SERIES ESTIMATION OF STOCHASTIC PROCESSES: RECENT DEVELOPMENTS AND ECONOMETRIC APPLICATIONS

By

Peter C.B. Phillips and Zhipeng Liao

September 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1871

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Series Estimation of Stochastic Processes: Recent Developments and Econometric Applications*

Peter C. B. Phillips† Zhipeng Liao‡

August, 2012

Abstract

This paper overviews recent developments in series estimation of stochastic processes and some of their applications in econometrics. Underlying this approach is the idea that a stochastic process may under certain conditions be represented in terms of a set of orthonormal basis functions, giving a series representation that involves deterministic functions. Several applications of this series approximation method are discussed. The first shows how a continuous function can be approximated by a linear combination of Brownian motions (BMs), which is useful in the study of the spurious regressions. The second application utilizes the series representation of BM to investigate the effect of the presence of deterministic trends in a regression on traditional unit-root tests. The third uses basis functions in the series approximation as instrumental variables (IVs) to perform efficient estimation of the parameters in cointegrated systems. The fourth application proposes alternative estimators of long-run variances in some econometric models with dependent data, thereby providing autocorrelation robust inference methods in these models. We review some work related to these applications and some ongoing research involving series approximation methods.

Keywords: Cointegrated System; HAC Estimation; Instrumental Variables; Lasso Regression; Karhunen-Loève Representation; Long-run Variance; Reproducing Kernel Hilbert Space; Oracle Effciency; Orthonormal System; Trend Basis.

JEL classification: C22

*Our thanks to the referee and Editor for helpful comments on the original version of this paper.

† *Yale University, University of Auckland, University of Southampton* and *Singapore Management University.* Support from the NSF under Grant No SES 09-56687 is gratefully acknowledged. Email: peter.phillips@yale.edu

‡Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu

1 Introduction

The explicit representation of stochastic processes has a long history in the probability literature with many applications in asymptotic statistics. For example, in early work Kac and Siegert (1947) showed that a Gaussian process can be decomposed as an infinite linear combination of deterministic functions. In fact, a much more powerful representation theory holds for any stochastic process that is continuous in quadratic mean, a result that was separately established in Karhunen (1946) and Loève (1955). In the modern literature, the explicit decomposition of a stochastic process in this way is known as the Karhunen-Loève (KL) representation or transformation. The deterministic functions used in this KL representation are orthonormal basis functions in a Hilbert space constructed on the same interval for which the stochastic process is defined.

The KL transformation was originally proposed to assist in determining the exact forms of certain asymptotic distributions associated with Cramér-von Mises type statistics. These asymptotic distributions typically take the form of a quadratic functional of a Brownian motion (BM) or Brownian Bridge process, such as the integral over some interval of the square of the process. For example, the KL transformation reveals that the integral of the square of a Gaussian process is distributed as a weighted infinite sum of independent chi-square variates with one degree of freedom. Other examples are given in the work of Anderson and Darling (1952), Watson (1961), and Stephens (1976); and Shorack and Wellner (1988) provide an overview of results of this kind.

The theory underlying the KL representation relies on Mercer's theorem, which represents the covariance function of any quadratic mean continuous stochastic process $\{X_t\}_{t\in\mathcal{T}}$ in terms of basis functions in a Hilbert space $L^2(\mathcal{T})$ defined under some measure on $\mathcal{T}$. The covariance function can be viewed as an inner product of the Hilbert space $L^2(X)$ generated by the stochastic process [1]. On the other hand, by Mercer's theorem, the covariance function has a representation which defines an inner product with respect to another Hilbert space $L^2_R(\mathcal{T})$. This new Hilbert space $L^2_R(\mathcal{T})$ has the attractive feature that any function in the space can be reproduced by its inner product with the covariance function. As a result, $L^2_R(\mathcal{T})$ is often called a reproducing kernel Hilbert space (RKHS) with the covariance function being the reproducing kernel. It was noted in Parzen (1959) that the two Hilbert spaces $L^2(X)$ and $L^2_R(\mathcal{T})$ are isometrically isomorphic, which implies

[1] The Hilbert space generated by the stochastic process $\{X_t\}_{t\in\mathcal{T}}$ is the completion of the space defined as the linear span of any finite elements $X_{t_1}, ..., X_{t_n}$, where $t_k \in \mathcal{T}$, $k = 1, ..., n$ and $n = 1, 2, ...$

that analysis of the stochastic process $\{X_t\}_{t\in\mathcal{T}}$ in $L^2(X)$ can be equivalently executed in $L^2_R(\mathcal{T})$. Sometimes a complicated problem in $L^2(X)$ space can be treated more easily in the RKHS space $L^2_R(\mathcal{T})$. More details about the analysis of time series in RKHS space can be found in Parzen (1959, 1961a, 1961b and 1963). Berlinet and Thomas-Agnan (2003) provide a modern introduction to RKHS techniques and their applications in statistics and probability.

While statisticians and probabilists have focussed on the roles of the KL representation in determining asymptotic distributions of functionals of stochastic processes or rephrasing time series analysis issues equivalently in different spaces, econometric research has taken these representations in a new direction. In particular, econometricians have discovered that empirical versions of the KL representation are a powerful tool for estimation and inference in many econometric models. This chapter reviews some of these recent developments of the KL representation theory and its empirical application in econometrics.

First, the KL representation provides a bridging mechanism that links underlying stochastic trends with various empirical representations in terms of deterministic trend functions. This mechanism reveals the channel by which the presence of deterministic trends in a regression can affect tests involving stochastic trends, such as unit root and cointegration tests. For example, Phillips (2001) showed how the asymptotic distributions of coefficient based unit root test statistics are changed in a material way as deterministic function regressors continue to be added to the empirical regression model. This work used KL theory to show that as the number of deterministic functions tends to infinity, the coefficient based unit root tests have asymptotic normal distributions after appropriate centering and scaling rather than conventional unit root distributions. These new asymptotics are useful in revising traditional unit root limit theory and ensuring that tests have size that is robust to the inclusion of many deterministic trend functions or trajectory fitting by deterministic trends or trend breaks.

Secondly, the KL theory not only directly represents stochastic trends in terms of deterministic trends, it also provides a basis for linking independent stochastic trends. This extension of the theory was studied in Phillips (1998) where it was established that a continuous deterministic function can be approximated using linear combinations of independent BMs with a corresponding result for the approximation of a continuous stochastic process. This latter result is particularly useful in analyzing and interpreting so-called spurious regressions involving the regression of an integrated process on other (possibly independent) integrated processes.

The KL theory and its empirical extensions in Phillips (1998) explain how regression of an integrated process on a set of basis functions can successfully reproduce the whole process when the number of basis functions expands to infinity with the sample size. An empirically important implication of this result that is explored in Phillips (2012) is that trend basis functions can themselves serve as instrumental variables because they satisfy both orthogonality and relevance conditions in nonstationary regression. For instance, in a cointegrated system this type of trend IV estimator of the cointegrating matrix does not suffer from high order bias problems because the basis functions are independent of the errors in the cointegrated system by virtue of their construction, thereby delivering natural orthogonality. Moreover, the IV estimator is asymptotically efficient because when the number of basis functions diverges to infinity, the integrated regressors in the cointegrating system are reproduced by the basis functions, thereby assuring complete relevance in the limit. In short, the long-run behavior of the endogenous variables in a cointegrated system is fully captured through a linear projection on basis functions in the limit while maintaining orthogonality of the instruments.

As the above discussion outlines, KL theory helps to answer questions about the asymptotic behavior of linear projections of integrated processes on deterministic bases. A related question relates to the properties of similar projections of the trajectory of a stationary process on deterministic bases. In exploring this question, Phillips (2005b) proposed a new estimator of the long-run variance (LRV) of a stationary time series. This type of estimator is by nature a series estimate of the LRV and has since been extensively studied in Chen, Liao and Sun (2012), Chen, Hahn and Liao (2012), Sun (2011, 2012) and Sun and Kim (2012a,b).

