification problem for nonstationary continuous diffusion processes with cointegration.

We use a simple restriction which works well to achieve identification in all cases. Assumption 1 provides a desirable restriction by concentrating Ah on the unique principal logarithm of F in any possible situation, and the support implied by the assumption is typically quite wide and covers empirically relevant cases especially when h is small.

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

Proposition 3.1 *Under Assumption 1, F has no eigenvalue on closed negative real axis, namely,*

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

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

When F has no eigenvalue on the closed negative real axis, F has a unique logarithm with eigenvalues in the open strip $\{z \in \mathbb{C}, -\pi < \mathrm{Im}(z) < \pi\}$ of the complex plane. This is a well-known result in the linear algebra literature (see, e.g., Bernstein, 2009, p.721). The unique logarithm, denoted by $\ln(F)$, is called the principal logarithm. Under Assumption 1, Ah is the principal logarithm of F, namely, $A = \frac{1}{h}\ln(F)$, which naturally leads to the estimation of

$$\hat{A} = \frac{1}{h}\ln(\hat{F}). \tag{3.3}$$

3.1 Traditional Estimator and Its Shortcomings

When all the eigenvalues of $(I - F)$ have modulus less than 1, there is a widely known relationship between matrix F and its principal logarithm which takes the form of

$$Ah = \ln F = -\sum_{j=1}^{\infty} \frac{1}{j} (I - F)^j . \tag{3.4}$$

Given $\hat{F}$, the above representation leads to a frequently used estimator of A as in

$$\tilde{A} = \frac{1}{h}\ln \hat{F} = -\frac{1}{h}\sum_{j=1}^{\infty} \frac{1}{j}\left(I - \hat{F}\right)^j, \tag{3.5}$$

and

$$Vec\left(\tilde{A} - A\right) = \frac{1}{h}\left\{\sum_{j=1}^{\infty}\frac{1}{j}\left[\sum_{s=0}^{j-1}(I - F)^s \otimes \left(I - \hat{F}'\right)^{j-1-s}\right]\right\} Vec\left(\hat{F} - F\right). \tag{3.6}$$

As $\tilde{A}$ is a measurable transformation of $\hat{F}$, (3.5) and (3.6) suggest that one can apply standard results, such as the delta method, to obtain the limit theory of $\tilde{A}$ once the limit theory of $\hat{F}$ is known, and the matrix

$$K = \sum_{j=1}^{\infty} \frac{1}{j} \left[\sum_{s=0}^{j-1} (I - F)^s \otimes \left(I - F'\right)^{j-1-s} \right] \tag{3.7}$$

can be uses in the sandwich form to obtain the asymptotic covariance of $\tilde{A} - A$. The estimator of K, denoted by $\hat{K}$, is constructed by replacing F with $\hat{F}$, ie,

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

Although the traditional estimator works well sometimes, it suffers from three drawbacks. Firstly, an infinite summation is involved in $\tilde{A}$ as well as in $\hat{K}$. Hence, the calculation of $\tilde{A}$ and its asymptotic covariance matrix requires the truncation of infinite sequences in practice. Not surprisingly, the number of terms in truncation for $\tilde{A}$ and $\hat{K}$ could be different since they have different rates of convergence. In general, the truncation rule should depend on the eigenvalues of $\hat{F}$ as they determine the rate of convergence in $\tilde{A}$ and $\hat{K}$. Unfortunately, it is not clear how to truncate $\tilde{A}$ and $\hat{K}$. If too few terms are used, the estimation error caused by truncation would be large. If two many terms are used, not only the computational cost increases, but also, the estimation can get worse, as indicated in Table 1 and Table 2. The reason is that the infinite series may not converge. Another important aspect that is worth pointing out is that once the truncation is done, $\tilde{A}$ becomes inconsistent. The same argument applies to $\hat{K}$.

Secondly, the condition that all the eigenvalues of $I - F$ have modulus less than 1 determines the domain of convergence of A in (3.4) as well as K in (3.7). When the condition is violated, the power series representations of A and K are undefined. This condition excludes many interesting cases in practice, such as the case in which F has purely imaginary eigenvalues. Moreover, even if the true value of F satisfies the condition, $I - \hat{F}$ may have eigenvalues whose modulus are bigger than 1. As a result, the power series representations of $\tilde{A}$ and $\hat{K}$ are undefined. When this happens, not surprisingly, the $\tilde{A}$ and $\hat{K}$ obtained from truncating the power series could be far away from the true value.

To support the arguments above, we examine two simple cases and report the results in Table 1 and Table 2. Table 1 compares the estimates of A using the traditional estimator (3.5) with different truncation numbers, and estimates using the proposed estimator (3.13). Table 2 shows the comparison of the estimates of K from $\hat{K}$ with different truncation numbers, and that from the proposed estimator (3.15). Several features are apparent. Firstly, truncation rules for $\tilde{A}$ and $\hat{K}$ clearly depend on the value of $\hat{F}$. When $\hat{F} = F$, the case in which F is perfectly estimated with no error, 500 is a good truncation number for $\tilde{A}$ to make an accurate estimation of A. While, for $\hat{F} = \hat{F}_1$, even 1000 does not seem good enough. Similar phenomenon appears in $\hat{K}$. Secondly, the truncation rules for $\tilde{A}$ and $\hat{K}$ are quite different. To get accurate estimation of K, much larger truncation number is needed comparing with what is required in estimation of A. Thirdly, when the fitted value $\hat{F}_2$ is used, for which the eigenvalues of $I - \hat{F}_2$ have modulus 1.0311, the estimates from $\tilde{A}$ and $\hat{K}$ are far away from the true values. The larger the truncation taken, the worse the estimates are. Fourthly, both for A and K, the proposed estimator in the paper works very well in all cases. The accuracy of the estimates from the proposed estimator trumps that from the traditional estimator even with truncation number 1000.

In practice, it is always able to test how many cointegrating relations a cointegrated system has. Based on the introduction of partially nonstationary model in Section 2, the knowledge of cointegrating rank can easily translate into the information about the rank of A. Then, knowledge like rank $(A) = \tau$ where $0 < \tau < m$ could be revealed. The third drawback of the traditional approach lies in the fact that it can not utilize the prior knowledge of rank (A) to simplify the estimating procedure of A and the corresponding asymptotic covariance matrix. However, the proposed estimator in this paper can take advantage of the prior knowledge of rank (A). Details are discussed in Section 3.3.

3.2 New Estimator

A new explicit formula for the principal logarithm was recently proposed in the linear algebra literature. For any matrix F, the new formula represents the principal logarithm as a polynomial in the matrix $I - F$ of finite order with integral formulae for the coefficients involving the coefficients of the characteristic polynomial of $I - F$. The proposed estimator of A is based on the new formula.

Lemma 3.1 *(Cardoso, 2005) Let $F \in \mathbb{R}^{m\times m}$, $\Im = \{\iota \in \mathbb{R}|\ spec\{I - (I - F)\iota\} \cap \mathbb{R}_0^- = \varnothing\}$. For all $\iota \in \Im$, we have*

$$\ln\left[I - (I - F)\iota\right] = f_1(\iota)I + f_2(\iota)\left(I - F\right) + \cdots + f_m(\iota)\left(I - F\right)^{m-1},$$

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

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

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

Remark 3.2 *The characteristic polynomial of $I - F$ takes the form*

$$P(z) = \det[zI - (I - F)] = z^m + C_1 z^{m-1} + \cdots + C_{m-1} z + C_m. \tag{3.8}$$

Table 1: Traditional Estimator vs. New Estimator of $\hat{A}$

	A_{11}	A_{12}	A_{21}	A_{22}
A's true value	0 +12.3604i	0	0	0 -12.3604i
$[\ln(F)]/h\_10$	0.9567 +12.3617i	0	0	0.9567 -12.3617i
$[\ln(F)]/h\_100$	0.0175 +12.3408i	0	0	0.0175 -12.3408i
$[\ln(F)]/h\_500$	-0.0000 +12.3604i	0	0	-0.0000 -12.3604i
$[\ln(F)]/h\_1000$	-0.0000 +12.3604i	0	0	-0.0000 -12.3604i
$[\ln(F)]/h\_new$	-0.0000 +12.3604i	0	0	-0.0000 -12.3604i
$[\ln(\hat{F}_1)]/h\_10$	1.1226 +12.6501i	0	0	1.1226 -12.6501i
$[\ln(\hat{F}_1)]/h\_100$	0.1158 +12.5612i	0	0	0.1158 -12.5612i
$[\ln(\hat{F}_1)]/h\_500$	-0.0087 +12.5808i	0	0	-0.0087 -12.5808i
$[\ln(\hat{F}_1)]/h\_1000$	0.0087 +12.5607i	0	0	0.0087 -12.5607i
$[\ln(\hat{F}_1)]/h\_new$	-0.0000 +12.5622i	0	0	-0.0000 -12.5622i
$[\ln(\hat{F}_2)]/h\_10$	1.5392 +13.3835i	0	0	1.5392 -13.3835i
$[\ln(\hat{F}_2)]/h\_100$	-0.5891 +15.5681i	0	0	-0.5891 -15.5681i
$[\ln(\hat{F}_2)]/h\_500$	$10^5\times$(0.1219 - 1.1026i)	0	0	$10^5\times$(0.1219 + 1.1026i)
$[\ln(\hat{F}_2)]/h\_1000$	$10^{11}\times$(1.7144 - 1.8067i)	0	0	$10^{11}\times$(1.7144 +1.8067i)
$[\ln(\hat{F}_2)]/h\_new$	-0.0000 +12.9997i	0	0	-0.0000 -12.9997i

Note: Assume monthly data is used, i.e., h=1/12, and i is the imaginary unit. The true value of F=exp{Ah} is Vec(F)=$\big($0.5148+0.8573i, 0, 0, 0.5148-0.8573i$\big)'$. Two reasonable estimates are Vec($\hat{\text{F}}_1$)=$\big($0.5003+0.8659i, 0, 0, 0.5003-0.8659i$\big)'$ and Vec($\hat{\text{F}}_2$)= $\big($0.4684+0.8835i, 0, 0, 0.4684-0.8835i$\big)'$, respectively. The eigenvalues of (I-F), (I-$\hat{\text{F}}_1$) and (I-$\hat{\text{F}}_2$) have modulus 0.9851, 0.9997 and 1.0311, respectively. $[\ln(\cdot)]$/h_s, s=10,100,500,1000, are estimations of A by using formula (3.5) with truncation number 10, 100, 500, 1000, respectively. $[\ln(\cdot)]$/h_new denotes the estimation of A from the proposed estimator (3.13).

Table 2: Traditional Estimator vs. New Estimator of $AsyVar(\hat{A})$

	K_{11}	K_{22}	K_{33}	K_{44}
true value	0.5148 - 0.8573i	1.2015	1.2015	0.5148 + 0.8573i
$\hat{K}(F)$_10	0.0344 - 1.5713i	1.2016	1.2016	0.0344 + 1.5713i
$\hat{K}(F)$_100	0.2967 - 0.9029i	1.1996	1.1996	0.2967 + 0.9029i
$\hat{K}(F)$_500	0.5146 - 0.8568i	1.2015	1.2015	0.5146 + 0.8568i
$\hat{K}(F)$_1000	0.5148 - 0.8573i	1.2015	1.2015	0.5148 + 0.8573i
$\hat{K}(F)$_new	0.5148 - 0.8573i	1.2015	1.2015	0.5148 + 0.8573i
$\hat{K}(\hat{F}_1)$_10	0.0006 - 1.7286i	1.2175	1.2175	0.0006 + 1.7286i
$\hat{K}(\hat{F}_1)$_100	0.0009 - 1.6977i	1.2089	1.2089	0.0009 + 1.6977i
$\hat{K}(\hat{F}_1)$_500	1.3568 - 0.9405i	1.2108	1.2108	1.3568 + 0.9405i
$\hat{K}(\hat{F}_1)$_1000	0.0254 - 1.4323i	1.2089	1.2089	0.0254 + 1.4323i
$\hat{K}(\hat{F}_1)$_new	0.5003 - 0.8659i	1.2090	1.2090	0.5003 + 0.8659i
$\hat{K}(\hat{F}_2)$_10	-0.0344 - 2.1455i	1.2623	1.2623	-0.0344 + 2.1455i
$\hat{K}(\hat{F}_2)$_100	20.7477 - 7.7164i	1.4684	1.4684	20.7477 + 7.7164i
$\hat{K}(\hat{F}_2)$_500	$10^6\times$(-4.0729 + 1.8843i)	$O(10^4)$	$O(10^4)$	$10^6\times$(-4.0729 - 1.8843i)
$\hat{K}(\hat{F}_2)$_1000	$10^{13}\times$(-1.9668 - 0.4329i)	$O(10^{11})$	$O(10^{11})$	$10^{13}\times$(-1.9668 + 0.4329i)
$\hat{K}(\hat{F}_2)$_new	0.4684 - 0.8835i	1.2261	1.2261	0.4684 + 0.8835i

Note:

1. Let AsyVar(Â) denotes the asymptotic variance of $\hat{A} - A$.

2. The asymptotic variance of the traditional estimator in (3.5) and the proposed estimator in (3.13) are KVar(F̂)K$'$ and ΓVar(F̂)Γ', respectively. Therefore, the accuracy of the estimations of K and Γ determines how accurate the estimations of KVar(F̂)K$'$ and ΓVar(F̂)Γ' could be, once the estimation of the common term Var(F̂) is obtained. When the traditional estimator is applicable, i.e., each eigenvalue of I-F has modulus less than 1, it is easy to see that K=Γ. Therefore, $\hat{\Gamma}$ as in (3.15) can be used to estimate K. K̂$(\cdot)$_s,s= 10,100,500,1000, are estimations of K by using K̂ with truncation number 10, 100, 500, 1000, respectively. K̂$(\cdot)$_new denotes the estimation of K from the proposed estimator (3.15).

3. Assume monthly data is used, i.e., h=1/12, and i is the imaginary unit. The true value of F=exp{Ah} is Vec(F)=$($0.5148+0.8573i, 0, 0, 0.5148-0.8573i$)'$. Two reasonable estimates are Vec($\hat{F}_1$)=$($0.5003+0.8659i, 0, 0, 0.5003-0.8659i$)'$and Vec($\hat{F}_2$)=$($0.4684+0.8835i, 0, 0, 0.4684-0.8835i$)'$, respectively. The eigenvalues of (I-F), (I-$\hat{F}_1$) and (I-$\hat{F}_2$) have modulus 0.9851, 0.9997 and 1.0311, respectively.

4. Let K_{jj}, for j = 1,2,3,4, denote diagonal elements of matrix K. For the particular form of F, $\hat{F}_1$ and $\hat{F}_2$ here, both the true value and the estimation values of other elements of matrix K are zero. Therefore, this table only reports the estimation results of K_{jj}, for j = 1,2,3,4.