The remainder of this chapter is organized as follows. Section 2 presents the KL representation theory for continuous stochastic processes together with some recent developments of this theory. Section 3 explores the implications of the KL theory for empirical practice, focusing on understanding and interpreting spurious regressions in econometrics. Section 4 investigates the implication of these representations for unit root tests when there are deterministic trends in the model. Section 5 considers the optimal estimation of cointegrated systems using basis functions as instruments. The optimal estimation method discussed in section 5 assumes that the cointegration space of the cointegration system is known from the beginning. In section 6, we present a new method which optimally estimates the cointegration system without even knowing the cointegration rank. Series estimation of LRVs and some of the recent applications of this theory are discussed in

section 7. Section 8 concludes and briefly describes some ongoing and future research in the field. Technical derivations are included in the Appendix.

2 Orthogonal Representation of Stochastic Processes

We start with a motivating discussion in Euclidean space concerned with the orthonormal representation of finite dimensional random vectors. Such representations provide useful intuition concerning the infinite dimensional case and are indicative of the construction of orthonormal representations of stochastic processes in Hilbert space.

Suppose X is a T-dimensional random vector with mean zero and positive definite covariance matrix Σ. Let $\{(\lambda_k, \varphi_k)\}_{k=1}^T$ be the pairs of eigenvalues and orthonormalized right eigenvectors of Σ. Define

$$Z_T' = X'\Phi_T = [z_1, ..., z_T],$$

where $\Phi_T = [\varphi_1, ..., \varphi_T]$, then Z_T is a T-dimensional random vector with mean zero and covariance matrix $\Lambda_T = diag(\lambda_1, ..., \lambda_T)$. We have the representation

$$X = \Phi_T Z_T = \sum_{k=1}^{T} z_k \varphi_k = \sum_{k=1}^{T} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k, \tag{2.1}$$

where the $\xi_k = \lambda_k^{-\frac{1}{2}} z_k$ have zero mean and covariances $E[\xi_k \xi_{k'}] = \delta_{kk'}$ where $\delta_{kk'}$ is the Kronecker delta. When X is a zero mean Gaussian random vector, $[\xi_1, ..., \xi_T]'$ is simply a T-dimensional standard Gaussian random vector. Expression (2.1) indicates that any T-dimensional ($T \in \mathbb{Z}_+ \equiv \{1, 2, ..., \}$) random vector can be represented by a weighted linear combination of T orthonormal real vectors, where the weights are random and uncorrelated across different vectors. Moreover, (2.1) shows that the spectrum of the covariance matrix of the random vector X plays a key role in the decomposition of X into a linear combination of deterministic functions with random coefficients.

The orthonormal representation of a random vector given in (2.1) can be generalized to a stochastic process $X(t)$ with $t \in [a, b]$ for $\infty < a < b < \infty$, and in this form it is known as the Kac-Siegert decomposition or KL representation. We can use heuristics based on those used to derive (2.1) to develop the corresponding KL representation of a general stochastic process. Without loss of generality, we assume the random variables $\{X(t) : t \in [a, b]\}$ live

on the same probability space $(\Omega, \mathcal{G}, P)$. The first and second moments of $X(t)$ for any $t \in [a, b]$ are given by

$$E\left[X(t)\right] = \int_{\Omega} X(t)dP \text{ and } E\left[X^2(t)\right] = \int_{\Omega} X^2(t)dP.$$

The following assumption is used to derive the KL representation of $X(t)$.

Assumption 2.1 *The stochastic process $X(t)$ satisfies $E\left[X(t)\right] = 0$ and $E\left[X^2(t)\right] < \infty$ for all $t \in [a, b]$.*

The zero mean assumption is innocuous as the process $X(t)$ can always be recentred about its mean. The second moment assumption is important because it allows us to embed $X(t)$ in a Hilbert space and use the Hilbert space setting to establish the representation. Accordingly, let $L^2(X)$ denote the Hilbert space naturally generated by $X(t)$ so that it is equipped with the following inner product and semi-norm

$$\langle X_1, X_2 \rangle \equiv \int_{\Omega} X_1 X_2 dP \text{ and } \|X_1\|^2 = \int_{\Omega} X_1^2 dP,$$

for any $X_1, X_2 \in L^2(X)$. Let $L^2[a, b]$ be the Hilbert space of square integrable functions on $[a, b]$ with the following inner product and semi-norm

$$\langle g_1, g_2 \rangle_e \equiv \int_a^b g_1(s) g_2(s) ds \text{ and } \|g_1\|_e^2 = \int_a^b g_1^2(s) ds, \tag{2.2}$$

for any $g_1, g_2 \in L^2[a, b]$.

Under Assumption 2.1, the covariance/kernel function $\gamma(\cdot, \cdot)$ of the stochastic process $X(t)$ can be defined as

$$\gamma(s, t) \equiv E\left[X(s)X(t)\right] \tag{2.3}$$

for any $s, t \in [a, b]$. Let $\{(\lambda_k, \varphi_k)\}_{k \in K}$ be the collection of all different pairs (λ, φ) which satisfy the following integral equation

$$\lambda \varphi(t) = \int_a^b \gamma(s, t) \varphi(s) ds \text{ with } \|\varphi\|_e = 1, \tag{2.4}$$

where λ and φ are called as the eigenvalue and normalized eigenfunction of the kernel $\gamma(\cdot, \cdot)$ respectively.

Using heuristics based on the procedure involved in deriving (2.1), one might expect to use the eigenvalues and eigenfunctions of the kernel function $\gamma(\cdot,\cdot)$ to represent the stochastic process $X(t)$ as a sum of the form

$$X(t) \stackrel{?}{=} \sum_{k=1}^{\bar{K}} \left[\int_a^b X(t)\varphi_k(t)dt \right] \varphi_k(t) = \sum_{k=1}^{\bar{K}} z_k \varphi_k(t) = \sum_{k=1}^{\bar{K}} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t) \tag{2.5}$$

where $z_k \equiv \int_a^b X(t)\varphi_k(t)dt$ and $\xi_k \equiv \lambda_k^{-\frac{1}{2}} z_k$ for $k = 1, ..., \bar{K}$ and some (possibly infinite) $\bar{K}$. To ensure that the expression in (2.5) is indeed an orthonormal representation of $X(t)$, we first confirm that the components ξ_k satisfy

$$E\left[\xi_k\right] = 0 \text{ and } E\left[\xi_k \xi_{k'}\right] = \delta_{kk'} \text{ for any } k, k' = 1, ..., \bar{K} \tag{2.6}$$

where $\delta_{kk'}$ is Kronecker's delta, and that the process $X(t)$ can be written as

$$X(t) = \sum_{k=1}^{\bar{K}} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t) \text{ a.s. } t \in [a,b] \text{ in quadratic mean} \tag{2.7}$$

The following condition is sufficient to show (2.6) and (2.7).

Assumption 2.2 *The stochastic process $X(t)$ is continuous in quadratic mean (q.m.) on $[a,b]$, i.e., for any $t_o \in [a,b]$*

$$\|X(t) - X(t_o)\|^2 = E\left\{[X(t) - X(t_o)]^2\right\} \to 0 \tag{2.8}$$

as $|t - t_o| \to 0$, where we require $t \in [a,b]$ such that $X(t)$ is well defined in (2.8).

In this assumption, continuity in q.m. is well defined at the boundary points a and b because we only need to consider the limits from the right to a and limits from the left to b. The following lemma is useful in deriving the KL representation of $X(t)$.

Lemma 2.1 *Suppose that Assumptions 2.1 and 2.2 are satisfied. Then the kernel function $\gamma(\cdot,\cdot)$ of the stochastic process $X(t)$ is symmetric, continuous, and bounded and it satisfies*

$$\int_a^b \int_a^b g(t)\gamma(t,s)g(s)dsdt \geq 0$$

for any $g \in L^2[a,b]$.

Under Assumptions 2.1 and 2.2, Lemma 2.1 implies that sufficient conditions for Mercer's theorem hold (see e.g., Shorack and Wellner, 1986, p. 208). Thus, we can invoke Mercer's theorem to deduce that the normalized eigenfunctions of the kernel function $\gamma(\cdot,\cdot)$ are continuous on $[a,b]$ and form an orthonormal basis for the Hilbert space $L^2[a,b]$. Mercer's theorem ensures that the kernel function $\gamma(\cdot,\cdot)$ has the following series representation in terms of this orthonormal basis

$$\gamma(s,t) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(s)\varphi_k(t) \tag{2.9}$$

uniformly in s and t. The following theorem justifies the orthonormal representation of $X(t)$ in (2.5) with $\bar{K} = \infty$ and (2.6) and (2.7) both holding.

Theorem 2.1 *Suppose the stochastic process $X(t)$ satisfies Assumptions 2.1 and 2.2. Then $X(t)$ has the following orthogonal expansion*

$$\begin{aligned} X(t) &= \sum_{k=1}^{\infty} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t) \\ \text{with } \xi_k &= \lambda_k^{-\frac{1}{2}} \int_a^b X(t)\varphi_k(t)dt, \end{aligned} \tag{2.10}$$

where $E[\xi_k \xi_{k'}] = \int_a^b \varphi_k(t)\varphi_{k'}(t)dt = \delta_{kk'}$ and $\delta_{kk'}$ denotes the Kronecker delta, if and only if λ_k and φ_k ($k \in \mathbb{Z}_+$) are the eigenvalues and normalized eigenfunctions of $\gamma(\cdot,\cdot)$. The series in (2.10) converges in q.m. uniformly on $[a,b]$.

Just as a continuous function in $L^2[a,b]$ can be represented by series involving Fourier basis functions, Theorem 2.1 indicates that a continuous (in q.m.) stochastic process can also be represented by orthonormal basis functions that lie in $L^2[a,b]$. However, unlike the series representation of a continuous function, the coefficients of the basis functions in the KL representation are random variables and uncorrelated with each other. The representation of $X(t)$ in (2.10) converges in q.m. but may not necessarily converge pointwise[2]. For this reason, the equivalence in (2.10) is sometimes represented by the symbol "$\sim$" or "$\stackrel{d}{=}$", signifying that the series is convergent in the L^2 sense and that distributional equivalence

[2] Similarly, the series representation of a continuous function may not converge pointwise unless the function has right and left derivatives at that point

applies. Importantly, the series (2.10) involves two sets of orthonormal components – the orthogonal random sequence $\{\xi_k\}$ and the orthogonal basis functions $\{\varphi_k\}$.

When the continuous time stochastic process $X(t)$ is covariance stationary, it is well-known that $X(t)$ has the following spectral (SP) representation

$$X(t) = \int_{-\infty}^{+\infty} \exp(i\lambda t) dZ(\lambda) \tag{2.11}$$

where i is the imaginary unit and $Z(\lambda)$ denotes the related complex spectral process which has orthogonal increments whose variance involve the corresponding increments in the spectral distribution function. In expression (2.11), $X(t)$ is represented as an uncountably infinite sum of the products of deterministic functions $\exp(i\lambda t)$ and random coefficients $dZ(\lambda)$ at different frequencies, which differs from the KL expression (2.10) in several ways. Most importantly, (2.10) represents in quadratic mean the trajectory of the process over a fixed interval $[a, b]$, whereas (2.11) is a representation of the entire stochastic process $X(t)$ in terms of the mean square limit of approximating Riemann Stieltjes sums (e.g. Hannan, 1970, p. 41).

When the stochastic process $X(t)$ is a BM, its KL representation has more structure. For example, the representation in (2.10) holds almost surely and uniformly in $[0, 1]$ and the random coefficients $\{\xi_k\}$ are *iid* normal. These special structures are summarized in the following corollary.

Corollary 2.2 *Let $B_\sigma(t)$ be a BM with variance σ^2, then (i) $B_\sigma(t)$ has the following orthogonal expansion*

$$B_\sigma(t) = \sum_{k=1}^{\infty} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t), \tag{2.12}$$

where

$$\xi_k = \lambda_k^{-\frac{1}{2}} \int_a^b B_\sigma(t) \varphi_k(t) dt \tag{2.13}$$

and the above representation converges almost surely uniformly on $[a, b]$; (ii) the random sequence $\{\xi_k\}_k$ is iid $N(0, \sigma^2)$; (iii) the random sequence $\{\eta_k\}_k$ defined by

$$\eta_k = \int_a^b \varphi_k(t) dB_\sigma(t) \tag{2.14}$$

is also iid $N(0, \sigma^2)$.

It is easy to verify that $B_\sigma(t)$ satisfies Assumptions 2.1 and 2.2. Thus by Theorem 2.1, $B_\sigma(t)$ has a KL representation which converges in q.m. uniformly on $[a, b]$. The q.m. convergence of the series in (2.9) is strengthened to almost sure convergence in (2.12) by applying the martingale convergence theorem to the martingale formed by finite sums of (2.12). The normality of ξ_k or η_k ($k \in \mathbb{Z}_+$) holds directly in view of the representations (2.13) and (2.14) (the normal stability theorem, Loève, 1976) and the independence of the sequence $\{\xi_k\}$ or $\{\eta_k\}$ follows by their orthogonality. It is clear that the expression in (2.10) links the stochastic trend $X(t)$ with a set of deterministic functions $\{\varphi_k(\cdot)\}_{k=1}^\infty$ which might be regarded as trend functions on the interval $[a, b]$. Since the random wandering behavior of the stochastic trend $X(t)$ over $[a, b]$ is fully captured by the deterministic functions in its KL representation, throughout this chapter we shall call $\{\varphi_k(\cdot) : k \in \mathbb{Z}_+\}$ the trend basis functions.

Example 2.3 *Let $B(\cdot)$ be a standard BM on $[0, 1]$. Then Corollary 2.2 ensures that $B(\cdot)$ has a KL representation. By definition, the kernel function of $B(\cdot)$ is $\gamma(s, t) = \min(s, t)$ and its eigenvalues and normalized eigenfunctions are characterized by the following integral equation*

$$\lambda\varphi(t) = \int_0^t s\varphi(s)ds + t\int_t^1 \varphi(s)ds \text{ with } \int_0^1 \varphi^2(s)ds = 1.$$

Direct calculation reveals that the eigenvalues and normalized eigenfunctions of $\gamma(\cdot, \cdot)$ are

$$\lambda_k = \frac{1}{(k - 1/2)^2\pi^2} \text{ and } \varphi_k(t) = \sqrt{2}\sin\left[(k - 1/2)\pi t\right] \tag{2.15}$$

respectively for $k \in \mathbb{Z}_+$. Applying Corollary 2.2, we have the following explicit orthonormal representation

$$B(t) = \sqrt{2}\sum_{k=1}^\infty \frac{\sin\left[(k - 1/2)\pi t\right]}{(k - 1/2)\pi}\xi_k \tag{2.16}$$

which holds almost surely and uniformly in $t \in [0, 1]$, where

$$\xi_k = \sqrt{2}(k - 1/2)\pi\int_0^1 B(t)\sin\left[(k - 1/2)\pi t\right]dt \text{ for } k \in \mathbb{Z}_+. \tag{2.17}$$

Invoking Corollary 2.2, we know that $\{\xi_k\}_{k=1}^\infty$ are iid standard normal random variables.