The coefficients $\{C_j\}_{j=1}^m$, which can be obtained by calculating the determinant of the matrix $zI - (I-F)$, are given by

$$C_1 = (-1)\sum_{s=1}^{m}(1-\lambda_s) = -tr(I-F),$$

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

$$\ldots$$

$$C_m = (-1)^m \prod_{s=1}^{m}(1-\lambda_s) = (-1)^m \det(I-F),$$

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

Remark 3.3 *The condition $\iota \in \Im$ is critical. If $\iota \notin \Im$, it is possible that $1 + C_1 S + \cdots + C_m S^m = 0$ for some $S \in [0,\iota]$ which makes $f_j(\iota)$, for $j = 1, \cdots, m$, undefined.*

Based on Proposition 3.1, Assumption 1 ensures that 1 belongs to $\Im$. Letting $\iota = 1$, from Lemma (3.1) we have

$$Ah = \ln(F) = f_1 I + f_2 (I-F) + \cdots + f_m (I-F)^{m-1}, \tag{3.9}$$

where

$$f_1 = \int_0^1 \frac{C_m S^{m-1}}{1 + C_1 S + \cdots + C_m S^m} dS, \tag{3.10}$$

$$f_j = \int_0^1 \frac{-S^{j-2} - C_1 S^{j-1} - \cdots - C_{m-j} S^{m-2}}{1 + C_1 S + \cdots + C_m S^m} dS, \text{ for } j = 2, \cdots, m-1, \tag{3.11}$$

$$f_m = \int_0^1 \frac{-S^{m-2}}{1 + C_1 S + \cdots + C_m S^m} dS. \tag{3.12}$$

The proposed estimator of A takes the form of

$$\hat{A} = \frac{1}{h}\ln\left(\hat{F}\right) = \frac{1}{h}\left\{\hat{f}_1 I + \hat{f}_2\left(I-\hat{F}\right) + \cdots + \hat{f}_m\left(I-\hat{F}\right)^{m-1}\right\}, \tag{3.13}$$

where

$$\hat{f}_1 = \int_0^1 \frac{\hat{C}_m S^{m-1}}{1 + \hat{C}_1 S + \cdots + \hat{C}_m S^m} dS,$$

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

$$\hat{f}_m = \int_0^1 \frac{-S^{m-2}}{1 + \hat{C}_1 S + \cdots + \hat{C}_m S^m} dS,$$

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

$$\hat{C}_j = (-1)^j \sum_{1\le s_1 < s_2 \ldots < s_j \le m}\left(1-\hat{\lambda}_{s_1}\right)\cdots\left(1-\hat{\lambda}_{s_j}\right). \tag{3.14}$$

Note that the formulae (3.4) and (3.9) are two different expressions of the principal logarithm of F, whose domains of definition are $\{F:$ eigenvalues of $(I-F)$ have modulus less than unity$\}$ and $\left\{F : spec\{F\} \cap \mathbb{R}_0^- = \varnothing\right\}$, respectively. The fact that

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

indicates that formulae (3.9) is more generally applicable than formulae (3.4). When (3.4) holds, formulae (3.9) is equivalent to formulae (3.4). Replacing F by $\hat{F}$, the same argument applies to $\tilde{A}$ given in formula (3.5) and $\hat{A}$ given in formula (3.13).

The consistency of the proposed estimator in (3.13) is easy to establish under the condition of $\hat{F} \xrightarrow{p} F$. Note that eigenvalues under ordering (with any ordering rule) are continuous functions of the elements of a matrix. Hence, the eigenvalues of $\hat{F}$, $\left\{\lambda_s\left(\hat{F}\right)\right\}_{s=1}^m$, converge to those of F, $\{\lambda_s(F)\}_{s=1}^m$, in probability, as long as $\hat{F} \xrightarrow{p} F$. Since, for $j=1,\cdots,m$, $\hat{C}_j$ are continuous in $\left\{\lambda_s\left(\hat{F}\right)\right\}_{s=1}^m$ and $\hat{f}_j$ are continuous in $\left\{\hat{C}_j\right\}_{j=1}^m$, the consistency of $\hat{A}$ is established immediately. We collect these results in the following theorem.

Theorem 3.4 *Let $\hat{A}$ be defined in (3.13), that Assumption 1 holds, h is fixed and $T \to \infty$. If $\hat{F} \xrightarrow{p} F$, then*

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

In order to draw a clear link between the limiting distribution of $\hat{A}-A$ and that of $\hat{F}-F$, a simplified relationship between $\hat{A}-A$ and $\hat{F}-F$ is presented in the next corollary. Some new notations appear. 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, whose ij^{th} element is given by $(-1)^{k+j}\left|(\Psi)_{kj}\right|$, where $\left|(\Psi)_{kj}\right|$ is determinant of the matrix.

Corollary 3.5 *Suppose Assumption 1 holds, h is fixed and $T \to \infty$. If $\hat{F} \xrightarrow{p} F$, we have*
(a)

$$hVec\left(\hat{A}-A\right) = \tilde{\Gamma} Vec\left(\hat{F}-F\right) + o_p\left(Vec\left(\hat{F}-F\right)\right)$$

with

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

(b) $\tilde{\Gamma}$ *converges in probability to a nonsingular matrix* Γ *as*

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

where, for $j = 1, \cdots, m$, $F'_j = \begin{bmatrix} \frac{\partial f_j}{\partial C_m} & \frac{\partial f_j}{\partial C_{m-1}} & \cdots & \frac{\partial f_j}{\partial C_1} \end{bmatrix}$ *with* f_j *taking the forms given in formulae (3.10), (3.11) and (3.12),* $L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & m & \cdots & m^{m-1} \end{bmatrix}$ *is a nonsingular matrix, and* $H = \begin{bmatrix} [Vec(H_1)]' \\ \vdots \\ [Vec(H_m)]' \end{bmatrix}$ *with* $H_z = [adj\,(zI - (I - F))]'$ *for* $z = 1, \cdots, m$.

Remark 3.6 *Clearly,* Γ *is the matrix which is typically used in the sandwich form to get the asymptotic covariance matrix of* $\hat{A} - A$ *from that of* $\hat{F} - F$*. The consistency of* $\hat{F}$ *and* $\hat{C}_j$, $j = 1, \cdots, m$, *ensures that*

$$\hat{\Gamma} = \sum_{j=1}^{m} Vec\left[\left(I - \hat{F}\right)^{j-1}\right]\hat{F}'_j L^{-1}\hat{H} - \sum_{j=2}^{m}\sum_{s=0}^{j-2}\hat{f}_j\left\{\left(I - \hat{F}\right)^{s} \otimes \left[I - \hat{F}'\right]^{j-2-s}\right\} \tag{3.15}$$

is a consistent estimation of Γ*, where* $\hat{F}'_j$, $\hat{f}_j$ *and* $\hat{H}$ *are obtained from* F'_j, f_j *and* H *by replacing* $\{C_j\}_{j=1}^{m}$ *and* F *with* $\left\{\hat{C}_j\right\}_{j=1}^{m}$ *and* $\hat{F}$.

Remark 3.7 *When all the eigenvalues of* $I - \hat{F}$ *have modulus less than unity,* $\hat{A}$ *in (3.13) is equivalent to* $\tilde{A}$ *given in (3.5). With the assumption that* $\hat{F} \xrightarrow{p} F$*, it is easy to see that* Γ *in Corollary 3.5 is equivalent to* K *in formula (3.7), but involves a finite summation only.*

To help understand Corollary 3.5, we examine the special case of $m = 1$, the univariate case. When $m = 1$, the exact ML estimator of A is $\hat{A} = \frac{1}{h}\ln\left(\hat{F}\right)$, and the first order Taylor expansion

$$h\left(\hat{A} - A\right) = \frac{1}{F}\left(\hat{F} - F\right) + o_p\left(\hat{F} - F\right) \tag{3.16}$$

is usually used to derive the asymptotic properties. In the univariate set up, Assumption 1 is satisfied when A is restricted to take real values. Hence, the new estimator in (3.13) is applicable and takes the form

$$\hat{A} = \frac{1}{h}\hat{f}_1 = \frac{1}{h}\int_0^1 \frac{\hat{C}_1}{1 + \hat{C}_1 S}dS,$$

where $\hat{C}_1 = \hat{F} - 1$. Straightforward calculation gives

$$\int_0^1 \frac{\hat{C}_1}{1 + \hat{C}_1 S}dS = \int_0^1 \frac{\hat{F} - 1}{1 + \left(\hat{F} - 1\right)S}dS = \ln\left(\hat{F}\right).$$

Therefore, as expected, the new estimator is just the commonly used ML estimator. To get the leading term of $h\left(\hat{A} - A\right)$, we first take the first order Taylor expansion of $\hat{f}_1$ at the point $\hat{C}_1 = C_1$, which leads to

$$\hat{f}_1 - f_1 = \frac{1}{1 + C_1}\left(\hat{C}_1 - C_1\right) + o_p\left(\hat{C}_1 - C_1\right),$$

for the reason that $\frac{\partial \hat{f}_1}{\partial \hat{C}_1}\Big|_{\hat{C}_1=C_1} = \int_0^1 \frac{1}{(1+C_1 s)^2} ds = \frac{1}{1+C_1}$. We then take the first order Taylor expansion of $\hat{C}_1$ at the point of $\hat{F}=F$, which leads to

$$\hat{C}_1 - C_1 = \hat{F} - F.$$

Putting the results together, we obtain $\frac{1}{F}\left(\hat{F}-F\right)$ as the leading term of $h\left(\hat{A}-A\right)$, as revealed by (3.16). Applying the same idea to the m-dimensional case, we get the results in Theorem 3.5.

To calculate the asymptotic covariance matrix of $h\left(\hat{A}-A\right)$, we need to obtain a more explicit expression of Γ. When $m=1$, it is also easy to obtain that

$$\tilde{\Gamma} = \Gamma = 1/F.$$

Low dimensional models, especially when $m=2,3$, are empirically very relevant. In the following two corollaries, we provide a more explicit expression of Γ when $m=2,3$. The proofs are omitted, because expanding the formulae of F_j', L and H given in Corollary 3.5 gives the results immediately.

Corollary 3.8 *When $m=2$,*

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

where

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

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

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

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

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

Corollary 3.9 *When $m=3$,*

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

where

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

$$\xi_3 = \int_0^1 \frac{(1+C_1S+C_2S^2)S^2}{(1+C_1S+C_2S^2+C_3S^3)^2}dS, \quad \xi_4 = \int_0^1 \frac{-(C_2+C_3S)S^3}{(1+C_1S+C_2S^2+C_3S^3)^2}dS,$$

$$\xi_5 = \int_0^1 \frac{(1+C_1S)S^2}{(1+C_1S+C_2S^2+C_3S^3)^2}dS, \quad \xi_6 = \int_0^1 \frac{(1+C_1S)S^3}{(1+C_1S+C_2S^2+C_3S^3)^2}dS,$$

$$\xi_7 = \int_0^1 \frac{S^2}{(1+C_1S+C_2S^2+C_3S^3)^2}dS, \quad \xi_8 = \int_0^1 \frac{S^3}{(1+C_1S+C_2S^2+C_3S^3)^2}dS,$$

$$\xi_9 = \int_0^1 \frac{S^4}{(1+C_1S+C_2S^2+C_3S^3)^2}dS,$$

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

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

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

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

$$f_2 = \int_0^1 \frac{-1-C_1S}{1+C_1S+C_2S^2+C_3S^3}dS, \quad f_3 = \int_0^1 \frac{-S}{1+C_1S+C_2S^2+C_3S^3}dS,$$

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

In empirical applications, sometimes extra restrictions on matrix A are available. One example is that all the eigenvalues of A are known to be distinct. Consequently, $F = e^{Ah}$ is diagonalizable with distinct eigenvalues. It is difficult to incorporate this prior knowledge to the estimation of A. However, as it is shown in the next corollary, the representation of matrix Γ becomes much simpler under this extra restriction. Taking advantage of the new expression of Γ not only makes the estimation of Γ easier, but also generate extra efficiency in estimating the asymptotic covariance of $\hat{A} - A$.

Before reporting Corollary 3.10, we first need to introduce a specific ordering rule of eigenvalues and a specific normalization rule of eigenvectors, to make the eigen-decomposition unique. Firstly, we let F's eigenvalues $\{\lambda_1, \cdots, \lambda_m\}$ be ordered according to

$$\mathrm{Re}(\lambda_1) \geq \cdots \geq \mathrm{Re}(\lambda_m).$$

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

$$|\mathrm{Im}(\lambda_j)| \geq |\mathrm{Im}(\lambda_{j+1})|\,.$$

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

$$p_j' p_j = 1$$

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

$$F = P diag \{\lambda_1, \cdots, \lambda_m\} Q,$$

where $P = \begin{bmatrix} p_1 & \cdots & p_m \end{bmatrix}$, $Q = P^{-1}$. Then, the matrix A has the ordered eigenvalues as $\{\eta_1, \cdots, \eta_m\} = \frac{1}{h} \{\ln(\lambda_1), \cdots, \ln(\lambda_m)\}$, and the decomposition of $A = P diag \{\eta_1, \cdots, \eta_m\} Q$.

Corollary 3.10 *Under Assumption 1, $\hat{F} \xrightarrow{p} F$ when h is fixed and $T \to \infty$. If A is diagonalizable with distinct eigenvalues, the matrix Γ can be expressed as*

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

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 $(v,v)^{th}$ element with $v \neq k$ is equal to $(e^{\eta_v h} - e^{\eta_k h}) / [(\eta_v - \eta_k)h]$.

Remark 3.11 *Note that the ordered eigenvalues and eigenvectors under normalization are continuous functions of the elements of a matrix. Once $\hat{F} \xrightarrow{p} F$ is established, a consistent estimate of Γ is easy to get as*

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

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 eigenvectors associated with the corresponding eigenvalues, and $\hat{Q} = \hat{P}^{-1}$. Comparing with the $\hat{\Gamma}$ constructed in Remark (3.6), $\hat{\Gamma}$ here is easier to get. Therefore, some extra efficiency in estimating the asymptotic covariance of $\hat{A} - A$ is expected.

Under the condition that all the eigenvalues of A are distinct and $\hat{F} \xrightarrow{a.s.} F$, $\hat{F}$ can be decomposed as $\hat{F} = \hat{P} diag \left\{\hat{\lambda}_1, \cdots, \hat{\lambda}_m\right\} \hat{Q}$ when the sample size T is large enough. By taking the principal logarithm, $\hat{A}$ in (3.13) has the eigenvalues, $\{\hat{\eta}_1, \cdots, \hat{\eta}_m\} = \frac{1}{h} \left\{\ln\left(\hat{\lambda}_1\right), \cdots, \ln\left(\hat{\lambda}_m\right)\right\}$, and can be decomposed as $\hat{A} = \hat{P} diag \{\hat{\eta}_1, \cdots, \hat{\eta}_m\} \hat{Q}$. Corollary 3.12 below provides an explicit relationship between $\{\hat{\eta}_1, \cdots, \hat{\eta}_m\} - \{\eta_1, \cdots, \eta_m\}$ and $\hat{A} - A$, facilitating the derivation of the joint limit distribution of eigenvalues. It can be shown that the elements of G reported in Corollary 3.12 are the same as the coefficient of the partial derivatives of the eigenvalues of A with respect to the elements of A as given in Phillips (1982).

Corollary 3.12 *Assume A is diagonalizable with distinct eigenvalues, and $\hat{F} \xrightarrow{a.s.} F$ when h is fixed and $T \to \infty$. We then have*

$$\left[\hat{\eta}_1 - \eta_1, \quad \cdots, \quad \hat{\eta}_m - \eta_m\right]' = G \cdot Vec\left(\hat{A} - A\right) + o_p\left(Vec\left(\hat{A} - A\right)\right),$$

where G is a $m \times m^2$ 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.

3.3 New Estimator under the Ranking Condition

In practice, it is possible to have some prior information regarding the rank of A. Sometimes, economic theory suggests the rank; sometimes, econometric tests suggests the rank. Take a cointegrated system as an example. If a cointegrated system is characterized by $0 < \tau < m$ cointegrating relations and cointegration tests such as Johansen (1988, 1991) confirm it, then it is known that $I - F$ has rank τ. From the discussion in Section 2, it is easy to see that rank $(A) = \tau$. While it is difficult to incorporate this prior information into the traditional estimator, it is straightforward to do so in the new estimator, as will be shown in this section. To facilitate the discussion, we focus on a bivariate cointegrated system and then provide the results for the general case.