Example 2.4 *Let $W(\cdot)$ be a Brownian bridge process corresponding to the standard BM $B(\cdot)$ on $[0, 1]$, i.e. $W(t) = B(t) - tB(1)$ for any $t \in [0, 1]$. It is easy to show that $W(\cdot)$ is*

continuous in q.m. on $[0,1]$. *Moreover,* $W(\cdot)$ *has kernel function* $\gamma(s,t) = \min(s,t) - st$, *which is continuous on* $[0,1]$. *The eigenvalues and normalized eigenfunctions are characterized by the following integral equation*

$$\lambda\varphi(t) = \int_0^t s\varphi(s)ds + t\int_t^1 \varphi(s)ds - \frac{t}{2} \text{ with } \int_0^1 \varphi^2(s)ds = 1.$$

Direct calculation shows that the eigenvalues and normalized eigenfunctions of $\gamma(\cdot,\cdot)$ *are*

$$\lambda_k = \frac{1}{k^2\pi^2} \text{ and } \varphi_k(t) = \sqrt{2}\sin(k\pi t),$$

respectively for $k \in \mathbb{Z}_+$. *Applying Theorem 2.1, we have the following orthonormal representation*

$$W(t) = \sqrt{2}\sum_{k=1}^{\infty} \frac{\sin(k\pi t)}{k\pi}\xi_k \tag{2.18}$$

where

$$\xi_k = \sqrt{2}k\pi \int_0^1 B(t)\sin(k\pi t)dt \text{ for } k \in \mathbb{Z}_+. \tag{2.19}$$

Using similar arguments as those in Corollary 2.2, the representation in (2.18) is convergent almost surely and uniformly in $t \in [0,1]$. *Moreover,* $\{\xi_k\}_{k=1}^{\infty}$ *are iid standard normal random variables.*

The KL representation of a BM can be used to decompose other stochastic processes that are functionals of BMs. The simplest example is the Brownian bridge process studied in the above example. From the representation in (2.16),

$$W(t) = B(t) - tB(1) = \sqrt{2}\sum_{k=1}^{\infty} \frac{\sin\left[(k-1/2)\pi t\right] + (-1)^k t}{(k-1/2)\pi}\xi_{1,k}$$

where $\xi_{1,k}$ $(k \in \mathbb{Z}_+)$ is defined in (2.17). Of course, one can also use the KL representation of the Brownian bridge process to decompose the process $B(t)$ into a series form, viz.,

$$B(t) = tB(1) + W(t) = t\xi_{2,0} + \sqrt{2}\sum_{k=1}^{\infty} \frac{\sin(k\pi t)}{k\pi}\xi_{2,k} \tag{2.20}$$

where $\xi_{2,0} = B(1)$ and the $\xi_{2,k}$ $(k \in \mathbb{Z}_+)$ are defined in (2.19).

The second example is the quadratic functional of a BM given by the integral $[B]_1 =$

$\int_0^1 B^2(t)dt$. Using the KL representation (2.16) the following series expression for the functional is readily obtained

$$[B]_1 = \int_0^1 B^2(t)dt = \sum_{k=1}^{\infty} \frac{1}{(k-1/2)^2\pi^2}\xi_k^2,$$

which implies that the random variable $[B]_1$ has a distribution equivalent to the weighted sum of independent chi-square random variables, each with unit degree of freedom.

The third example is the series representation of an Ornstein–Uhlenbeck (O-U) process. We provide two illustrations of how to construct such as series.

Example 2.5 *Let $J_c(t)$ be a stochastic process on $t \in [0,1]$ satisfying the following stochastic differential equation*

$$dJ_c(t) = cJ_c(t)dt + \sigma dB(t) \tag{2.21}$$

where c and $\sigma > 0$ are constants and $B(\cdot)$ denotes a standard BM. Set $\sigma = 1$ for convenience in what follows. It is clear that when $c = 0$, the process $J_c(t)$ reduces to standard BM $B(t)$. Under the initial condition $J_c(0) = B(0) = 0$, the above differential equation has the following solution

$$J_c(t) = B(t) + c\int_0^t \exp[(t-s)c]B(s)ds. \tag{2.22}$$

Using the series representation (2.20) and the solution (2.22), one obtains for $t \in [0,1]$

$$\begin{aligned} J_c(t) &= \frac{e^{ct}-1}{c}\xi_{2,0} + \sum_{k=1}^{\infty}\left[\sqrt{2}e^{ct}\int_0^t e^{-cs}\cos(k\pi s)ds\right]\xi_k \\ &= \frac{e^{ct}-1}{c}\xi_{2,0} + \sqrt{2}\sum_{k=1}^{\infty}\frac{ce^{ct}+k\pi\sin(k\pi t)-c\cos(k\pi t)}{c^2+k^2\pi^2}\xi_k, \end{aligned} \tag{2.23}$$

where ξ_k ($k \in \mathbb{Z}_+$) are iid standard normal random variables. The series representation (2.23) involves the orthogonal sequence $\{\xi_k\}$ associated with the Brownian bridge $W(t)$. An alternative representation that uses the series (2.16) is given in Phillips (1998) and in (8.2) below.

Example 2.6 *Suppose $X(t)$ is an O-U process with covariance kernel $\gamma(s,t) = e^{-|s-t|}$. In this case the process $X(t)$ satisfies the stochastic differential equation (2.21) with $c = -1$*

and $\sigma = \sqrt{2}$. *Then the KL representation of* $X(t)$ *over* $t \in [0,1]$ *is*

$$X(t) = \sqrt{2}\sum_{k=0}^{\infty} \frac{\sin\left\{\omega_k\left(t-\frac{1}{2}\right)+\left(k+1\right)\frac{\pi}{2}\right)\right\}}{\left(1+\lambda_k\right)^{1/2}}\xi_k, \tag{2.24}$$

where ξ_k $(k \in \mathbb{Z}_+)$ *are iid standard normal random variables,* $\lambda_k = 2\left(1+\omega_k^2\right)^{-1}$, *and* $\omega_0, \omega_1, ...$ *are the positive roots of the equation*

$$\tan(\omega) = -2\frac{\omega}{1-\omega^2}$$

(Pugachev, 1959; see also Bosq, 2000, p. 27).

3 New Tools for Understanding Spurious Regression

Spurious regression refers to the phenomenon that arises when fitted least squares regression coefficients appear statistically significant even when there is no true relationship between the dependent variable and the regressors. In simulation studies, Granger and Newbold (1974) showed that the phenomenon occurs when independent random walks are regressed on one another. Similar phenomena occur in regressions of stochastic trends on deterministic polynomial regressors, as shown in Durlauf and Phillips (1988). Phenomena of this kind were originally investigated by Yule (1926) and the first analytic treatment and explanation was provided in Phillips (1986).

As seen in the previous section, the orthonormal representation (2.10) links the random function $X(\cdot)$ to deterministic basis functions $\varphi_j(\cdot)$ $(j \in \mathbb{Z}_+)$ on the Hilbert space $L^2[a,b]$. This linkage provides a powerful tool for studying relations between stochastic trends and deterministic trends, as demonstrated in Phillips (1998). The orthonormal representation (2.10) also provides useful insights in studying relations among stochastic trends.

Consider the normalized time series $B_n\left(\frac{t}{n}\right) = n^{-\frac{1}{2}}\sum_{s=1}^{t} u_s$, whose components u_t satisfy the following assumption.

Assumption 3.1 *For all* $t \geq 0$, u_t *has Wold representation*

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j\varepsilon_{t-j}, \ \sum_{j=0}^{\infty} j\left|c_j\right| < \infty \text{ and } C(1) \neq 0 \tag{3.1}$$

with $\varepsilon_t = iid(0, \sigma_\varepsilon^2)$ *with* $E\left(\left|\varepsilon_t\right|^p\right) < \infty$ *for some* $p > 2$.

Under the above assumption, one can invoke Lemma 3.1 of Phillips (2007) which shows that in a possibly expanded probability space we have the (in probability) approximation

$$\sup_{0\leq t\leq n}\left|B_n\left(\frac{t}{n}\right)-B_{\sigma_u}\left(\frac{t}{n}\right)\right|=o_p(n^{-\frac{1}{2}+\frac{1}{p}}) \tag{3.2}$$

where $B_{\sigma_u}(\cdot)$ denotes a BM with variance $\sigma_u^2=2\pi f_u(0)$ and $f_u(\cdot)$ is the spectral density of u_t. Using the KL representation[3] in (2.12) and the uniform approximation in (3.2), we can deduce that

$$\sup_{0\leq t\leq n}\left|B_n\left(\frac{t}{n}\right)-\sum_{k=1}^{\infty}\lambda_k^{\frac{1}{2}}\varphi_k\left(\frac{t}{n}\right)\xi_k\right|=o_p(1) \tag{3.3}$$

where $\{(\lambda_k,\varphi_k(\cdot))\}_{k=1}^{\infty}$ is the set of all pairs of eigenvalues and orthonormalized eigenfunctions of the kernel function $\gamma(s,t)=\sigma_u^2\min(s,t)$, and where ξ_k $(k\in\mathbb{Z}_+)$ are independent Gaussian random variables.

The result in (3.2) implies that the scaled partial sum $B_n\left(\frac{t}{n}\right)=n^{-\frac{1}{2}}\sum_{s=1}^{t}u_s$ can be uniformly represented in terms of the basis functions $\varphi_k(\cdot)$ $(k\in\mathbb{Z}_+)$ in $L^2[a,b]$ for all $t\leq n$. Such a uniform approximation motivates us to study empirical LS regression estimation in which the scaled partial sum $B_n\left(\frac{t}{n}\right)$ is fitted using K orthonormal basis functions $\varphi_k(\cdot)$ $(k=1,...,K)$, i.e.

$$B_n\left(\frac{t}{n}\right)=\sum_{k=1}^{K}\widehat{a}_{k,n}\varphi_k\left(\frac{t}{n}\right)+\widehat{u}_{t,K}, \tag{3.4}$$

where

$$\widehat{A}_K=(\widehat{a}_{1,n},...,\widehat{a}_{K,n})'=\left[\sum_{t=1}^{n}\Phi_K\left(\frac{t}{n}\right)\Phi_K'\left(\frac{t}{n}\right)\right]^{-1}\left[\sum_{t=1}^{n}\Phi_K\left(\frac{t}{n}\right)B_n\left(\frac{t}{n}\right)\right]$$

and $\Phi_K(\cdot)=[\varphi_1(\cdot),...,\varphi_K(\cdot)]$. There are several interesting questions we would like to ask about the regression in (3.4). First, what are the asymptotic properties of the estimator $\widehat{A}_K$? More specifically, if we rewrite the uniform approximation (3.3) in the form

$$B_n\left(\frac{t}{n}\right)=\Phi_K\left(\frac{t}{n}\right)\Lambda_K\overline{\xi}_K+\sum_{k=K+1}^{\infty}\lambda_k^{\frac{1}{2}}\varphi_k\left(\frac{t}{n}\right)\xi_k,$$

[3] The specific orthonormal representation of BM given in (2.16) can of course be used here. But we use the representation in (2.12) to make the results of this section applicable to general basis functions.

where $\Lambda_K \equiv diag(\lambda_1, ..., \lambda_K)$ and $\overline{\xi}_K = (\xi_1, ..., \xi_K)$, will the estimate $\widehat{A}_K$ replicate the random vector $\Lambda_K \overline{\xi}_K$ in the limit? In practical work an econometrician might specify a regression that represents an integrated time series such as $y_t = \sum_{s=1}^{t} u_s$ in terms of deterministic trends. Upon scaling, such a regression takes the form

$$B_n\left(\frac{t}{n}\right) = \Phi_K\left(\frac{t}{n}\right) A_{o,K} + v_{nk} \tag{3.5}$$

which may be fitted by least squares to achieve trend elimination. To test the significance of the regressors $\Phi_K(\cdot)$ in such a trend regression, a natural approach would be to use a t-statistic for a linear combination of the coefficients $c_K' A_{o,K}$, such as

$$t_{c_K'\widehat{A}_K} = \frac{c_K'\widehat{A}_K}{\sqrt{\left(n^{-1}\sum_{i=1}^{n}\widehat{u}_{t,K}^2\right) c_K' \left[\sum\limits_{t=1}^{n} \Phi_K\left(\frac{t}{n}\right)\Phi_K'\left(\frac{t}{n}\right)\right]^{-1} c_K}}$$

for any $c_K \in R^K$ with $c_K' c_K = 1$. Corresponding robust versions of $t_{c_K'\widehat{A}_K}$ using conventional HAC or HAR estimates of the variance of $c_K'\widehat{A}_K$ might also be used, options that we will discuss later. For now, what are the asymptotic properties of the statistic $t_{c_K'\widehat{A}_K}$ and how adequate is the test? Further, we might be interested in measuring goodness of fit using the estimated coefficient of determination

$$\widehat{R}_K^2 = \frac{\widehat{A}_K' \left[\sum\limits_{t=1}^{n} \Phi_K\left(\frac{t}{n}\right)\Phi_K'\left(\frac{t}{n}\right)\right] \widehat{A}_K}{n^{-1}\sum_{t=1}^{n} B_n^2\left(\frac{t}{n}\right)}.$$

What are the asymptotic properties of $\widehat{R}_K^2$ and how useful is this statistic as a measure of goodness of fit in the regression? The following theorem from Phillips (1998) answers these questions.

Theorem 3.1 *As $n \to \infty$, we have*

(a) $c_K'\widehat{A}_K \to_d c_K' \int_0^1 \Phi_K(r)B(r)dr \overset{d}{=} N\left(0, c_K'\Lambda_K c_K\right)$;

(b) $n^{-\frac{1}{2}} t_{c_K'\widehat{A}_K} \to_d c_K' \left[\int_0^1 \Phi_K(r)B(r)dr\right] \left[\int_0^1 B_{\varphi_K}^2(r)dr\right]^{-\frac{1}{2}}$;

(c) $\widehat{R}_K^2 \to_d 1 - \left[\int_0^1 B_{\varphi_K}^2(r)dr\right] \left[\int_0^1 B^2(r)dr\right]^{-1}$;

where $B_{\varphi_K}(\cdot) = B(\cdot) - \left[\int_0^1 B(r)\Phi_K(r)dr\right]\Phi_K'(\cdot)$ *is the projection residual of* $B(\cdot)$ *on* $\Phi_K(\cdot)$.