Example 3.13 *Consider a bivariate diffusion process $X_t = (x_{1t}, x_{2t})'$ taking the form of (2.1). Its exact discrete time representation is (2.2). Based on the formulae from (3.9) to (3.12), we get*

$$Ah = \ln(F) = f_1 I + f_2 (I - F),$$

where

$$f_1 = \int_0^1 \frac{C_2 S}{1 + C_1 S + C_2 S^2} dS, \quad f_2 = \int_0^1 \frac{-1}{1 + C_1 S + C_2 S^2} dS,$$

$C_1 = tr\,(F - I)$ *and* $C_2 = \det(I - F)$. *Given $\hat{F}$, we have*

$$\hat{A}h = \ln\left(\hat{F}\right) = \hat{f}_1 I + \hat{f}_2\left(I - \hat{F}\right),$$

with

$$\hat{f}_1 = \int_0^1 \frac{\hat{C}_2 S}{1 + \hat{C}_1 S + \hat{C}_2 S^2} dS, \quad \hat{f}_2 = \int_0^1 \frac{-1}{1 + \hat{C}_1 S + \hat{C}_2 S^2} dS,$$

$\hat{C}_1 = tr\left(\hat{F} - I\right)$ *and* $\hat{C}_2 = \det\left(I - \hat{F}\right)$. *The matrix Γ, which is used in the sandwich form to get the asymptotic covariance of $(\hat{A} - A)$, is presented in Corollary (3.8).*

If there exists the prior knowledge that both x_{1t} and x_{2t} are random walks but cointegrated, then rank $(A) =$ rank $(I - F) = 1$. Hence, $C_2 = 0$, $f_1 = 0$, and,

$$Ah = \ln(F) = f_2 (I - F),$$

where $f_2 = \int_0^1 \frac{-1}{1+C_1 S} ds$, *and* $C_1 = tr(F-I)$. *With the simplified expression of A, the new proposed estimator becomes*

$$\hat{A}^* h = \ln\left(\hat{F}\right) = \hat{f}_2^* \left(I - \hat{F}\right)$$

where

$$\hat{f}_2^* = \int_0^1 \frac{-1}{1+\hat{C}_1 S} dS \quad \text{and} \quad \hat{C}_1 = tr\left(\hat{F} - I\right).$$

Obviously, $\hat{A}^*$ *is simpler than* $\hat{A}$*, which is embodied not only by missing the estimation of* f_1 *but also by the simplified formula of* f_2*. When* $\hat{F} \xrightarrow{p} F$*, straightforward algebra gives that*

$$hVec\left(\hat{A}^* - A\right) = \Gamma^* Vec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right),$$

where

$$\Gamma^* = Vec\left[\varphi_3 (I - F)\right] \Delta_1 - f_2 I_4,$$

and φ_3*,* Δ_1*,* f_2*,* I_4 *are defined as in Corollary (3.8) with the condition that* $C_2 = 0$*. Comparing with* Γ *in Corollary (3.8),* Γ^* *takes a simpler representation and is easier to calculate.*

In general settings, prior ranking knowledge is assumed to be

$$\text{rank}\ (A) = \tau, \quad \text{where} \ \ 0 < \tau < m.$$

What is revealed by the prior ranking condition is that A has at least $m - \tau$ zero eigenvalues [2]. Therefore, $F = e^{Ah}$ has at least $m - r$ eigenvalues equaling 1. Note that C_j, $j = 1, \cdots, m$, can be expressed as

$$C_j = (-1)^j \sum_{1 \le s_1 < s_2 \ldots < s_j \le m} (1 - \lambda_{s_1}) \cdots \left(1 - \lambda_{s_j}\right),$$

where $\{\lambda_s\}_{s=1}^m$ is the eigenvalue of F and that $C_j = 0$ whenever $j > \tau$. Based on the formulae from (3.9) to (3.12), the principal logarithm of F can be rewritten as

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

where

$$f_j = \int_0^1 \frac{-S^{j-2} - C_1 S^{j-1} - \cdots - C_\tau S^{j-2+\tau}}{1 + C_1 S + \cdots + C_\tau S^\tau} dS, \quad \text{for} \quad 2 \le j < m - \tau,$$
$$f_j = \int_0^1 \frac{-S^{j-2} - C_1 S^{j-1} - \cdots - C_{m-j} S^{m-2}}{1 + C_1 S + \cdots + C_\tau S^\tau} dS, \quad \text{for} \quad m - \tau \le j \le m-1,$$
$$f_m = \int_0^1 \frac{-S^{m-2}}{1 + C_1 S + \cdots + C_\tau S^\tau} dS.$$

[2] Having $m - \tau$ zero eigenvalues is the most that can be guaranteed by rank $(A) = \tau$. But, the number of zero eigenvalues could be larger than $m - \tau$. For example, $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, whose rank equals 1, possesses 2 zero eigenvalues. Note that $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \alpha\beta'$ makes x_{1t} be a $I(2)$ process which is stationary after differencing two times. This case is not of much practical interest, and is always excluded from partially nonstationary continuous model by the restriction that all eigenvalues of $\beta'\alpha$ should have negative real parts.

Consequently, the new proposed estimator under prior ranking knowledge is defined as

$$\hat{A}^* = \frac{1}{h}\ln\left(\hat{F}\right) = \frac{1}{h}\left\{\hat{f}_2^*\left(I-\hat{F}\right) + \cdots + \hat{f}_m^*\left(I-\hat{F}\right)^{m-1}\right\}, \tag{3.17}$$

where

$$\hat{f}_j^* = \int_0^1 \frac{-S^{j-2} - \hat{C}_1 S^{j-1} - \cdots - \hat{C}_\tau S^{j-2+\tau}}{1+\hat{C}_1 S + \cdots + \hat{C}_\tau S^\tau} dS, \quad \text{for} \quad 2 \le j < m-\tau,$$
$$\hat{f}_j^* = \int_0^1 \frac{-S^{j-2} - \hat{C}_1 S^{j-1} - \cdots - \hat{C}_{m-j} S^{m-2}}{1+\hat{C}_1 S + \cdots + \hat{C}_\tau S^\tau} dS, \quad \text{for} \quad m-\tau \le j \le m-1,$$
$$\hat{f}_m^* = \int_0^1 \frac{-S^{m-2}}{1+\hat{C}_1 S + \cdots + \hat{C}_\tau S^\tau} dS,$$

and $\hat{C}_j$, for $j = 1, \cdots, \tau$, are defined as in formula (3.14). Clearly, $\hat{A}^*$ is simpler than $\hat{A}$ given in formula (3.13) in which no prior ranking knowledge is incorporated. Missing estimates of f_1 and simplified estimating formula of f_j, for $j = 1, \cdots, m$, work together to fulfill the simplification of $\hat{A}^*$.

In the next corollary, a simplified relationship between $\hat{A}^* - A$ and $\hat{F} - F$ is derived to connect the limiting distribution of $\hat{A}^* - A$ to that of $\hat{F} - F$. Comparison between Γ^* in the next corollary and Γ in Corollary 3.5 indicates clearly that $\hat{A}^*$ has a simpler expression of the asymptotic covariance matrix.

Corollary 3.14 *Assume (1) the prior knowledge rank $(A) = \tau$ with $0 < \tau < m$ is given; (2) Assumption 1 holds; (3) h is fixed and $T \to \infty$. If $\hat{F} \xrightarrow{p} F$, then*
(a)

$$hVec\left(\hat{A}^* - A\right) = \tilde{\Gamma}^* Vec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right),$$

with

$$\tilde{\Gamma}^* = \sum_{j=2}^{m} Vec\left[\left(I-\hat{F}\right)^{j-1}\right] F_j^{*\prime}\Theta L^{-1} H - \sum_{j=2}^{m}\sum_{s=0}^{j-2} f_j \left\{(I-F)^s \otimes \left[I - \hat{F}'\right]^{j-2-s}\right\};$$

(b) $\tilde{\Gamma}^*$ converges in probability to a nonsingular matrix Γ^* as

$$\tilde{\Gamma}^* \xrightarrow{p} \Gamma^* = \sum_{j=2}^{m} Vec\left[(I-F)^{j-1}\right] F_j^{*\prime}\Theta L^{-1} H - \sum_{j=2}^{m}\sum_{s=0}^{j-2} f_j \left\{(I-F)^s \otimes \left[I-F'\right]^{j-2-s}\right\},$$

where $\Theta = \left[0_{\tau\times(m-\tau)}, I_\tau\right]_{\tau\times m}$, $F_j^{*\prime} = \left[\frac{\partial f_j}{\partial C_\tau} \quad \frac{\partial f_j}{\partial C_{\tau-1}} \quad \cdots \quad \frac{\partial f_j}{\partial C_1}\right]$, for $j = 2, \cdots, m$, f_j take the forms of formulae (3.10), (3.11) and (3.12) with condition that $C_j = 0$ whenever $j > \tau$, L and H are defined as in Corollary 3.5.

Remark 3.15 *Two differences between* Γ *in Corollary 3.5 and* Γ^* *in Corollary 3.14 deserve to be highlighted. Firstly,* $Vec\left[(I-F)^{j-1}\right]F_j'L^{-1}H$ *in* Γ *disappears in* Γ^**. The second difference is that* F_j' *in* Γ *is replaced by* $F_j^{*\prime}\Theta$ *in* Γ^**. As a result, there is no further need to calculate* $\partial f_j/\partial C_s$*, for* $s=\tau+1,\cdots,m$*,* $j=2,\cdots,m$*. These two differences allow* Γ^* *to enjoy a simpler expression relative to* Γ*, which in turn allows* $\hat{A}^*$ *to have a simpler expression of the asymptotic covariance matrix than* $\hat{A}$*.*

4 Asymptotics of New Estimator

4.1 Asymptotics for Stationary Model

This subsection develops the limit theory for $\hat{A}$ defined in (3.13) when the diffusion process $X(t)$ is stationary.

Assumption 2: All eigenvalues of A have negative real parts.

This commonly used condition makes all the eigenvalues of $F=\exp(Ah)$ have modulus less than 1, and consequently ensures the discrete time representation (2.2) of the diffusion process (2.1) to be a covariance stationary VAR(1) model. Note that A has full rank here and the discrete time model (2.2) can be rewritten as

$$X_t = e^{Ah}X_{t-1} + A^{-1}\left[e^{Ah}-I\right]b + \varepsilon_t = FX_{t-1} + g + \varepsilon_t,$$

where ε_t are MDS$(0,\Omega)$.

Under constant initial condition, the ML/LS estimator $\hat{F}$ defined in (3.1) has the following standard limit theory (see Hannan, 1970, p.329)[3].

Lemma 4.1 *If Assumption 2 holds, h is fixed and sample size T goes to infinity, then*

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

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

where $V_F=\Omega\otimes(V_X)^{-1}$, $V_X=Var(X_t)=\sum_{i=0}^{\infty}F^i\Omega F'^i$ *and* $\Omega=E(\varepsilon_t\varepsilon_t')$.

A direct application of the results in Corollary 3.5, Corollary 3.12 and Corollary 3.10 gives the results in the next theorem.

Theorem 4.2 *Let Assumption 1 and Assumption 2 hold,* $\hat{A}$ *is defined by (3.13) and,* $\{\eta_j\}_{j=1}^m$ *and* $\{\hat{\eta}_j\}_{j=1}^m$ *are the ordered eigenvalues of* A *and* $\hat{A}$*, respectively. If* h *is fixed and* $T\to\infty$*,*

[3] If $b=0$ and hence $g=0$ is prior knowledge, then $\hat{F}$ defined by formula (3.2) not (3.1) is going to be used to estimate F. It has the same asymptotics listed in Lemma 4.1. Therefore, the limit theory of $\hat{A}$ given by Theorem 4.2 still be true even in the case where $b=0$ is known a priori.

(a) we have:

$$h\sqrt{T}Vec\left(\hat{A}-A\right)\xrightarrow{d}N\left(0,\Gamma V_F\Gamma'\right).$$

(b) If A is diagonalizable with distinct eigenvalues, we have

$$h\sqrt{T}\left[\hat{\eta}_1-\eta_1,\quad\cdots,\quad\hat{\eta}_m-\eta_m\right]'\xrightarrow{d}N\left(0,G\Gamma V_F\Gamma'G'\right).$$

where G and V_F are defined in Corollary 3.12 and Corollary 4.1. In general, the matrix Γ takes the form reported in Corollary 3.5. When A is diagonalizable with distinct eigenvalues, Γ can be expressed by the simplified formula reported in Corollary 3.10.

4.2 Asymptotics for Pure Unit Roots Model

This subsection develops the limit theory of $\hat{A}$ defined in (3.13) for the model in which the rank of A is zero.

When rank $(A)=0$, $A=0_{m\times m}$. Consequently, the continuous time model (2.1) is

$$dX(t)=b\cdot dt+\Sigma^{1/2}dW(t),\tag{4.1}$$

whose exact discrete time representation is

$$X_t=X_{t-1}+bh+\varepsilon_t=FX_{t-1}+g+\varepsilon_t,\tag{4.2}$$

where $F=I$, $g=bh$, $\varepsilon_t=(\varepsilon_{1t},\cdots,\varepsilon_{mt})'$ is a Gaussian MDS $(0,\Omega=\Sigma h)$. Hence, X_t is a nonstationary process with no cointegration. We also use the ML/LS estimator of $[F,g]$ defined in formula (3.1) to estimate the model.

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

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t\Rightarrow B_0\left(r\right),$$

where $r\in[0,1]$, $B_0\left(r\right)$ is the 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 $T\to\infty$ with fixed h. For notational convenience, we often eliminate function arguments and write, for example, B_0 in place of $B_0\left(r\right)$ and $\int_0^1 B_0$ in place of $\int_0^1 B_0\left(r\right)dr$.

To facilitate the representation of the limit theory of $\left[\hat{F},\hat{g}\right]$, Park and Phillips (1988) introduced the functional

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

where B is the vector Brownian motion, M is a process with continuous sample paths such that $\int_0^1 MM'>0$ a.s., and R is a (possibly random) matrix of conformable dimension.

When $g=bh=0$, it is known that (e.g., Theorem 3.2 of Park and Phillips (1988))

$$T\left(\hat{F}-F\right)\Rightarrow f\left(B_0,B_0^*,\Delta_{20}\right), \tag{4.3}$$

where $B_0^*=B_0-\int_0^1 B_0$ and $\Delta_{20}=0_{m\times m}$.

When $g=bh\neq 0$, 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 $m\times m$ orthogonal matrix. We further define $\underline{B}_0=U_2'B_0$ and $\underline{\Delta}_{20}=U_2'\Delta_{20}=0_{m\times m}$. From Theorem 3.6 of Park and Phillips (1988) we get

$$T\left(\hat{F}-F\right)\Rightarrow f\left(B_0,\underline{B}_0^{**},\underline{\Delta}_{20}\right)U_2', \tag{4.4}$$

$$T^{3/2}\left(\hat{F}-F\right)\mu_1\Rightarrow\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right), \tag{4.5}$$

where $\underline{B}_0^{**}=\underline{B}_0-4\left(\int_0^1\underline{B}_0-(3/2)\int_0^1 s\underline{B}_0\right)+6r\left(\int_0^1\underline{B}_0-2\int_0^1 s\underline{B}_0\right)$, $\underline{\delta}=0_{1\times m}$, and $\underline{P}=r-1/2-\left(\int_0^1 s\underline{B}_0'-(1/2)\int_0^1\underline{B}_0'\right)\left(\int_0^1\underline{B}_0\underline{B}_0'-\int_0^1\underline{B}_0\int_0^1\underline{B}_0'\right)^{-1}\left(\underline{B}_0-\int_0^1\underline{B}_0\right)$.

By using the limit theory of the discrete time model reported above, the asymptotic distribution of $\hat{A}$ is obtained and reported in the following theorem.

Theorem 4.3 *Assume that $X(t)$ follows Model (4.1), and that $\hat{A}$ is defined as in (3.13) in which $\hat{F}$ is defined by (3.1). If h is fixed and $T\to\infty$, we have:*
(a) when $b=0$,

$$Th\left(\hat{A}-A\right)\xrightarrow{d}f\left(B_0,B_0^*,\Delta_{20}\right),$$

(b) when $b\neq 0$,

$$Th\left(\hat{A}-A\right)\xrightarrow{d}f\left(B_0,\underline{B}_0^{**},\underline{\Delta}_{20}\right)U_2',$$

and

$$T^{3/2}h\left(\hat{A}-A\right)\mu_1\xrightarrow{d}\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right),$$

where $f\left(B_0,B_0^*,\Delta_{20}\right)$, $f\left(B_0,\underline{B}_0^{**},\underline{\Delta}_{20}\right)U_2'$ and $\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right)$ are defined as in (4.3), (4.4) and (4.5), respectively, $g=bh$ and $\mu_1=g/\left(g'g\right)^{1/2}=b/\left(b'b\right)^{1/2}$.

Remark 4.4 *Consider the case where $b\neq 0$. Although $Th\left(\hat{A}-A\right)\xrightarrow{d}f\left(B_0,\underline{B}_0^{**},\underline{\Delta}_{20}\right)U_2'$ characterizes the asymptotic theory for each element of the matrix $h\left(\hat{A}-A\right)$, some particular column linear combinations like $h\left(\hat{A}-A\right)\mu_1$ may possess higher convergence rates, such as $T^{3/2}h\left(\hat{A}-A\right)\mu_1\xrightarrow{d}\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right)$.*

Remark 4.5 *For the case in which $m=1$, the results in Theorem 4.3 turn out to be:*

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

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

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

Remark 4.6 *If $A = 0_{m\times m}$ and $F = I_m$, but cointegrating tests suggest that rank $(I_m - F) =$ rank $(A) = \tau$ with $0 < \tau < m$, the estimator $\hat{A}^*$ defined in (3.17) is used. The same idea used to prove Theorem 4.3 can be applied to show that $\hat{A}^* - A$ and $\hat{A} - A$ share the same limit theory when $A = 0_{m\times m}$.*

Theorem 4.7 *Assume that $X(t)$ follows the model (4.1), that $\hat{A}$ is defined as (3.13) in which $\hat{F}$ is defined by (3.1), and that $\{\hat{\eta}_j\}_{j=1}^m$ are eigenvalues of $\hat{A}$. Let h is fixed and $T\to\infty$,*
(a) when $b = 0$,

$$Th\sum_{j=1}^{m}\hat{\eta}_j\xrightarrow{d}\Delta\cdot Vec\left[f\left(B_0,B_0^*,\Delta_{20}\right)\right],$$

(b) when $b\neq 0$,

$$Th\sum_{j=1}^{m}\hat{\eta}_j\xrightarrow{d}\Delta\cdot Vec\left[f\left(B_0,\underline{B}_0^{**},\underline{\Delta}_{20}\right)U_2'\right],$$

where Δ is a $1\times m^2$ row vector whose $1^{st},[m+2]^{th},\cdots,[(m-1)m+m]^{th}$ elements are 1, and 0 elsewhere, $f\left(B_0,B_0^*,\Delta_{20}\right)$, $f\left(B_0,\underline{B}_0^{**},\underline{\Delta}_{20}\right)U_2'$ are defined as in Theorem 4.3.

When $b = 0$ is known a priori, the discrete representation of the diffusion process (4.1) changes to an AR(1) model without drift. As a result, the estimator of

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

is used to estimate F. From Park and Phillips (1988), we have

$$T\left(\hat{F}-F\right)\xrightarrow{d}f\left(B_0,B_0,\Delta_{20}\right).$$

The approach used in this subsection can be easily applied to this simple case, and results similar to those reported in Theorem 4.3 and Theorem 4.7 can be obtained.

4.3 Asymptotics for Partially Non-stationary Model

In this subsection, we study the limit theory of $\hat{A}$ defined in (3.13) for the model where A has reduced rank τ and can be decomposed as $A = \alpha\beta'$, where α and β are $m\times\tau$ matrices with full column rank. Throughout the subsection, the following assumption is made.

Assumption 3: All the eigenvalues of $\left(\beta'\alpha\right)$ have negative real parts.

Under Assumption 3 and the condition of rank$(A) = \tau$, the process $X(t)$ defined in (2.1) is partially nonstationary and has two equivalent representations, namely, the exact discrete time representation

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

and an error correction representation

$$\Delta X_t = \underline{\alpha}\beta' X_{t-1} + g + \varepsilon_t,$$

where $F = \exp\{Ah\}$, $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{mt})'$ is a Gaussian MDS $(0, \Omega)$. The explicit expressions for $\underline{\alpha}$, g and Ω are also given in Section 2.

Let $Z_t = \left[X_t', 1\right]'$, the ML/LS estimator of $[F, g]$ defined in formula (3.1) as

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

is used to estimate the model. However, for the reasons that will be clear, it is difficult to establish the limit theory based on this expression.

From Section 2, X_t can be decomposed into an ergodic part and a Brownian motion

$$X_t = \Phi_1 Y_{1(t-1)} + \Phi_2 Y_{2(t-1)} + g + \varepsilon_t, \tag{4.6}$$

where $Y_{1t} = \beta' X_t$, $Y_{2t} = \alpha_\perp' X_t$, $\Phi_1 = \alpha e^{(\beta'\alpha)h} \left(\beta'\alpha\right)^{-1}$ and $\Phi_2 = \beta_\perp \left(\alpha_\perp' \beta_\perp\right)^{-1}$ which satisfies the equation of $\Phi_1 \beta' + \Phi_2 \alpha_\perp' = F$, and $\alpha_\perp$ and $\beta_\perp$ are the orthogonal complementary matrices of α and β, respectively. The exact discrete time representations of the diffusions (2.6) and (2.5) will give the dynamic functions of Y_{1t} and Y_{2t} as

$$Y_{1t} = e^{(\beta'\alpha)h} Y_{1(t-1)} + g_1 + \nu_{1t}, \tag{4.7}$$

with $g_1 = \left[e^{(\beta'\alpha)h} - I_\tau\right] \left(\beta'\alpha\right)^{-1} \beta' b$, $\quad \nu_{1t} = \int_{(t-1)h}^{th} e^{(\beta'\alpha)(th-s)} \beta' \Sigma^{1/2} dW(s)$, and

$$Y_{2t} = Y_{2(t-1)} + g_2 + \nu_{2t}, \tag{4.8}$$

with $g_2 = \alpha_\perp' bh$, $\quad \nu_{2t} = \int_{(t-1)h}^{th} \alpha_\perp' \Sigma^{1/2} dW(s)$. The LS estimator of $[\Phi_1, \Phi_2, g]$ is

$$\left[\hat{\Phi}_1, \hat{\Phi}_2, \hat{g}\right] = \left[\sum_{t=1}^{T} X_t \widetilde{Z}_{t-1}'\right] \times \left[\sum_{t=1}^{T} \widetilde{Z}_{t-1} \widetilde{Z}_{t-1}'\right]^{-1},$$

where

$$\widetilde{Z}_t = \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ 1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \beta' \\ \alpha_\perp' \end{bmatrix} & \\ & 1 \end{bmatrix} \begin{bmatrix} X_t \\ 1 \end{bmatrix} = \Pi Z_t.$$

From the non-singularity of the matrix $[\beta, \alpha_\perp]$ we get

$$\left[\hat{F} - F, \hat{g} - g\right] = \left[\hat{\Phi}_1 - \Phi_1, \hat{\Phi}_2 - \Phi_2, \hat{g} - g\right] \Pi,$$

and as a result,

$$\hat{F} - F = \left(\hat{\Phi}_1 - \Phi_1\right) \beta' + \left(\hat{\Phi}_2 - \Phi_2\right) \alpha'_\perp,$$

from which the limit theory of $\hat{F} - F$ can be derived using the limit theory of $\hat{\Phi}_1 - \Phi_1$ and $\hat{\Phi}_2 - \Phi_2$. It is known that $\hat{\Phi}_1 - \Phi_1$ and $\hat{\Phi}_2 - \Phi_2$ have different rates of convergence, and that their asymptotics depend on whether the intercept terms g_1 and g_2 appear in the processes of Y_{1t} and Y_{2t}. Therefore, not surprisingly, the limit theory of $\left[\hat{F}, \hat{g}\right]$, and consequently the corresponding limit theory of $\hat{A}$, vary across different cases.

- Case 1: $b = 0_{m\times 1}$ which implies X_t, Y_{1t} and Y_{2t} have no intercept term;
  - Case 1.1: $b = 0_{m\times 1}$ is known a priori;
  - Case 1.2: $b = 0_{m\times 1}$ is not a prior knowledge;
- Case 2: $b \neq 0_{m\times 1}$, $\alpha'_\perp b = 0_{(m-\tau)\times 1}$, $\beta' b \neq 0_{\tau\times 1}$ which only implies Y_{2t} have no intercept term;
- Case 3: $b \neq 0_{m\times 1}$, $\alpha'_\perp b \neq 0_{(m-\tau)\times 1}$, $\beta' b = 0_{\tau\times 1}$ which only implies Y_{1t} have no intercept term;
  - Case 3.1: $m - \tau = 1$;
  - Case 3.2: $m - \tau > 1$;
- Case 4: $b \neq 0_{m\times 1}$, $\alpha'_\perp b \neq 0_{(m-\tau)\times 1}$, $\beta' b \neq 0_{\tau\times 1}$ which implies each of X_t, Y_{1t}, and Y_{2t} includes an intercept term;
  - Case 4.1: $m - \tau = 1$;
  - Case 4.2: $m - \tau > 1$.

The condition $m - \tau = 1$ corresponds to a cointegrated system with $\tau = m - 1$ cointegrating relationships. The classification of situations above is complete by noticing the fact that the non-singularity of the matrix $[\beta, \alpha_\perp]$ ensures $b = 0_{m\times 1}$ is the unique condition to make neither Y_{1t} nor Y_{2t} have an intercept term.

To derive the limit theory of $\left[\hat{F}, \hat{g}\right]$ for Case 3.2 and Case 4.2 in which the pure unit root process Y_{2t} has dimension greater than 1 and a nonzero intercept term, the decomposition as in (4.6) is not good enough any more. Another decomposition on Y_{2t} is necessary. Let $d_1 = \left(g'_2 g_2\right)^{-1/2} g_2$ and D_2 be the orthogonal complementary matrix of d_1. Then, $D = [d_1, D_2]$ be an $(m-\tau) \times (m-\tau)$ orthogonal matrix. From formula (4.6), we get

$$X_t = \Phi_1 Y_{1(t-1)} + \Phi_{2d} Y_{2d(t-1)} + \Phi_{2D} Y_{2D(t-1)} + g + \varepsilon_t, \tag{4.9}$$

where $[\Phi_{2d}, \Phi_{2D}] = \Phi_2 (D')^{-1}$, $Y_{2dt} = d_1' Y_{2t}$ and $Y_{2Dt} = D_2' Y_{2t}$. The LS estimator takes the form of

$$\left[\hat{\Phi}_1, \hat{\Phi}_{2d}, \hat{\Phi}_{2D}, \hat{g}\right] = \left[\sum_{t=1}^{T} X_t Z_{t-1}^{\dagger\prime}\right] \times \left[\sum_{t=1}^{T} Z_{t-1}^{\dagger} Z_{t-1}^{\dagger\prime}\right]^{-1},$$

where

$$Z_t^{\dagger} = \begin{bmatrix} Y_{1t} \\ Y_{2dt} \\ Y_{2Dt} \\ 1 \end{bmatrix} = \begin{bmatrix} I_\tau & & \\ & D' & \\ & & 1 \end{bmatrix} \widetilde{Z}_t = \begin{bmatrix} I_\tau & & \\ & D' & \\ & & 1 \end{bmatrix} \Pi Z_t.$$

The nonsingularity of D and Π leads to

$$\left[\hat{F} - F, \hat{g} - g\right] = \left[\hat{\Phi}_1 - \Phi_1, \hat{\Phi}_{2d} - \Phi_{2d}, \hat{\Phi}_{2D} - \Phi_{2D}, \hat{g} - g\right] \begin{bmatrix} I_\tau & & \\ & D' & \\ & & 1 \end{bmatrix} \Pi.$$

Consequently, we have

$$\hat{F} - F = \left(\hat{\Phi}_1 - \Phi_1\right)\beta' + \left(\hat{\Phi}_{2d} - \Phi_{2d}\right) d_1' \alpha_\perp' + \left(\hat{\Phi}_{2D} - \Phi_{2D}\right) D_2' \alpha_\perp'.$$

The limit theory of $\hat{F} - F$ and the column linear combinations could be obtained based on the relationship above and the fact that

$$\begin{bmatrix} \left(\hat{F} - F\right)\alpha \\ \left(\hat{F} - F\right)\beta_\perp \left(\alpha_\perp' \beta_\perp\right)^{-1} d_1 \\ \left(\hat{F} - F\right)\beta_\perp \left(\alpha_\perp' \beta_\perp\right)^{-1} D_2 \end{bmatrix} = \begin{bmatrix} \left(\hat{\Phi}_1 - \Phi_1\right)\beta'\alpha \\ \left(\hat{\Phi}_{2d} - \Phi_{2d}\right) d_1' d_1 \\ \left(\hat{\Phi}_{2D} - \Phi_{2D}\right) D_2' D_2 \end{bmatrix}.$$

Lemma 4.8 provides the limit theory of $\left[\hat{\Phi}_1, \hat{\Phi}_{2d}, \hat{\Phi}_{2D}, \hat{g}\right]$ which plays an essential role in getting the limit theory of $\hat{F} - F$ in the most general case (case 4.2). In the remarks following it, we explain how the asymptotics of $\hat{F} - F$ for other cases can be derived as special cases of the results in Lemma 4.8.

Lemma 4.8 *Consider the diffusion process (2.1) in which the conditions in Case 4.2 are satisfied. When h is fixed and $T \to \infty$, the LS estimator of the regression function (4.9) has the limit theory as*

$$\begin{aligned} &\left[\sqrt{T}\left(\hat{\Phi}_1 - \Phi_1\right),\ T^{3/2}\left(\hat{\Phi}_{2d} - \Phi_{2d}\right),\ T\left(\hat{\Phi}_{2D} - \Phi_{2D}\right),\ \sqrt{T}\left(\hat{g} - g\right)\right] \\ \Rightarrow\ &\left[N\left(0, \Omega \otimes \left[\mu_{Y_1}\mu_{Y_1}' + V_{Y_1^0}\right]\right),\ (g_2' g_2)^{1/2} \int_0^1 r dB_0,\ \int_0^1 dB_0 \underline{B}_2',\ \int_0^1 dB_0\right] \\ &\times \begin{bmatrix} \mu_{Y_1}\mu_{Y_1}' + V_{Y_1^0} & \frac{1}{2}(g_2' g_2)^{-1/2}\mu_{Y_1} & \mu_{Y_1}\int_0^1 \underline{B}_2' & \mu_{Y_1} \\ \frac{1}{2}(g_2' g_2)^{-1/2}\mu_{Y_1}' & \frac{1}{3} g_2' g_2 & (g_2' g_2)^{1/2}\int_0^1 r\underline{B}_2' & \frac{1}{2}(g_2' g_2)^{1/2} \\ \int_0^1 \underline{B}_2 \mu_{Y_1}' & \int_0^1 r\underline{B}_2 (g_2' g_2)^{1/2} & \int_0^1 \underline{B}_2\underline{B}_2' & \int_0^1 \underline{B}_2 \\ \mu_{Y_1}' & \frac{1}{2}(g_2' g_2)^{1/2} & \int_0^1 \underline{B}_2' & 1 \end{bmatrix}^{-1}, \end{aligned}$$

where $\Omega = E(\varepsilon_t \varepsilon_t')$, $\mu_{Y_1} = -(\beta'\alpha)^{-1}\beta' b$ *which is the long run mean of the process* Y_{1t}, $V_{Y_1^0} = Var(Y_{1t}) = \sum_{j=0}^{\infty} e^{(\beta'\alpha)hj}\Omega_{\nu_1}e^{(\alpha'\beta)hj}$ *with* $\Omega_{\nu_1} = E(\nu_{1t}\nu_{1t}')$, *and, for any* $r \in [0,1]$, $T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} \nu_{2t} \Rightarrow B_2(r)$, $\underline{B}_2 = D_2' B_2$, $T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \Rightarrow B_0(r)$.