Theorem 3.1 explains the spurious regression phenomenon that arises when an integrated process is regressed on a set of trend basis functions. Part (a) implies that the OLS estimate $\widehat{a}_{k,n}$ has a limit that is equivalent to $\lambda_k^{\frac{1}{2}}\xi_k$ for $k = 1, ..., K$. Note that the weak convergence in part (a) leads to pointwise functional limits. In particular, it leads directly to the following pointwise functional convergence

$$\Phi_K(t)\widehat{A}_K \rightarrow_d \sum_{k=1}^{K} \lambda_k^{\frac{1}{2}} \varphi_k(t)\, \xi_k, \text{ for any } t \in [0,1]. \tag{3.6}$$

A corresponding uniform weak approximation, i.e.

$$\sup_{t\in[\mathbf{0},\mathbf{1}]} \left| \Phi_K(t)\widehat{A}_K - \sum_{k=1}^{K} \lambda_k^{\frac{1}{2}} \varphi_k(t)\, \xi_k \right| = o_p(1) \tag{3.7}$$

can be proved using bracketing entropy arguments and the rate of pointwise convergence in (3.6). We leave the theoretical justification of such a uniform approximation to future research. Part (b) confirms that trend basis functions are always significant when used in regressions to explain an integrated process because the related t-statistics always diverge as the sample size $n \rightarrow \infty$.[4] From the KL representation (2.10), we observe that for large K the Hilbert space projection residual $B_{\varphi_K}(\cdot)$ is close to zero with high probability. From Part (c), we see that in such a case, $\widehat{R}_K^2$ is also close to 1 with large probability.

The results in Theorem 3.1 are derived under the assumption that the number of trend basis functions is fixed. A natural question to ask is: what are the asymptotic properties of $c_K'\widehat{A}_K$, $t_{c_K'\widehat{A}_K}$ and $\widehat{R}_K^2$ if the number of the trend basis functions K diverges to infinity with the sample size n. Note that if $K \rightarrow \infty$, then

$$\left[\int_0^1 B(r)\Phi_K(r)dr\right] \Phi_K'(t) = \sum_{k=1}^{K} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t) \rightarrow_{a.s.} \sum_{k=1}^{\infty} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t) = B(t) \tag{3.8}$$

[4] The divergent behavior of the t-statistics might be thought to be a consequence of the use of OLS standard errors based on $n^{-1}\sum_{i=1}^{n} \widehat{u}_{t,K}^2$ which do not take account of serial dependence in the residuals. However, Phillips (1998) confirmed that divergence at a reduced rate continues to apply when HAC standard errors are used (employing an estimate of the long run variance (LRV)). On the other hand, if HAR estimates rather than HAC estimates are used (for example, a series LRV estimate with fixed number of basis functions, see section 7 for details), the t-statistics no longer diverge in general. Theorem 3.1 simply illustrates the spurious regression phenomenon when standard testing procedures based on OLS are employed.

where the almost sure convergence follows by the martingale convergence theorem. The convergence in (3.8) immediately implies

$$B_{\varphi_K}(t) = B(t) - \left[\int_0^1 B(r)\Phi_K(r)dr\right]\Phi_K'(t) \to_{a.s.} 0 \tag{3.9}$$

as $K \to \infty$. Now, using (3.9) and sequential asymptotic arguments, we deduce that

$$\Phi_K(t)\widehat{A}_K \quad \to_d \sum_{k=1}^{\infty} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t) = B(t), \tag{3.10}$$

$$\left|n^{-\frac{1}{2}} t_{c_K'\widehat{A}_K}\right| \quad \to_p \infty \text{ and } \widehat{R}_K^2 \to_p 1, \tag{3.11}$$

as $n \to \infty$ followed by $K \to \infty$. The result (3.10) indicates that the fitted value $\Phi_K(\cdot)\widehat{A}_K$ based on the OLS estimate $\widehat{A}_K$ fully replicates the BM $B(\cdot)$ as K goes to infinity. Moreover, (3.10) implies that all fitted coefficients are significant even when infinitely many trend basis functions are used in (3.3). Note that when more trend basis functions are added to the regression, the fitted coefficients become more significant, instead of being less significant, because the residual variance in the regression (3.4) converges to zero in probability when both K and n diverge to infinity. The second result in (3.11) implies that the model is perfectly fitted when $K \to \infty$, which is anticipated in view of (3.10).

The following theorem is due to Phillips (1998) and presents asymptotic properties of $c_K'\widehat{A}_K$, $t_{c_K'\widehat{A}_K}$ and $\widehat{R}_K^2$ under joint asymptotics when n and K pass to infinity jointly.

Theorem 3.2 *Suppose that $K \to \infty$, then $c_K'\Lambda_K c_K$ converges to a positive constant $\sigma_c^2 = c'\Lambda c$, where $c = (c_1, c_2, ...)$, $\Lambda \equiv diag(\lambda_1, \lambda_2, ...)$ and $c'c = 1$. Moreover, if $K \to \infty$ and $K/n \to 0$ as $n \to \infty$, then we have (a) $c_K'\widehat{A}_K \to_d N\left(0, \sigma_c^2\right)$; (b) $n^{-\frac{1}{2}} t_{c_K'\widehat{a}_K}$ diverges; and (c) $\widehat{R}_K^2 \to_p 1$.*

From Theorem 3.2 it follows that the asymptotic properties of $c_K'\widehat{A}_K$, $t_{c_K'\widehat{A}_K}$ and $\widehat{R}_K^2$ under joint limits are very similar to their sequential asymptotic properties. Thus, the above discussion about the results in (3.10) and (3.11) also applies to Theorem 3.2.

As this analysis shows, the KL representation is a powerful tool in interpreting regressions of stochastic trends on deterministic trends. The KL representation can also link different BMs, because different BMs can themselves each be represented in terms of the same set of orthonormal basis functions. This intuition explains spurious regressions that arise when an integrated process is regressed on other (possibly independent) integrated

processes. The following theorem, again from Phillips (1998), indicates that any BM can be represented in terms of infinitely many independent standard BMs. This theory assists our understanding of empirical regressions among integrated processes that may be of full rank (or non-cointegrating). Such regressions are considered prototypical spurious regressions following the simulation study of Granger and Newbold (1974).

Theorem 3.3 *Let $B_\sigma(\cdot)$ be a BM on $[0,1]$ with variance σ^2 and let $\varepsilon > 0$ be arbitrarily small. Then we can find a sequence of independent BMs $\{B_i^*(\cdot)\}_{i=1}^N$ that are independent of $B_\sigma(\cdot)$ and a sequence of random variables $\{d_i\}_{i=1}^N$ defined on an augmented probability space $(\Omega, \mathcal{F}, P)$, such that as $N \to \infty$,*

(a) $\sup_{t\in[0,1]} \left| B_\sigma(t) - \sum_{i=1}^N d_i B_i^*(t) \right| < \varepsilon$ *a.s. P;*

(b) $\int_0^1 \left[B_\sigma(t) - \sum_{i=1}^N d_i B_i^*(t) \right]^2 dt < \varepsilon$ *a.s. P;*

(c) $B_\sigma(t) \stackrel{d}{=} \sum_{i=1}^\infty d_i B_i^*(t)$ *in* $L^2[a,b]$ *a.s. P.*

Part (c) of Theorem 3.3 shows that an arbitrary BM $B_\sigma(\cdot)$ has an L_2 representation in terms of independent standard BMs with random coefficients. It also gives us a model for the classic spurious regression of independent random walks. In this model, the role of the regressors and the coefficients becomes reversed. The coefficients d_i are random and they are co-dependent with the dependent variable $B_\sigma(t)$. The variables $B_i^*(t)$ are functions that take the form of BM sample paths, and these paths are independent of the dependent variable, just like the fixed coefficients in a conventional linear regression model. Thus, instead of a spurious relationship, we have a model that serves as a representation of one BM in terms of a collection of other BMs. The coefficients in this model provide the connective tissue that relates these random functions.