Remark 4.9 *It is clear that, by letting* $\mu_{Y_1} = 0_{\tau\times 1}$, *the results in Lemma 4.8 translate into the limit theory for the Case 3.2 immediately.*

Remark 4.10 *In Case 4.1,* $m - \tau = 1$ *makes* Y_{2t} *be a scalar unit root process with drift. Then,* d_1 *is a nonzero scalar who has no orthogonal complementary matrix. Hence,* D_2 *does not exist. The decomposition (4.9) degenerates to*

$$X_t = \Phi_1 Y_{1(t-1)} + \Phi_{2d} Y_{2d(t-1)} + g + \varepsilon_t.$$

The LS estimator of this regression model and $\hat{F} - F$ *is connected by*

$$\hat{F} - F = \left(\hat{\Phi}_1 - \Phi_1\right)\beta' + \left(\hat{\Phi}_{2d} - \Phi_{2d}\right) d_1' \alpha_\perp'.$$

Deleting the corresponding columns and rows in the results given in Lemma 4.8, the limit theory for Case 4.1 can be derived as

$$\begin{aligned}
& \left[\sqrt{T}\left(\hat{\Phi}_1 - \Phi_1\right), \quad T^{3/2}\left(\hat{\Phi}_{2d} - \Phi_{2d}\right), \quad \sqrt{T}\left(\hat{g} - g\right)\right] \\
\Rightarrow & \left[N\left(0, \Omega \otimes \left[\mu_{Y_1}\mu_{Y_1}' + V_{Y_1^0}\right]\right), \quad (g_2' g_2)^{1/2}\int_0^1 r dB_0, \quad \int_0^1 dB_0\right] \\
& \times \begin{bmatrix} \mu_{Y_1}\mu_{Y_1}' + V_{Y_1^0} & \frac{1}{2}(g_2'g_2)^{-1/2}\mu_{Y_1} & \mu_{Y_1} \\ \frac{1}{2}(g_2'g_2)^{-1/2}\mu_{Y_1}' & \frac{1}{3}g_2'g_2 & \frac{1}{2}(g_2'g_2)^{1/2} \\ \mu_{Y_1}' & \frac{1}{2}(g_2'g_2)^{1/2} & 1 \end{bmatrix}^{-1}.
\end{aligned}$$

Letting $\mu_{Y_1} = 0_{\tau\times 1}$, *the results above translate into the limit theory for Case 3.1, which can be expressed more explicitly as*

$$Vec\left[\sqrt{T}\left(\hat{\Phi}_1 - \Phi_1\right)\right] \Rightarrow N\left(0, \Omega \otimes V_{Y_1^0}^{-1}\right),$$

and

$$T^{3/2}\left(\hat{\Phi}_{2d} - \Phi_{2d}\right) \Rightarrow N\left(0, \Omega \cdot \frac{12}{(g_2)^2}\right).$$

Remark 4.11 *In Case 2,* $g_2 = \alpha_\perp' bh = 0_{(m-\tau)\times 1}$, *and* Y_{2t} *is nonstationary process with no drift. Then,* $I_{m-\tau}$ *is orthogonal complementary matrix of* $d_1 = 0_{(m-\tau)\times 1}$. *Hence,* $D = D_2 = I_{m-\tau}$, $Y_{2Dt} = D_2' Y_{2t} = Y_{2t}$ *and* $\Phi_{2D} = \Phi_2$. *The decomposition (4.9) degenerates to decomposition (4.6) as*

$$X_t = \Phi_1 Y_{1(t-1)} + \Phi_{2D} Y_{2D(t-1)} + g + \varepsilon_t = \Phi_1 Y_{1(t-1)} + \Phi_2 Y_{2(t-1)} + g + \varepsilon_t.$$

The LS estimator of this regression model and $\hat{F} - F$ *is connected by*

$$\hat{F} - F = \left(\hat{\Phi}_1 - \Phi_1\right)\beta' + \left(\hat{\Phi}_{2D} - \Phi_{2D}\right) D_2' \alpha_\perp'.$$

Note that $\underline{B}_2 = D_2' B_2 = B_2$. Then, deleting the corresponding columns and rows in the results given in Lemma 4.8, the limit theory for Case 2 can be derived as

$$\begin{aligned} & \left[\sqrt{T}\left(\hat{\Phi}_1 - \Phi_1\right), \quad T\left(\hat{\Phi}_{2D} - \Phi_{2D}\right), \quad \sqrt{T}\left(\hat{g} - g\right)\right] \\ \Rightarrow \quad & \left[N\left(0, \Omega \otimes \left[\mu_{Y_1}\mu_{Y_1}' + V_{Y_1^0}\right]\right), \quad \int_0^1 dB_0 B_2', \quad \int_0^1 dB_0\right] \\ & \times \begin{bmatrix} \mu_{Y_1}\mu_{Y_1}' + V_{Y_1^0} & \mu_{Y_1}\int_0^1 B_2' & \mu_{Y_1} \\ \int_0^1 B_2 \mu_{Y_1}' & \int_0^1 B_2 B_2' & \int_0^1 B_2 \\ \mu_{Y_1}' & \int_0^1 B_2' & 1 \end{bmatrix}^{-1}. \end{aligned}$$

Letting $\mu_{Y_1} = 0_{\tau\times 1}$, the results above translate into the limit theory for Case 1.2, which can be expressed separately as

$$Vec\left[\sqrt{T}\left(\hat{\Phi}_1 - \Phi_1\right)\right] \Rightarrow N\left(0, \Omega \otimes V_{Y_1^0}^{-1}\right),$$

and

$$T\left(\hat{\Phi}_{2D} - \Phi_{2D}\right) \Rightarrow f\left(B_0, B_2^*, \Delta_{20}\right),$$

where $B_2^* = B_2 - \int_0^1 B_2$ and $\Delta_{20} = 0_{(m-\tau)\times m}$.

Remark 4.12 *In Case 1.1, the regression model has no drift and takes the form of*

$$X_t = F X_{t-1} + \varepsilon_t,$$

with the LS estimator of

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

We continue to have the decomposition such as

$$X_t = \Phi_1 Y_{1(t-1)} + \Phi_{2D} Y_{2D(t-1)} + \varepsilon_t = \Phi_1 Y_{1(t-1)} + \Phi_2 Y_{2(t-1)} + \varepsilon_t,$$

and the relationship such as

$$\hat{F} - F = \left(\hat{\Phi}_1 - \Phi_1\right)\beta' + \left(\hat{\Phi}_{2D} - \Phi_{2D}\right) D_2' \alpha_\perp'.$$

with $D = D_2 = I_{m-\tau}$. Deleting the corresponding columns and rows in the results given in Lemma 4.8, we have

$$\left[\sqrt{T}\left(\hat{\Phi}_1 - \Phi_1\right), \quad T\left(\hat{\Phi}_{2D} - \Phi_{2D}\right)\right] \Rightarrow \left[N\left(0, \Omega \otimes V_{Y_1^0}\right), \quad \int_0^1 dB_0 B_2',\right] \times \begin{bmatrix} V_{Y_1^0} & \\ & \int_0^1 B_2 B_2' \end{bmatrix}^{-1}.$$

The results for Case 1.1 and Case 1.2 coincide with whose reported in Theorems 3.1 and 3.2 in Park and Phillips (1989).

Theorem 4.13 *Consider the diffusion process (2.1) in which the conditions in Case 4.2 are satisfied. When h is fixed and $T \to \infty$, $\hat{A}$ defined in (3.13) has the limit theory*
(a)

$$h\sqrt{T}\Gamma^{-1}Vec\left(\hat{A}-A\right) \overset{d}{\sim} \sqrt{T}\left(I_m \otimes \beta\right)Vec\left(\hat{\Phi}_1-\Phi_1\right);$$

(b)

$$h\sqrt{T}\left(I_m \otimes \alpha'\right)\Gamma^{-1}Vec\left(\hat{A}-A\right) \overset{d}{\sim} \sqrt{T}\left(I_m \otimes \left[\alpha'\beta\right]\right)Vec\left(\hat{\Phi}_1-\Phi_1\right);$$

(c)

$$hT\left(I_m \otimes \left[\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}D_2\right]'\right)\Gamma^{-1}Vec\left(\hat{A}-A\right) \overset{d}{\sim} T\left(I_m \otimes \left[D_2'D_2\right]\right)Vec\left(\hat{\Phi}_{2D}-\Phi_{2D}\right);$$

(d)

$$hT\left(I_m \otimes \left[\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}d_1\right]'\right)\Gamma^{-1}Vec\left(\hat{A}-A\right) \overset{d}{\sim} T^{3/2}\left(I_m \otimes \left[d_1'd_1\right]\right)Vec\left(\hat{\Phi}_{2d}-\Phi_{2d}\right),$$

where $\overset{d}{\sim}$ denotes asymptotic equivalence in distribution, Γ is defined as in Theorem 3.5. The limit theory of $\hat{\Phi}_1-\Phi_1$, $\hat{\Phi}_{2d}-\Phi_{2d}$ and $\hat{\Phi}_{2D}-\Phi_{2D}$ are reported in Lemma 4.8.

Remark 4.14 *Note that β is an $m\times\tau$ matrix whose rows are linearly dependent. Hence, $h\sqrt{T}\Gamma^{-1}Vec\left(\hat{A}-A\right)$ has a singular asymptotic covariance matrix. However, the fact that the $\tau\times\tau$ matrix $\alpha'\beta$ has the full rank ensures that the asymptotic covariance matrix of $h\sqrt{T}\left(I_m \otimes \alpha'\right)\Gamma^{-1}Vec\left(\hat{A}-A\right)$ is non-singular.*

Remark 4.15 *The remarks behind the Lemma 4.8 have demonstrated clearly how the limit theory of $\hat{\Phi}_1-\Phi_1$, $\hat{\Phi}_{2d}-\Phi_{2d}$ and $\hat{\Phi}_{2D}-\Phi_{2D}$ for the cases from Case 1.1 to Case4.1 can be derived and what they look like. Taking those asymptotic results to replace their counterparts in Theorem 4.13, the corresponding limit theory of $\hat{A}$ for the cases from Case 1.1 to Case 4.1 would be obtained straightforwardly. For example in Case 1.2, $d_1 = 0_{(m-\tau)\times 1}$ and $D_2 = I_{m-\tau}$, $\hat{A}$ possesses the limit theory*

$$h\sqrt{T}\Gamma^{-1}Vec\left(\hat{A}-A\right) \Rightarrow N\left(0, \Omega \otimes \left[\beta V_{Y_1^0}^{-1}\beta'\right]\right),$$

$$h\sqrt{T}\left(I_m \otimes \alpha'\right)\Gamma^{-1}Vec\left(\hat{A}-A\right) \Rightarrow N\left(0, \Omega \otimes \left[\left(\alpha'\beta\right)V_{Y_1^0}^{-1}\left(\beta'\alpha\right)\right]\right)$$

and

$$hT\left(I_m \otimes \left[\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}\right]'\right)\Gamma^{-1}Vec\left(\hat{A}-A\right) \Rightarrow Vec\left[f\left(B_0, B_2^*, \Delta_{20}\right)\right].$$

Remark 4.16 *The limit theory of the estimator $\hat{A}^*$ defined in (3.17) could be obtained immediately from the results in Theorem 4.13 by substituting Γ^* defined in Theorem 3.14 for Γ.*

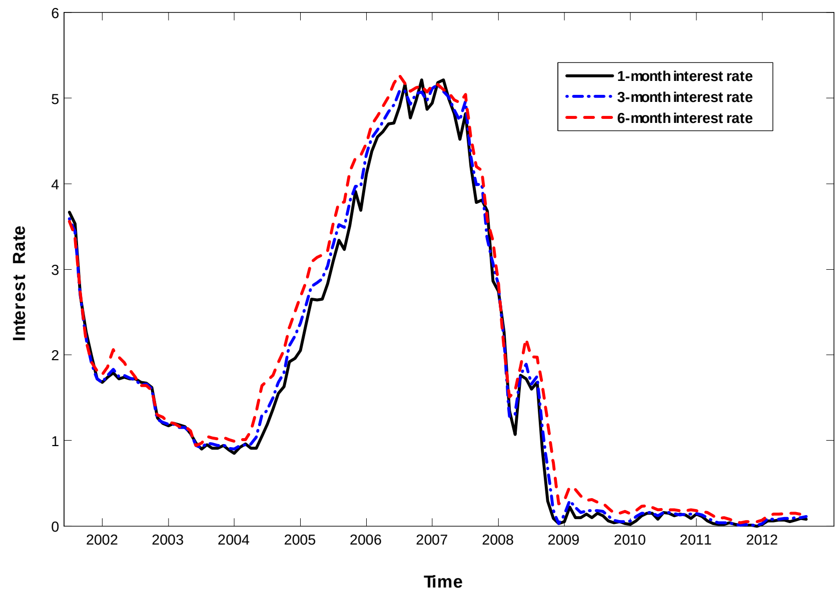

Figure 1: Time series plot of 1-month, 3-month and 6-month interest rates of U.S. Government Treasuries from July, 2009 to September, 2012.

5 An Empirical Illustration

To illustrate the implementation and the advantage of the proposed method, we estimate an affine term structure model of interest rates using the traditional method and the proposed method. The affine term structure model was introduced in Duffie and Kan (1996).

The observed data, $X_t = (X_{1t}, X_{2t}, X_{3t})'$, represent 1-month, 3-month and 6-month interest rates of U.S. Government Treasuries sampled monthly from July 2001 to September 2012 (i.e., $h = 1/12$ and $T = 135$). The time series plots of the interest rates data are given in Figure 1. The unit root test of Phillips and Perron (1988) suggests that the unit root hypothesis for each process X_{st}, for $s = 1, 2, 3$, cannot be rejected. The cointegrating rank tests of Johansen (1988) suggests that there is only one cointegrating relationship among the three interest rate series.

In this paper, we describe the dynamics of the process X_t by using a three factor Gaussian affine term structure model which takes the form of

$$dX(t) = [AX(t) + b]\, dt + \Sigma^{1/2} dW(t), \tag{5.1}$$

where A, b are 3×3 and 3×1 matrices. The exact discrete time representation is

$$X_t = FX_{t-1} + g + \varepsilon_t \quad \text{with} \quad F = e^{A/12}. \tag{5.2}$$

Table 3: Estimates of A in Affine Term Structure Model of Interest Rates

	$\hat{A}_{11}$	$\hat{A}_{12}$	$\hat{A}_{13}$	$\hat{A}_{21}$	$\hat{A}_{22}$	$\hat{A}_{23}$	$\hat{A}_{31}$	$\hat{A}_{32}$	$\hat{A}_{33}$
$\tilde{A}(\hat{F})$_10	-55.741	84.413	-29.194	-21.691	29.658	-8.307	-17.788	26.231	-8.675
$\tilde{A}(\hat{F})$_100	-73.011	113.192	-40.790	-28.563	41.132	-12.921	-23.340	35.484	-12.403
$\tilde{A}(\hat{F})$_1000	-73.205	113.515	-40.920	-28.640	41.266	-12.973	-23.403	35.587	-12.445
$\hat{A}(\hat{F})$_new	-73.205	113.515	-40.920	-28.640	41.266	-12.973	-23.403	35.587	-12.445

Note:
1. $\hat{F}$ is the ML/LS estimates of F in the regression model (5.2), which is the exact discrete time representation of the affine term structure model (5.1). The observed data, $X_t=(X_{1t}, X_{2t}, X_{3t})'$, represent 1-month, 3-month and 6-month interest rates of U.S. Government Treasuries sampled monthly from July 2001 to September 2012.
2. $\tilde{A}(\hat{F})$_j, j = 10,100,1000, denote the estimates of A by using $\tilde{A}$ in (3.5) with truncation number 10,100,1000, respectively. $\hat{A}(\hat{F})$_new indicates the estimate of A from the proposed estimator (3.13).

The ML/LS estimates of F is given by

$$\hat{F} = \begin{pmatrix} -0.6540 & 2.1349 & -0.5078 \\ -0.6221 & 1.5632 & 0.0401 \\ -0.5218 & 0.6200 & 0.8880 \end{pmatrix}.$$

The eigenvalues of $I - \hat{F}$ are $(0.0062, 0.2290, 0.9677)$, all have modulus less than unity. In this case, as discussed in Subsection 3.1 and 3.2, the new proposed representation of ML estimator, $\hat{A}$, given in formula (3.13) and the traditional expression of ML estimator, $\tilde{A}$, given in formula (3.5) are equivalent but take significantly different forms.

Table 3 reports $\hat{A}$ and $\tilde{A}$ with different truncation numbers, $10, 100, 1000$. It can be seen clearly that the truncation number affects the estimation results of $\tilde{A}$ in (3.5) significantly. If the truncation number picked is too small, say 10, the estimates could be far away from the true value. The new proposed representation of ML estimator $\hat{A}$ as in (3.13) needs no truncation and provides the exact value of $\frac{1}{h}\ln(\hat{F})$. Comparing the estimates from $\tilde{A}$ and $\hat{A}$, 1000 is a good truncation number to chose for $\tilde{A}$ in this case. However, as argued in Subsection 3.1, the truncation number depends on the value of $\hat{F}$, hence 1000 may not be good in other cases.

Assuming that all eigenvalues of $(I - F)$ have modulus less than unity, $\tilde{A} - A$ and $\hat{A} - A$ have the same limit theory. However, the procedures to estimate the asymptotic covariance matrices of $\tilde{A} - A$ and $\hat{A} - A$ are different. For $\tilde{A} - A$, the estimate of asymptotic covariance matrix depends on the estimation of K as in (3.4) which involves an infinite summation. For $\hat{A} - A$, it depends on the estimation of Γ whose formula is given in Corollary 3.9 which only includes a finite summation.