4 New Unit Root Asymptotics with Deterministic Trends

Since the mid 1980s it has been well understood that the presence of deterministic functions in a regression affects tests involving stochastic trends even asymptotically. This dependence has an important bearing on the practical implementation of unit root and cointegration tests. For example, the following model involves both an autoregressive component and some auxiliary regressors which include a trend component

$$Y_t = \rho_o Y_{t-1} + b_o' X_t + u_t. \tag{4.1}$$

Here Y_t and u_t are scalars and X_t is a p-vector of deterministic trends. Suppose that u_t is $iid(0, \sigma^2)$ and X_t, Y_t satisfy

$$D_n \sum_{s=1}^{\lfloor nt \rfloor} X_s \to_d X(t) \text{ and } n^{-\frac{1}{2}} Y_{\lfloor nt \rfloor} \to_d B_\sigma(t) \tag{4.2}$$

for any $t \in [0, 1]$ as $n \to \infty$, where D_n is a suitable $p \times p$ diagonal scaling matrix, $X(\cdot)$ is a p-dimensional vector of piecewise continuous functions and $B_\sigma(\cdot)$ is a BM with variance σ^2. By standard methods the OLS estimate $\widehat{\rho}_n$ of ρ_o in (4.1) has the following limiting distribution

$$n(\widehat{\rho}_n - \rho_o) \to_d \left[\int_0^1 B_X(t) dB_\sigma(t) \right] \left[\int_0^1 B_X^2(t) dt \right]^{-1},$$

where

$$B_X(\cdot) \equiv B_\sigma(\cdot) - X'(\cdot) \left[\int_0^1 X(t) X'(t) dt \right]^{-1} \left[\int_0^1 X(t) B_\sigma(t) dt \right]$$

is the Hilbert space projection residual of $B_\sigma(\cdot)$ on $X(\cdot)$.

Figure 4.1 (from Phillips, 2001) depicts the asymptotic density of $n(\widehat{\rho}_n - \rho_o)$ with different numbers of deterministic (polynomial) trend functions. It is clear that the shape and location of the asymptotic density of $n(\widehat{\rho}_n - \rho_o)$ are both highly sensitive to the trend degree p. This sensitivity implies that critical values of the tests change substantially with the specification of the deterministic trend functions, necessitating the use of different statistical tables according to the precise specification of the fitted model. As a result, if the approach to modelling the time series were such that one contemplated increasing p as the sample size n increased, and to continue to do so as n goes to infinity, then a limit theory in which $p \to \infty$ as $n \to \infty$ may be more appropriate. In fact, even the moderate degree $p \sim 5$ produces very different results from $p = 0, 1$, and the large p asymptotic theory in this case produces a better approximation to the finite sample distribution. Entirely similar considerations apply when the regressor X_t includes trend breaks.

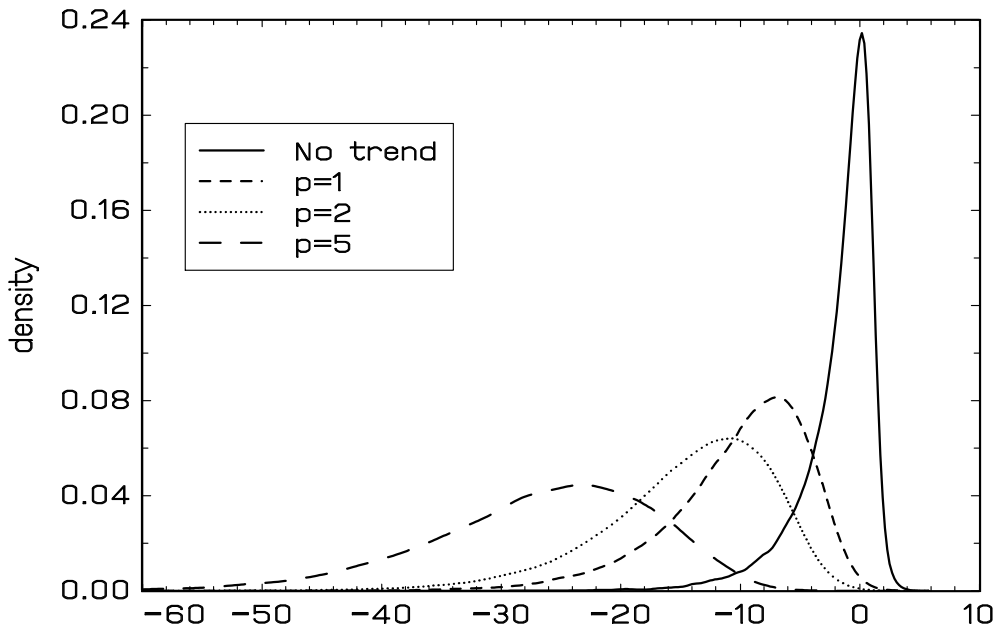

Figure 4.1. Densities of $\int_0^1 B_X(t)dB_\sigma(t) \Big/ \int_0^1 B_X^2(t)dt$ for $X = (1, t, ..., t^p)$

As we have seen in the previous section, the KL representation (2.10) of a stochastic process links the random function $B_\sigma(t)$ ($t \in [a, b]$) with the trend basis functions $\varphi_k(t)$ ($k \in \mathbb{Z}_+$) of the Hilbert space $L^2[a, b]$, thereby enabling us to study the effects of deterministic functions on tests involving the stochastic trends. The present section reviews some of the findings in Phillips (2001), which shows how the asymptotic theory of estimation in unit root models changes when deterministic trends co-exist with the stochastic trend.

Specifically, consider the following typical autoregression with a trend component

$$\frac{1}{\sqrt{n}} Y_t = \frac{\widehat{\rho}_n}{\sqrt{n}} Y_{t-1} + \sum_{k=1}^{K} \widehat{a}_{k,n} \varphi_k \left(\frac{t}{n} \right) + \widehat{u}_{t,K} \tag{4.3}$$

where $\varphi_k(\cdot)$ ($k \in \mathbb{Z}_+$) are trend basis functions, $\widehat{\rho}_n$ and $\widehat{a}_{k,n}$ are the OLS estimates by regressing $n^{-\frac{1}{2}} Y_t$ on the lagged variable $n^{-\frac{1}{2}} Y_{t-1}$ and $\varphi_k\left(\frac{t}{n}\right)$ ($k = 1, ..., K$). The scaling in (4.3) is entirely innocuous and used only to assist in the asymptotics. As is apparent from regression (3.4) and Theorem 3.1, when there is no lagged dependent variable $n^{-\frac{1}{2}} Y_{t-1}$ in (4.3), the fitted value from the trend basis $\sum_{k=1}^{K} \widehat{a}_{k,n} \varphi_k(t)$ reproduces the KL component $\sum_{k=1}^{K} \lambda_k^{\frac{1}{2}} \xi_k \varphi_k(t)$ of the BM limit process of $n^{-\frac{1}{2}} Y_t$ as the sample size $n \to \infty$.

In particular, as the scaled partial sum $n^{-\frac{1}{2}} Y_t$ satisfies the functional central limit

theorem (FCLT) in (4.2), we can invoke (3.2) to deduce that

$$\sup_{0\leq t\leq n}\left|\frac{1}{\sqrt{n}}Y_t-\sum_{k=1}^{\infty}\lambda_k^{\frac{1}{2}}\varphi_k\left(\frac{t}{n}\right)\xi_k\right|=o_p(1). \tag{4.4}$$

From the partitioned regression in (4.3) and the series representation in (4.4) we see that $\widehat{\rho}_n$ is the fitted coefficient in the regression of $n^{-\frac{1}{2}}Y_t$ on the projection residual of $n^{-\frac{1}{2}}Y_{t-1}$ on the trend basis functions $\varphi_k(\cdot)$ $(k=1,...,K)$. The stochastic trend variable Y_{t-1} and the trend basis functions are highly correlated with large K and there is a collinearity problem in the regression (4.3) as $K\to\infty$ because the lagged regressor is perfectly fitted by the trend basis. The asymptotic properties of $\widehat{\rho}_n$ are correspondingly affected by the presence of the deterministic trends and their influence is severe when $K\to\infty$. As a result unit root tests and limit theory based on $\widehat{\rho}_n$ are affected by the presence of deterministic trends, the effects being sufficiently important as to alter the convergence rate. This point is confirmed in the next theorem. First, we have the following Lemma (Phillips, 2001) which shows the effect of a finite number K of deterministic trends on the limit theory of semiparametric Z tests (Phillips, 1987; Phillips and Perron, 1988; and Ouliaris, Park and Phillips, 1988). These tests are either coefficient based (denoted here by $Z_{\rho,n}$) or t-ratio tests (denoted by $Z_{t,n}$). Readers may refer to the above references for their construction.