Before comparing the estimates of the asymptotic covariance matrices of $\tilde{A} - A$ and $\hat{A} - A$, we first run the regression

$$X_{1t} = \beta_2 X_{2t} + \beta_3 X_{3t} + \nu_{1t}$$

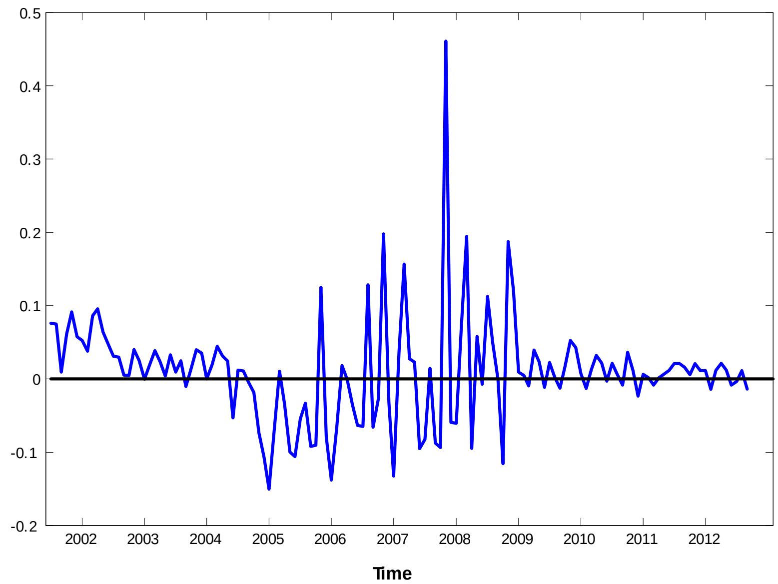

Figure 2: Time series plot of $X_{1t} - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}$ where X_{1t}, X_{2t} and X_{3t} are 1-month, 3-month and 6-month interest rates of U.S. Government Treasuries from July, 2009 to September, 2012, respectively.

to obtain a consistent estimation of the cointegrating relationship as

$$\hat{\beta} = (1, -\hat{\beta}_2, -\hat{\beta}_3) = (1, -1.5189, 0.5222). \tag{5.3}$$

The usual t-test for the sample mean of $X_{1t} - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}$ shows that no drift term appears in this cointegrated process, as suggested in Figure 2.

Then, from Theorem 4.8 and Theorem 4.13, it is obtained that the asymptotic distribution of the proposed estimator $\hat{A}$ given in (3.13) is

$$h\sqrt{T}Vec\left(\hat{A} - A\right) \Rightarrow N\left(0, \Gamma\left[\Omega \otimes \left(\beta V_{Y_1^0}^{-1} \beta'\right)\right] \Gamma'\right),$$

where $\Omega = E\left(\varepsilon_t \varepsilon_t'\right)$, $V_{Y_1^0} = Var\left(\beta' X_t\right)$, β is the cointegrating vector. From the formula (3.6) and Theorem 4.8, it is easy to see that

$$h\sqrt{T}Vec\left(\tilde{A} - A\right) \Rightarrow N\left(0, K\left[\Omega \otimes \left(\beta V_{Y_1^0}^{-1} \beta'\right)\right] K'\right),$$

where K takes the formula given in (3.7). As argued above, $\hat{A} - A$ and $\tilde{A} - A$ share the same limit theory

$$K\left[\Omega \otimes \left(\beta V_{Y_1^0}^{-1} \beta'\right)\right] K' = \Gamma\left[\Omega \otimes \left(\beta V_{Y_1^0}^{-1} \beta'\right)\right] \Gamma',$$

where K and Γ are equivalent but with different expressions.

We estimate the asymptotic covariance matrix as

$$\hat{K}\left[\hat{\Omega}\otimes\left(\hat{\beta}\hat{V}_{Y_1^0}^{-1}\hat{\beta}'\right)\right]\hat{K}'=\hat{\Gamma}\left[\hat{\Omega}\otimes\left(\hat{\beta}\hat{V}_{Y_1^0}^{-1}\hat{\beta}'\right)\right]\hat{\Gamma}',$$

where $\hat{K}$, $\hat{\Gamma}$ are obtained by letting $\hat{F}$ replace F in the formulae of K and Γ, respectively, $\hat{\beta}$ is given in (5.3),

$$\hat{\Omega}=\frac{1}{T-1}\sum_{t=2}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t' \quad \text{with } \hat{\varepsilon}_t=X_t-\hat{F}X_{t-1}-\hat{g}, \tag{5.4}$$

$$\text{and}\quad \hat{V}_{Y_1^0}=\frac{1}{T}\sum_{t=1}^{T}\hat{\beta}'X_t\left(\hat{\beta}'X_t\right)'. \tag{5.5}$$

Table 4 reports the estimates of the asymptotic covariance matrix using two approaches. One is based on $\hat{K}$ with different truncation numbers, 10, 100, 1000. The other is based on $\hat{\Gamma}$. It can be seen clearly that the estimation results of the asymptotic covariance matrix using $\hat{K}$ depends heavily on the truncation number. If the truncation number is picked too small, say 10, the estimates could be far away from the true value. The new proposed representation of ML estimator $\hat{A}$ as in (3.13) enables us to estimate the asymptotic covariance matrix through $\hat{\Gamma}$, the estimation of Γ given in Corollary 3.9. The new estimation procedure needs no truncation, hence gives an accurate estimate of $\Gamma = K$, and consequently, an accurate estimate of asymptotic covariance matrix. In Table 4, we find that, for some elements, the accuracy of the proposed estimator is superior to the traditional one even with truncation number 1000. The errors introduced in the traditional one with truncation number 10 and 100 are generally very large and should not be used.

6 Conclusion and Further Remarks

This paper provides a new representation 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. Comparing with the traditional representation of the ML/LS estimator, the new representation enjoys three major advantages. First, while the traditional matrix logarithmic representation involves an infinite polynomial series, the new representation involves only finite polynomial series whose degree is the same as the dimension of the multivariate process, facilitating the use of the delta method and the calculation of the covariance matrix in the limit distribution. Second, the new representation has a larger domain of convergence than the traditional method. Third, the new representation can use prior knowledge about the model such as the cointegrating rank to simplify the estimation procedure, whereas it is hard to do so in the traditional method.

For all the cases in which the multivariate diffusion process is stationary, pure unit root or partially nonstationary, the limit theory of the ML/LS estimator of the "mean reversion matrix" with the new representation is established by using the limit distribution of the estimated VAR coefficient

Table 4: Estimated Variance-Covariance matrix of $h\sqrt{T}Vec(\hat{A}-A)$

		dA_{11}	dA_{12}	dA_{13}	dA_{21}	dA_{22}	dA_{23}	dA_{31}	dA_{32}	dA_{33}
dA_{11}	·_10	149.33								
	·_100	2200.7								
	·_1000	2396.3								
	·_new	2395.7								
dA_{12}	·_10	-240.86	388.48							
	·_100	-3628.4	5982.4							
	·_1000	-3952.7	6519.9							
	·_new	-3952.2	6519.9							
dA_{13}	·_10	92.21	-148.73	56.94						
	·_100	1438.4	-2371.7	940.3						
	·_1000	1568.0	-2586.5	1026.1						
	·_new	1567.8	-2586.5	1026.1						
dA_{21}	·_10	57.97	-93.20	35.50	47.18					
	·_100	738.3	-1215.7	481.0	281.7					
	·_1000	807.4	-1330.2	526.7	306.0					
	·_new	807.4	-1330.2	526.7	306.0					
dA_{22}	·_10	-93.52	150.36	-57.26	-75.82	121.84				
	·_100	-1221.1	2010.8	-795.6	-464.2	765.0				
	·_1000	-1336.0	2201.2	-871.6	-504.6	832.0				
	·_new	-1336.2	2201.7	-871.8	-504.6	832.2				
dA_{23}	·_10	35.81	-57.58	21.93	28.85	-46.37	17.65			
	·_100	486.5	-801.1	317.0	183.8	-303.0	120.0			
	·_1000	532.6	-877.5	347.5	200.0	-329.9	130.8			
	·_new	532.5	-877.5	347.5	200.0	-329.9	130.8			
dA_{31}	·_10	42.75	-68.70	26.14	39.50	-63.44	24.12	40.77		
	·_100	587.0	-966.7	382.5	228.5	-376.4	149.0	198.9		
	·_1000	642.8	-1059.0	419.4	248.1	-409.0	162.1	214.8		
	·_new	642.7	-1059.0	419.4	248.1	-409.0	162.1	214.8		
dA_{32}	·_10	-69.00	110.88	-42.19	-63.47	101.94	-38.76	-65.48	105.18	
	·_100	-971.3	1599.6	-633.0	-376.5	620.3	-245.6	-327.3	538.6	
	·_1000	-1063.9	1753.1	-694.3	-409.1	674.3	-267.3	-353.7	582.4	
	·_new	-1063.9	1753.1	-694.3	-409.1	674.4	-267.3	-353.7	582.4	
dA_{33}	·_10	26.44	-42.49	16.17	24.15	-38.79	14.75	24.90	-39.99	15.20
	·_100	387.2	-637.7	252.4	149.1	-245.7	97.3	129.3	-212.9	84.2
	·_1000	424.4	-699.2	277.0	162.2	-267.4	106.0	139.9	-230.4	91.2
	·_new	424.0	-698.7	276.8	162.1	-267.3	105.9	139.9	-230.4	91.2

Note: dA_{ij} denotes the ijth element of the matrix $h\sqrt{T}(\hat{A}-A)$. The results in the rows with ·_10, ·_100 and ·_1000 are the estimates of asymptotic covariance matrix based on the estimation of K as in (3.7) with truncation number 10,100,1000, respectively. The estimating results based on the estimation of Γ given in Corollary 3.9 are reported in the rows with ·_new.

matrix only. Special attention has been paid on the expression of the asymptotic covariance matrix. Different situations have been discussed to get an explicit expression of the covariance matrix in the limit theory.

The limit theory for explosive continuous time model is not covered in the paper, but, it should be feasible once the limiting distribution of the estimated VAR coefficient matrix is known (see, Phillips and Magdalinos, 2008, 2011). The method proposed in this paper can also help us extend the limit theory of the ML estimator to the continuous time models which are driven by Lévy process as long as the estimated VAR coefficient matrix is available. If it is assumed that the mean and the variance of the Lévy process are finite, the error term in the discrete time model would be an independent sequence with a finite mean and a finite variance and the limit theory given in the paper is still be applied.

The new representation is illustrated in an empirical application to an affine term structure model of 1-month, 3-month and 6-month U.S. interest rates. It has been shown that the estimates of the "mean reversion matrix" and its asymptotic covariance matrix based on the traditional representation of ML/LS estimator highly depends on the choice of the truncation number. When small truncation numbers are used, the estimates and the elements in the estimated covariance matrix could be far away from the correct values. Although it is clear that a large truncation numbers are required to make the estimates accurate, there are no guidelines for making such choices. However, the new representation is free from this difficulty.

Although the paper focuses on multivariate Ornstein-Uhlenbeck process, our method is applicable to continuous time models with a linear drift and with more flexible diffusion functions, i.e.,

$$dX(t) = (AX(t) + b)\,dt + \Sigma^{1/2} q(X(t);\phi)\,dW(t),$$

where $\Sigma^{1/2} q(X(t);\phi)$ is a general diffusion function with parameter vector ϕ. The Nowman approximation (Nowman, 1997), which approximates the diffusion function within each unit interval $[(j-1)h, jh)$ by its left end point value, lead to the approximate model

$$dX(t) = (AX(t) + b)\,dt + \Sigma^{1/2} q\left(X_{(j-1)h};\phi\right) dW(t) \quad \text{for } t \in [(j-1)h, jh).$$

The correspondingly approximate discrete time model is

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

where $F = e^{Ah}$, $g = \int_0^h e^{As} b\,ds$ and $\varepsilon_t = \int_{(t-1)h}^{th} e^{As} \Sigma^{1/2} q(X_{t-1};\phi)\,ds$. Based on this approximate discrete time model, one can first estimate F, then estimate A from $\hat{F}$ through a nonlinear matrix logarithmic mapping. At this point, the new representation proposed in this paper can be applied to obtain improved estimates.

APPENDIX

A Proofs in Section 2

Proof of formula (2.8). With the non-singularity of $\beta'\alpha$, it is easy to get that

$$\begin{aligned} g &= \int_0^h e^{As} b ds = \left\{ \int_0^h \left(e^{As} - I_m \right) ds + I_m h \right\} b \\ &= \left\{ \int_0^h \alpha \left[e^{(\beta'\alpha)s} - I_\tau \right] \left(\beta'\alpha \right)^{-1} \beta' ds + I_m h \right\} b \\ &= \left\{ \alpha \left(\int_0^h e^{(\beta'\alpha)s} ds \right) \left(\beta'\alpha \right)^{-1} \beta' - \alpha \left(\beta'\alpha \right)^{-1} \beta' h + I_m h \right\} b \\ &= \left\{ \alpha \left(\beta'\alpha \right)^{-1} \left[e^{(\beta'\alpha)h} - I_\tau \right] \left(\beta'\alpha \right)^{-1} \beta' - \alpha \left(\beta'\alpha \right)^{-1} \beta' h + I_m h \right\} b = \vartheta(h) b. \end{aligned}$$

■

Proof of formula (2.9). Let $\Upsilon = \left(\beta'\alpha\right)^{-1} \beta'\Sigma\beta \left(\alpha'\beta\right)^{-1}$, we could have

$$\begin{aligned} \Omega &= \int_0^h e^{As} \Sigma e^{A's} ds = \int_0^h \left(e^{As} - I_m \right) \Sigma \left(e^{A's} - I_m \right) ds + \int_0^h \left(e^{As}\Sigma + \Sigma e^{A's} - \Sigma \right) ds \\ &= \int_0^h \alpha \left[e^{(\beta'\alpha)s} - I_\tau \right] \Upsilon \left[e^{(\alpha'\beta)s} - I_\tau \right] \alpha' ds + \vartheta(h) \Sigma + \Sigma \left[\vartheta(h) \right]' - \Sigma h \\ &= \alpha \left(\int_0^h e^{(\beta'\alpha)s} \Upsilon e^{(\alpha'\beta)s} ds \right) \alpha' - \alpha \left(\int_0^h e^{(\beta'\alpha)s} ds \right) \Upsilon \alpha' - \alpha \Upsilon \left(\int_0^h e^{(\beta'\alpha)s} ds \right)' \alpha' \\ &\quad + \alpha \Upsilon \alpha' h + \vartheta(h) \Sigma + \Sigma \left[\vartheta(h) \right]' - \Sigma h \\ &= \alpha \left(\int_0^h e^{(\beta'\alpha)s} \Upsilon e^{(\alpha'\beta)s} ds \right) \alpha' - \underline{\alpha} \Upsilon \alpha' - \alpha \Upsilon \underline{\alpha}' + \alpha \Upsilon \alpha' h + \vartheta(h) \Sigma + \Sigma \left[\vartheta(h) \right]' - \Sigma h, \end{aligned}$$

where the final equality comes from

$$\alpha \left(\int_0^h e^{(\beta'\alpha)s} ds \right) \Upsilon \alpha' = \alpha \left[e^{(\beta'\alpha)h} - I_\tau \right] \left(\beta'\alpha \right)^{-1} \Upsilon \alpha' = \underline{\alpha} \Upsilon \alpha'.$$

Noting that $\beta'\alpha$ is non-singular, it is easy to find out that

$$\left(\beta'\alpha \right) \Xi + \Xi \left(\beta'\alpha \right)' = e^{(\beta'\alpha)h} \Upsilon e^{(\alpha'\beta)h} - \Upsilon$$

where $\Xi = \int_0^h e^{(\beta'\alpha)s} \Upsilon e^{(\alpha'\beta)s} ds$ (Phillips, 1973, derived a similar expression for $\int_0^h e^{As} \Sigma e^{A's} ds$ when A is non-singular). The fact that

$$Vec(\Xi) = \left\{ \left(\beta'\alpha \right) \otimes I_\tau + I_\tau \otimes \left(\beta'\alpha \right) \right\}^{-1} \left\{ e^{(\beta'\alpha)h} \otimes e^{(\beta'\alpha)h} - I_\tau \otimes I_\tau \right\} Vec(\Upsilon).$$

completes the proof immediately. ■

Proof of formula (2.10). Under the condition that all the eigenvalues of $\left(\beta'\alpha\right)$ have negative real parts, it is not hard to show that $\begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}$ and $\begin{bmatrix}\beta & \alpha_\perp\end{bmatrix}$ are nonsingular matrices. Therefore,

$$\begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}\begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}^{-1} = I_m = \begin{bmatrix}\beta' \\ \alpha'_\perp\end{bmatrix}^{-1}\begin{bmatrix}\beta' \\ \alpha'_\perp\end{bmatrix}.$$

Hence,

$$\begin{aligned} I_m &= \begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}\begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}^{-1}\begin{bmatrix}\beta' \\ \alpha'_\perp\end{bmatrix}^{-1}\begin{bmatrix}\beta' \\ \alpha'_\perp\end{bmatrix} = \begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}\left(\begin{bmatrix}\beta' \\ \alpha'_\perp\end{bmatrix}\begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}\right)^{-1}\begin{bmatrix}\beta' \\ \alpha'_\perp\end{bmatrix} \\ &= \begin{bmatrix}\alpha & \beta_\perp\end{bmatrix}\begin{bmatrix}\beta'\alpha & 0_{\tau\times(m-\tau)} \\ 0_{(m-\tau)\times\tau} & \alpha'_\perp\beta_\perp\end{bmatrix}^{-1}\begin{bmatrix}\beta' \\ \alpha'_\perp\end{bmatrix} = \alpha\left(\beta'\alpha\right)^{-1}\beta' + \beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\alpha'_\perp, \end{aligned}$$

where the third equality comes from the fact that $\alpha'\alpha_\perp = 0_{\tau\times(m-\tau)}$ and $\beta'\beta_\perp = 0_{\tau\times(m-\tau)}$. ■

B proofs in Section 3

Proof of Theorem 3.5. (a) Based on the formulae (3.9) and (3.13), straightforward calculations 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

$$\begin{aligned} hVec\left(\hat{A}-A\right) &= \sum_{j=1}^{m}\left\{Vec\left[\left(I-\hat{F}\right)^{j-1}\right]\left(\hat{f}_j - f_j\right)\right\} \\ &\quad - \sum_{j=2}^{m}\sum_{s=0}^{j-2} f_j\left\{(I-F)^s \otimes \left[\left(I-\hat{F}\right)'\right]^{j-2-s}\right\}Vec\left(\hat{F}-F\right). \end{aligned} \tag{B.1}$$

Note the fact that $\{f_j\}_{j=1}^m$ defined as in (3.10), (3.11) and (3.12), are differentiable functions on $\{C_j\}_{j=1}^m$, and $\{C_j\}_{j=1}^m$ are continuous functions of elements of F. Let $\digamma_j = \begin{bmatrix}\frac{\partial f_j}{\partial C_m} & \frac{\partial f_j}{\partial C_{m-1}} & \cdots & \frac{\partial f_j}{\partial C_1}\end{bmatrix}'$, for $j = 1, 2, \cdots, m$, the first order Taylor expansion provides us

$$\hat{f}_j - f_j = \digamma'_j\begin{bmatrix}\hat{C}_m - C_m \\ \vdots \\ \hat{C}_1 - C_1\end{bmatrix} + o_p\left(Vec\left(\hat{F}-F\right)\right), \quad \text{for } j = 1, 2, \cdots, m.$$

Let $|\cdot|$ denotes determinant of a 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

$$\begin{aligned}
&\begin{bmatrix} 1 & z & z^2 & \cdots & z^{m-1} \end{bmatrix} \begin{bmatrix} \hat{C}_m - C_m \\ \vdots \\ \hat{C}_1 - C_1 \end{bmatrix} \\
&= \det\left[zI - \left(I - \hat{F}\right)\right] - \det\left[zI - (I - F)\right] = \left|\hat{\psi}_z\right| - |\psi_z| \\
&= \left.\frac{\partial \left|\hat{\psi}_z\right|}{\partial \left[Vec\left(\hat{\psi}_z\right)\right]'}\right|_{\hat{\psi}_z = \psi_z} Vec\left(\hat{\psi}_z - \psi_z\right) + o_p\left(Vec\left(\hat{\psi}_z - \psi_z\right)\right) \\
&= \left.\frac{\partial \left|\hat{\psi}_z\right|}{\partial \left[Vec\left(\hat{\psi}_z\right)\right]'}\right|_{\hat{\psi}_z = \psi_z} Vec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right) \\
&= \left[Vec\left(H_z\right)\right]' Vec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right)
\end{aligned}$$

where $H_z = \left[adj\left(\psi_z\right)\right]'$. The first equation comes from the representation of characteristic polynomial in (3.8). 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 $H = \begin{bmatrix} Vec\left(H_1\right) & \cdots & Vec\left(H_m\right) \end{bmatrix}'$. 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 of $\hat{f}_j - f_j$, we could get

$$\hat{f}_j - f_j = \digamma_j' L^{-1} HVec\left(\hat{F} - F\right) + o_p\left(Vec\left(\hat{F} - F\right)\right).$$

Taking this result into formula (B.1) will finally complete the proof.

(b) Under the condition that $\hat{F} \stackrel{p}{\longrightarrow} F$, the result $\tilde{\Gamma} \stackrel{p}{\longrightarrow} \Gamma$ is straightforward. What needs to be proved is the nonsingularity of Γ. Note that $\hat{F} \stackrel{p}{\longrightarrow} F$ is equivalent to the statement that every subsequence of $\hat{F}$, denoted by $\left\{\hat{F}_{T_k}\right\}$, has subsequence of its own as $\left\{\hat{F}_{T_{k'}}\right\}$, which satisfies

$\hat{F}_{T_{k'}} \xrightarrow{a.s.} F$ when $T_{k'} \longrightarrow \infty$. Consequently, the corresponding sequence $\left\{\tilde{\Gamma}_{T_{k'}}\right\}$ satisfies $\tilde{\Gamma}_{T_{k'}} \xrightarrow{a.s.} \Gamma$. Therefore, if we could show the nonsingularity of Γ under the condition of $\hat{F} \xrightarrow{a.s.} F$, the conclusion should still be true automatically in the situation where $\hat{F} \xrightarrow{p} F$. As a result, in the later context, we complete our proof under the condition of $\hat{F} \xrightarrow{a.s.} F$.

As eigenvalues are continuous functions of elements of a matrix, when $\hat{F} \xrightarrow{a.s.} F$, we have

$$\hat{\lambda}_j\left(\hat{F}\right) \xrightarrow{a.s} \lambda_j\left(F\right), \qquad \text{for} \quad j = 1, 2, \cdots, m.$$

Under Assumption 1, $spec\left\{F\right\} \cap R_0^- = \varnothing$. Hence, when sample size T is large enough, we could get

$$spec\left\{\hat{F}\right\} \cap R_0^- = \varnothing.$$

Therefore, based on Lemma 3.1, $\hat{A}$ represented in (3.13) is the principal logarithm of $\hat{F}$. We could rewrite the relationship between $\hat{A}$ and $\hat{F}$ as $\hat{F} = \exp\left\{\hat{A}h\right\} = \sum_{j=0}^{\infty}\left(\hat{A}h\right)^j/j!$. As a result,

$$\hat{F} - F = \left(\hat{A} - A\right)h + \sum_{j=2}^{\infty}\frac{\left(\hat{A}h\right)^j - (Ah)^j}{j!} = \left(\hat{A} - A\right)h + \sum_{j=2}^{\infty}\frac{(h)^j}{j!}\left\{\sum_{s=0}^{j-1}A^s\left(\hat{A} - A\right)\left(\hat{A}\right)^{j-1-s}\right\},$$

which leads to

$$Vec\left(\hat{F} - F\right) = \underbrace{\left\{I_{m^2} + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[A^s \otimes \left(\hat{A}'\right)^{j-1-s}\right]\right\}}_{\widetilde{E}} hVec\left(\hat{A} - A\right). \tag{B.2}$$

From $\hat{F} \xrightarrow{a.s} F$, it is easy to get $\hat{A} \xrightarrow{a.s} A$. Letting $\|\cdot\|$ denote the *Frobenius* norm of a matrix, and

$$E = I_{m^2} + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[A^s \otimes \left(A'\right)^{j-1-s}\right],$$

we have

$$\begin{aligned}
\left\|\widetilde{E} - E\right\| &= \left\|\sum_{j=2}^{\infty}\sum_{s=0}^{j-2}\frac{(h)^{j-1}}{j!}\left\{A^s \otimes \left[\left(\hat{A}'\right)^{j-1-s} - \left(A'\right)^{j-1-s}\right]\right\}\right\| \\
&= \left\|\sum_{j=2}^{\infty}\sum_{s=0}^{j-2}\sum_{\tau=0}^{j-2-s}\frac{(h)^{j-1}}{j!}\left\{A^s \otimes \left[\left(A'\right)^{\tau}\left(\hat{A}' - A'\right)\left(\hat{A}'\right)^{j-2-s-\tau}\right]\right\}\right\| \\
&\leq \sum_{j=2}^{\infty}\sum_{s=0}^{j-2}\sum_{\tau=0}^{j-2-s}\left\|\frac{(h)^{j-1}}{j!}\left\{A^s \otimes \left[\left(A'\right)^{\tau}\left(\hat{A}' - A'\right)\left(\hat{A}'\right)^{j-2-s-\tau}\right]\right\}\right\| \\
&\leq \sum_{j=2}^{\infty}\sum_{s=0}^{j-2}\sum_{\tau=0}^{j-2-s}\frac{(h)^{j-1}}{j!}\left\|A^s\right\|\left\|\left(A'\right)^{\tau}\left(\hat{A}' - A'\right)\left(\hat{A}'\right)^{j-2-s-\tau}\right\| \\
&\leq \left\{\sum_{j=2}^{\infty}\sum_{s=0}^{j-2}\sum_{\tau=0}^{j-2-s}\frac{(h)^{j-1}}{j!}\left\|A\right\|^s\left\|A\right\|^{\tau}\left\|\hat{A}\right\|^{j-2-s-\tau}\right\}\left\|\hat{A} - A\right\| \longrightarrow 0,
\end{aligned}$$

for $\left\|\hat{A}-A\right\| \longrightarrow 0$ and

$$\begin{aligned}
& \sum_{j=2}^{\infty}\sum_{s=0}^{j-2}\sum_{\tau=0}^{j-2-s}\frac{(h)^{j-1}}{j!}\|A\|^{s}\|A\|^{\tau}\left\|\hat{A}\right\|^{j-2-s-\tau} \\
= & \sum_{j=2}^{\infty}\sum_{s=0}^{j-2}\sum_{\tau=0}^{j-2-s}\frac{(h)^{j-1}}{j!}\|A\|^{s}\|A\|^{\tau}\|A\|^{j-2-s-\tau}+o(1) \\
= & \sum_{j=2}^{\infty}\sum_{s=0}^{j-2}(j-1-s)\frac{(h)^{j-1}}{j!}\|A\|^{s}\|A\|^{j-2-s}+o(1) \\
\leq & \sum_{j=2}^{\infty}\frac{(h)^{j-1}}{(j-2)!}\|A\|^{j-2}+o(1)=h\exp\left\{\|A\|h\right\}+o(1) \text{ is bounded.}
\end{aligned}$$

Therefore, $\widetilde{E}\xrightarrow{a.s}E$. Based on formulae (B.2) and (B.1), we have

$$\begin{aligned}
Vec\left(\hat{F}-F\right) &= \widetilde{E}\left\{hVec\left(\hat{A}-A\right)\right\}=E\left\{hVec\left(\hat{A}-A\right)\right\}+o_{p}\left(Vec\left(\hat{A}-A\right)\right) \\
&= E\left\{\widetilde{\Gamma}Vec\left(\hat{F}-F\right)+o_{p}\left(Vec\left(\hat{F}-F\right)\right)\right\}+o_{p}\left(Vec\left(\hat{A}-A\right)\right) \\
&= E\Gamma Vec\left(\hat{F}-F\right)+o_{p}\left(Vec\left(\hat{F}-F\right)\right).
\end{aligned}$$

As $Vec\left(\hat{F}-F\right)$ is a random vector whose elements can take any value, it must be true that $E\Gamma=I$. Therefore, Γ is nonsingular. ■

Proof of Theorem 3.10. The proof is only given under the condition of $\hat{F}\xrightarrow{a.s}F$, but can be applied to the case where $\hat{F}\xrightarrow{p}F$ because of the reason argued in the proof of (b) in Theorem 3.5.

As A has the Jordan decomposition form of

$$A=Pdiag\left\{\eta_{1},\cdots,\eta_{m}\right\}Q=PVQ,$$

the coefficient matrix E mentioned in the proof of (b) in Theorem 3.5 can be rewritten as

$$\begin{aligned}
E &= I_{m^2}+\sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[A^{s}\otimes\left(A'\right)^{j-1-s}\right]=I_{m^2}+\sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[\left(PV^{s}Q\right)\otimes\left(Q'V^{j-1-s}P'\right)\right] \\
&= I_{m^2}+\sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[\left(P\otimes Q'\right)\left(V^{s}\otimes V^{j-1-s}\right)\left(Q\otimes P'\right)\right] \\
&= \left(P\otimes Q'\right)\left\{I_{m^2}+\sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[\left(V^{s}\otimes V^{j-1-s}\right)\right]\right\}\left(Q\otimes P'\right) \\
&= \left(P\otimes Q'\right)diag\left\{\Lambda_{1},\cdots,\Lambda_{m}\right\}\left(Q\otimes P'\right),
\end{aligned}$$

where, for $k = 1, \cdots, m$,

$$\Lambda_k = diag\left(\left\{1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^s\eta_v^{j-1-s}\right\}_{v=1}^{m}\right).$$

When $k = v$, it is easy to get

$$1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^s\eta_v^{j-1-s} = 1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^{j-1} = 1 + \sum_{j=2}^{\infty}\frac{(h)^{j-1}}{(j-1)!}\eta_k^{j-1} = e^{\eta_k h}.$$

When $k \neq v$, as all the eigenvalues are distinct, we assume $|\eta_k| < |\eta_v|$ (the same result is easy to get when $|\eta_k| > |\eta_v|$). Then

$$\begin{aligned}
&1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\eta_k^s\eta_v^{j-1-s} = 1 + \sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left(\frac{\eta_k}{\eta_v}\right)^s\eta_v^{j-1} = 1 + \sum_{j=2}^{\infty}\frac{(h)^{j-1}\,\eta_v^{j-1}}{j!}\frac{1-(\eta_k/\eta_v)^j}{1-(\eta_k/\eta_v)} \\
&= 1 + \sum_{j=2}^{\infty}\frac{(h)^{j-1}}{j!}\frac{\eta_v^j - \eta_k^j}{\eta_v - \eta_k} = 1 + \frac{1}{(\eta_v - \eta_k)\,h}\sum_{j=2}^{\infty}\frac{(h)^j}{j!}\left(\eta_v^j - \eta_k^j\right) \\
&= 1 + \frac{1}{(\eta_v - \eta_k)\,h}\left\{(\exp\{\eta_v h\} - 1 - \eta_v h) - (\exp\{\eta_k h\} - 1 - \eta_k h)\right\} \\
&= \frac{e^{\eta_v h} - e^{\eta_k h}}{(\eta_v - \eta_k)\,h}.
\end{aligned}$$

The fact that $E\Gamma = I$ obtained in the proof of (b) in Theorem 3.5 completes the proof. ■

Proof of Theorem 3.12. Letting $\{\eta_1, \cdots, \eta_m\} = V$ and $\{\hat{\eta}_1, \cdots, \hat{\eta}_m\} = \hat{V}$, when T is large enough, we have

$$\begin{aligned}
\hat{A} - A = dA &= \hat{P}\hat{V}\hat{P}^{-1} - PVP^{-1} \\
&= \left(\hat{P} - P\right)\hat{V}\hat{P}^{-1} + P\left(\hat{V} - V\right)\hat{P}^{-1} + PV\left(\hat{P}^{-1} - P^{-1}\right) \\
&= \left(\hat{P} - P\right)\hat{V}\hat{P}^{-1} + P\left(\hat{V} - V\right)\hat{P}^{-1} - PVP^{-1}\left(\hat{P} - P\right)\hat{P}^{-1} \\
&= (dP)\,\hat{V}\hat{P}^{-1} + P\,(dV)\,\hat{P}^{-1} - PVP^{-1}\,(dP)\,\hat{P}^{-1},
\end{aligned}$$

and

$$\begin{aligned}
P^{-1}\,(dA)\,\hat{P} &= P^{-1}\,(dP)\,\hat{V} + dV - VP^{-1}\,(dP) \\
&= dV + P^{-1}\,(dP)\left(\hat{V} - V\right) + P^{-1}\,(dP)\,V - VP^{-1}\,(dP)
\end{aligned}$$

Noting that the diagonal elements of $P^{-1}\,(dP)\,V$ and $VP^{-1}\,(dP)$ are identical (c.f., Phillips, 1982), it is true that

$$\left(P^{-1}\,(dA)\,\hat{P}\right)_{(j,j)} = \left(\hat{\eta}_j - \eta_j\right) + \left(P^{-1}\,(dP)\right)_{(j,j)}\left(\hat{\eta}_j - \eta_j\right).$$

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

$$\left\{1+\left[P^{-1}\left(dP\right)\right]_{(j,j)}\right\}\left(\hat{\eta}_j-\eta_j\right)=\left(p^j\right)'\left(dA\right)\hat{p}_j=\left[\left(p^j\right)'\otimes\hat{p}_j'\right]Vec\left(dA\right).$$

As ordered eigenvalues and eigenvectors under normalization are continuous functions of elements of the matrix, it is achieved that

$$\hat{\eta}_j-\eta_j=\left[\left(p^j\right)'\otimes p_j'\right]Vec\left(\hat{A}-A\right)+o_p\left(Vec\left(\hat{A}-A\right)\right).$$

■

Proof of Theorem 3.14. The proof can be completed by taking the same steps in the proof of the Theorem 3.5 after changing a few notations. Hence, it is omitted. ■

C proofs in Section 4

Proof of Theorem 4.3. We only give the proof of

$$T^{3/2}h\left(\hat{A}-A\right)\mu_1\xrightarrow{d}\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right).$$

Other results in the Theorem can be proved immediately in a similar way.

Firstly, we give the proof for the case in which $m>1$. As $A=0_{m\times m}$ and $F=I$, simple 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 $s=2,\cdots,m$. Hence, it is obtained that

$$h\left(\hat{A}-A\right)=h\hat{A}=\ln\left(\hat{F}\right)=\hat{f}_1I+\hat{f}_2\left(I-\hat{F}\right)+\cdots+\hat{f}_m\left(I-\hat{F}\right)^{m-1}.$$

Note that

$$T^{3/2}\left(\hat{F}-I\right)\mu_1=T^{3/2}\left(\hat{F}-F\right)\mu_1\Rightarrow\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right).$$

Hence

$$T^{3/2}\left(\hat{F}-I\right)^j\mu_1\xrightarrow{p}0\ \text{ for } j>1.$$

From the consistency of $\hat{f}_j$, $j=1,\cdots,m$, the following expression is obtained

$$T^{3/2}h\left(\hat{A}-A\right)\mu_1=T^{3/2}\hat{f}_1I\cdot\mu_1-\hat{f}_2T^{3/2}\left(\hat{F}-I\right)\mu_1+o_p(1).$$

Note that

$$\hat{f}_1=\int_0^1\frac{\hat{C}_mS^{m-1}}{1+\hat{C}_1S+\cdots+\hat{C}_mS^m}dS,$$

with

$$\hat{C}_m=(-1)^m\det\left(I-\hat{F}\right)=(-1)^m\sum_{j=1}^{m!}\zeta_j.$$

where ζ_j, for each j, is a multiplication in terms of elements in matrix $\left(I-\hat{F}\right)$ with the number of m. Because of

$$T\left(\hat{F}-I\right)=T\left(\hat{F}-F\right)\Rightarrow f\left(B_0,\underline{B}_0^{**},\underline{\Delta}_{21}\right)U_2',$$

it it easy to get $\zeta_j\sim O_p\left(T^{-m}\right)$. Therefore, $\hat{C}_m\sim O_p\left(T^{-m}\right)$ and $T^{3/2}\hat{C}_m\xrightarrow{p}0$ when $m>1$. Based on the consistency of $\hat{C}_j$, $j=1,\cdots,m$, it is easy to get $T^{3/2}\hat{f}_1\xrightarrow{p}0$. Consequently,

$$\begin{aligned}T^{3/2}h\left(\hat{A}-A\right)\mu_1 &= -\hat{f}_2T^{3/2}\left(\hat{F}-I\right)\mu_1+o_p(1)\\ &\xrightarrow{d} -f_2\cdot\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right)=\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right).\end{aligned}$$

For the case $m=1$, the proposed estimator possesses the form of

$$h\hat{A}=\hat{f}_1,$$

and

$$\hat{f}_1=\hat{C}_1\int_0^1\frac{1}{1+\hat{C}_1S}dS,\ \text{with}\ \hat{C}_1=(-1)\left(1-\hat{F}\right)\xrightarrow{p}C_1=0.$$

As $\int_0^1\frac{1}{1+\hat{C}_1S}dS\xrightarrow{p}\int_0^1\frac{1}{1+C_1S}dS=1$, we can have

$$\begin{aligned}T^{3/2}h\left(\hat{A}-A\right)\mu_1 &= T^{3/2}h\hat{A}\mu_1=T^{3/2}\hat{C}_1\mu_1\int_0^1\frac{1}{1+\hat{C}_1S}dS\\ &= T^{3/2}\left(\hat{F}-1\right)\mu_1+o_p\left(1\right)\xrightarrow{d}\left(g'g\right)^{-1/2}f\left(B_0,\underline{P},\underline{\delta}\right).\end{aligned}$$

■

Proof of Theorem 4.7. When $b=0$, it is known from the Theorem 4.3 that $Th\left(\hat{A}-A\right)\xrightarrow{d}f\left(B_0,B_0^*,\Delta_{21}\right)$. Then, it is easy to get

$$Th\sum_{j=1}^m\hat{\eta}_j=Th\times tr\left(\hat{A}\right)=Th\Delta Vec\left(\hat{A}\right)\xrightarrow{d}\Delta\cdot Vec\left[f\left(B_0,B_0^*,\Delta_{21}\right)\right],$$

where Δ is a $1\times m^2$ row vector whose $1^{st},[m+2]^{th},\cdots,[(m-1)m+m]^{th}$ elements are 1 and 0 elsewhere. The other parts of the theorem can be easily proved by using the same method. ■

Proof of Lemma 4.8. It is easy to get

$$\left[\hat{\Phi}_1-\Phi_1,\hat{\Phi}_{2d}-\Phi_{2d},\hat{\Phi}_{2D}-\Phi_{2D},\hat{g}-g\right]=\left[\sum_{t=1}^T\varepsilon_tZ_{t-1}^{\dagger\prime}\right]\times\left[\sum_{t=1}^TZ_{t-1}^{\dagger}Z_{t-1}^{\dagger\prime}\right]^{-1},$$

where $Z_t^{\dagger}=\left[Y_{1t},\ \ Y_{2dt},\ \ Y_{2Dt},\ \ 1\right]$, $Y_{2dt}=d_1'Y_{2t}$ and $Y_{2Dt}=D_2'Y_{2t}$. The processes X_t,Y_{1t} and Y_{2t} possess error terms as $\varepsilon_t=\int_{(t-1)h}^{th}e^{(\alpha\beta')(th-s)}\Sigma^{1/2}dW\left(s\right)$, $\nu_{1t}=\int_{(t-1)h}^{th}e^{(\beta'\alpha)(th-s)}\beta'\Sigma^{1/2}dW\left(s\right)$ and $\nu_{2t}=\int_{(t-1)h}^{th}\alpha_\perp'\Sigma^{1/2}dW\left(s\right)$, respectively. Note the important fact that $\{\varepsilon_t\}_{t=1}^\infty$, $\{\nu_{1t}\}_{t=1}^\infty$ and $\{\nu_{2t}\}_{t=1}^\infty$ have no correlations across time index t with each other. Then, simply using FCLT could completes the proof. (The asymptotics for some of the elements in $\sum_{t=1}^{T}\varepsilon_t Z_{t-1}^{\dagger\prime}$ and $\sum_{t=1}^{T} Z_{t-1}^{\dagger} Z_{t-1}^{\dagger\prime}$ are well-known (see, for example, Park and Phillips, 1988, 1989).) ■

Proof of Theorem 4.13. The findings listed in Theorem 3.5 can be rewritten as

$$hVec\left(\hat{A}-A\right)=\Gamma Vec\left(\hat{F}-F\right)+o_p\left(Vec\left(\hat{F}-F\right)\right).$$

Then, based on the Theorem 4.8 and the relationship of

$$\hat{F}-F=\left(\hat{\Phi}_1-\Phi_1\right)\beta'+\left(\hat{\Phi}_{2d}-\Phi_{2d}\right)d_1'\alpha_\perp'+\left(\hat{\Phi}_{2D}-\Phi_{2D}\right)D_2'\alpha_\perp',$$

the results in (a) and (b) are proved immediately.

However, the appearance of the term as $o_p\left(Vec\left(\hat{F}-F\right)\right)$ puts some difficulty on deriving the results in (c) and (d). For example, it is not easy to prove the condition as

$$T\left(I_m\otimes\left[\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}D_2\right]'\right)\Gamma^{-1}o_p\left(Vec\left(\hat{F}-F\right)\right)=o_p\left(1\right),$$

which is a quite necessary condition for getting the result in (c). In the following, a method is applied which manages to avoid this sort of difficulty.

Let $\Psi_T(\cdot)$ denote the cumulative distribution function (cdf) of

$$hT\left(I_m\otimes\left[\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}D_2\right]'\right)\Gamma^{-1}Vec\left(\hat{A}-A\right).$$

And, let $\Psi(\cdot)$ be the asymptotic cdf of

$$T\left(I_m\otimes\left[D_2'D_2\right]\right)Vec\left(\hat{\Phi}_{2D}-\Phi_{2D}\right)=T\left(I_m\otimes\left[\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}D_2\right]'\right)Vec\left(\hat{F}-F\right).$$

The result in (c) would be obtained immediately if, at any right continuous point of $\Psi(\cdot)$, say, $\eth_0$, we have $\Psi_T(\eth_0)\to\Psi(\eth_0)$ as $T\to\infty$. Let $\{\Psi_{T_k}(\eth_0)\}$ be ANY subsequence of $\{\Psi_T(\eth_0)\}$. Then, to get $\Psi_T(\eth_0)\to\Psi(\eth_0)$ as $T\to\infty$, the only thing needed to do is to prove that there EXIST a sequence like $\{\Psi_{T_{ks}}(\eth_0)\}$, which is a subsequence of $\{\Psi_{T_k}(\eth_0)\}$, satisfying $\Psi_{T_{ks}}(\eth_0)\to\{\Psi(\eth_0)\}$.

Note that, for ANY subsequence of $\hat{F}$ denoting by $\left\{\hat{F}_{T_k}\right\}$, there always EXIST a sequence like $\left\{\hat{F}_{T_{ks}}\right\}$, which is a subsequence of $\left\{\hat{F}_{T_k}\right\}$, satisfying

$$\hat{F}_{T_{ks}}\xrightarrow{a.s.}F.$$

From the proof of Theorem 3.5, when T_{ks} is large enough, we have

$$Vec\left(\hat{F}_{T_{ks}}-F\right)=\underbrace{\left\{I_{m^2}+\sum_{j=2}^{\infty}\sum_{s=0}^{j-1}\frac{(h)^{j-1}}{j!}\left[A^s\otimes\left(\hat{A}_{T_{ks}}'\right)^{j-1-s}\right]\right\}}_{\widetilde{E}_{T_{ks}}}hVec\left(\hat{A}_{T_{ks}}-A\right),$$

and

$$\widetilde{E}_{T_{ks}} \xrightarrow{a.s.} \Gamma^{-1},$$

where $\left\{\hat{A}_{T_{ks}}\right\}$ is a sequence of $\hat{A}$ corresponding to $\left\{\hat{F}_{T_{ks}}\right\}$. Therefore,

$$T\left(I_m \otimes \left[\beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} D_2\right]'\right) Vec\left(\hat{F}_{T_{ks}} - F\right) \overset{d}{\sim} hT\left(I_m \otimes \left[\beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} D_2\right]'\right) \Gamma^{-1} Vec\left(\hat{A}_{T_{ks}} - A\right).$$

Note that $\{\Psi_{T_{ks}}(\cdot)\}$ can be regarded as the cdf of $hT\left(I_m \otimes \left[\beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} D_2\right]'\right) \Gamma^{-1} Vec\left(\hat{A}_{T_{ks}} - A\right)$. While, $\Psi(\cdot)$ is the asymptotic cdf of $T\left(I_m \otimes \left[\beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} D_2\right]'\right) Vec\left(\hat{F}_{T_{ks}} - F\right)$. As a result, $\Psi_{T_{ks}}(\eth_0) \to \{\Psi(\eth_0)\}$ is obtained immediately.

The results in (d) can be proved straightforwardly by using the same approach to prove (c). ■

REFERENCES

Ahn, S.K., and G.C., Reinsel, 1990, Estimation for Partially Nonstationary Multivariate Autoregressive Models. *Journal of the American Statistical Association*, 85, 813–823.

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.

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.

Bernstein, D.S., 2009, *Matrix Mathematics: Theory, Facts, and Formulas.* Princeton University Press, Princeton.

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

Comte, F., 1999, Discret and Continuous Time Cointegration. *Journal of Econometrics*, 88, 207–226.

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

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

Duffie, D., and R. Kan, 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.

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

Engle, R.F., and C.W.J., Granger, 1987, Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55, 251–276.

Franchi, M., 2007, The Integration Order of Vector Autoregressive Processes. *Econometric Theory*, 23, 546–553.

Granger, C.W.J., 1981, Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*, 28, 121–130.

Granger, C.W.J., 1983, Co-integrated Variables and Error Correcting Models. Unpublished discussion paper no. 83-13a (University of California, San Diego, CA).

Granger, C.W.J., and A.A., Weiss, 1983, Time Series Analysis of Error Correction Models. In: *Studies in Econometric Time-series and Multivariate Statistics*, S. Karlin, T. Amemiya and L.A. Goodman, eds., New York: Academic Press, pp. 255–278.

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

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

Hansen, L. P., and T. J., Sargent, 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.

Johansen, S., M., 1988, Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12, 231–254.

Johansen, S., M., 1991, Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector. *Econometrica*, 59, 1551–1580.

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.

Kessler, M., and A. Rahbek, 2004, Identification and Inference for Cointegrated and Ergodic Gaussian Diffusions. *Statistical Inference for Stochastic Process*, 36, 153–188.

McCrorie, J.R., 2003, The Problem of Aliasing in Indentifying Finite Parameter Continuous Time Stochastic Models. *Acta Applicandae Mathematicae*, 79, 9–16.

McCrorie, J.R., 2009, Estimating Continuous Time Models on the Basis of Discrete Data via an Exact Discrete Analog. *Econometric Theory*, 25, 1120–1137.

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

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.

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

Phillips, P. C. B., 1973, The Problem fo 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., 1991, Error Correction and Long-Run Equilibria in Continuous Time. *Econometrica*, 59, 967–980.

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

Phillips, P.C.B., and T. Magdalinos, 2008, Limit Theory for Explosively Cointegrated Systems *Econometric Theory*, 24, 865–887.

Phillips, P.C.B., and T. Magdalinos, 2011, Inconsistent VAR Regression with Common Explosive Roots. Working paper.

Phillips, P. C. B., and S. Ouliaris, 1990, Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica*, 58, 165–193.

Phillips, P. C. B., and P. Perron, 1988, Testing for a Unit Root in Time Series Regression. *Biometrika*, 75, 335–346.

Sims, C.A., J.H., Stock and M.W., Watson, 1990, Inference in Linear Time Series Models with some Unit Roots. *Econometrica*, 58, 113–144.

Stock, J.H., 1987, Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors. *Econometrica*, 55, 1035–1056.

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

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

Zellner, A., and H. Theil, 1962, Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica*, 30, 54–78.