Lemma 4.1 *Suppose that u_t satisfies Assumption 3.1 and $Y_t=\sum_{s=1}^{t}u_s$. Then the unit root test statistic $Z_{\rho,n}$ and the t-ratio test statistic $Z_{t,n}$ satisfy*

$$Z_{\rho,n}\to_d\frac{\int_0^1 B_{\varphi_K}(r)dB_\sigma(r)}{\int_0^1 B_{\varphi_K}^2(r)dr}\ \textit{and}\ Z_{t,n}\to_d\frac{\int_0^1 B_{\varphi_K}(r)dB_\sigma(r)}{\left[\int_0^1 B_{\varphi_K}^2(r)dr\right]^{\frac{1}{2}}}$$

where $B_{\varphi_K}(\cdot)=B_\sigma(\cdot)-\left[\int_0^1 B_\sigma(r)\Phi_K(r)dr\right]\Phi_K'(\cdot)$.

From the KL representation, we see that

$$\int_0^1 B_{\varphi_K}^2(r)dr=\int_0^1\left[\sum\nolimits_{k=K+1}^{\infty}\lambda_k^{\frac{1}{2}}\varphi_k(r)\xi_k\right]^2dr=\sum\nolimits_{k=K+1}^{\infty}\lambda_k\xi_k^2\to_{a.s.}0\text{ as }K\to\infty$$

which implies that when K is large, the asymptotic distributions of $Z_{\rho,n}$ and $Z_{t,n}$ are materially affected by a denominator that tends to zero and integrand in the numerator that tends to zero. This structure explains why the asymptotic distributions of $Z_{\rho,n}$ and

$Z_{t,n}$ are drawn towards minus infinity with larger K. One may conjecture that when $K \to \infty$, $Z_{\rho,n}$ and $Z_{t,n}$ will diverge to infinity as $\int_0^1 B_{\varphi_K}^2(r)dr \to_p 0$ as $K \to \infty$. This conjecture is confirmed in the following theorem from Phillips (2001).

Theorem 4.1 *Suppose that u_t satisfies Assumption 3.1. If $K \to \infty$ and $K^4/n \to 0$ as $n \to \infty$, then*

$$K^{-\frac{1}{2}}\left(Z_{\rho,n} + \frac{\pi^2 K}{2}\right) \to_d N\left(0, \pi^4/6\right)$$

and

$$Z_{t,n} + \frac{\pi\sqrt{K}}{2} \to_d N\left(0, \pi^2/24\right).$$

When the lagged dependent variable and deterministic trend functions are included in the LS regression to model a stochastic trend, they are seen to jointly compete for the explanation of the stochastic trend in a time series. In such a competition, Theorem 4.1 implies that the deterministic functions will be successful in modelling the trend even in the presence of an autoregressive component. The net effect of including K deterministic functions in the regression is that the rate of convergence to unity of the autoregressive coefficient $\widehat{\rho}_n$ is slowed down. In particular, the theorem implies that $\widehat{\rho}_n = 1 - \frac{\pi^2}{2}\frac{K}{n} + o_p\left(\frac{K}{n}\right) \to_p 1$ as $(n, K \to \infty)$. Thus, $\widehat{\rho}_n$ is still consistent for $\rho = 1$, but has a slower rate of approach to unity than when K is fixed. The explanation for the nonstationarity in the data is then shared between the deterministic trend regressors and the lagged dependent variable.

5 Efficient Estimation of Cointegrated Systems

The trend basis functions in the KL representation (2.10) are deterministic and accordingly independent of any random variables. Moreover, as shown in Theorem 3.2, a stochastic trend can be fully reproduced by its projection on the trend basis functions. These two properties indicate that trend basis functions provide a natural set of valid instrumental variables (IVs) to model stochastic processes that appear as endogenous regressors. This feature of the KL basis functions was pointed out in Phillips (2012), who proposed using trend basis functions as IVs to efficiently estimate cointegrated systems. We outline the essential features of this work in what follows.

Consider the cointegrated system

$$Y_t = A_o X_t + u_{y,t} \tag{5.1}$$

$$\Delta X_t = u_{x,t}, \tag{5.2}$$

where the time series Y_t is $m_y \times 1$ and X_t is $m_x \times 1$ with initial conditions $X_0 = O_p(1)$ at $t = 0$. The composite error $u_t = (u'_{y,t}, u'_{x,t})'$ is a weakly dependent time series generated as a linear process

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \sum_{j=0}^{\infty} j^a \|c_j\| < \infty, a > 3, \tag{5.3}$$

where $\varepsilon_t = iid(0, \Sigma)$ with $\Sigma > 0$ and $E\left[||\varepsilon_t||^p\right] < \infty$ for some $p > 2$ and matrix norm $\|\cdot\|$. The long-run moving average coefficient matrix $C(1)$ is assumed to be nonsingular, so that X_t is a full rank integrated process. Under (5.3), the scaled partial sum $\frac{1}{\sqrt{n}} \sum_{s=0}^{t} u_t$ satisfies the following FCLT

$$\frac{1}{\sqrt{n}} \sum_{s=0}^{\lfloor nt \rfloor} u_t \to_d B_u(t) \equiv \begin{pmatrix} B_y(t) \\ B_x(t) \end{pmatrix}, \tag{5.4}$$

for any $t \in [0, 1]$. The long-run variance matrix $\Omega = C(1)\Sigma C'(1)$ is partitioned conformably with u_t as

$$\Omega = \begin{bmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{bmatrix}.$$

The conditional long-run covariance matrix of u_y on u_x is $\Omega_{yy \cdot x} = \Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}$. In a similar way we define the one-sided long run covariance matrix

$$\Delta = \sum_{j=0}^{\infty} E\left(u_0 u'_{-j}\right) = \begin{bmatrix} \Delta_{yy} & \Delta_{yx} \\ \Delta_{xy} & \Delta_{xx} \end{bmatrix}$$

The rest of this section discusses and compares several different estimates of A_o. The comparison of different estimates helps in understanding the role that trend basis functions play in efficient estimation. For ease of notation and without loss of generality we henceforth assume that X_t and Y_t are scalar random variables. We first consider the OLS estimate of A_o, which is defined as $\widehat{A}_n = \left(\sum_{t=1}^{n} Y_t X'_t\right)\left(\sum_{t=1}^{n} X_t X'_t\right)^{-1}$. Under (5.3) it is easily seen

that

$$n(\widehat{A}_n - A_o) = \frac{n^{-1}\sum_{t=1}^{n} u_{y,t}X_t}{n^{-2}\sum_{t=1}^{n} X_t^2} \to_d \frac{\int_0^1 B_x(t)dB_y(t) + \Delta_{yx}}{\int_0^1 B_x^2(t)dt}$$

where B_x and B_y are defined in (5.4). In view of the contemporaneous and serial correlation between $u_{x,t}$ and $u_{y,t}$ it is well-known that OLS estimation suffers from two sources of high-order bias - endogeneity bias from the corresponding correlation of B_x and B_y and serial correlation bias that manifests in the one sided long run covariance Δ_{yx}.

We next consider the IV estimation of the augmented regression equation with K trend IVs (basis functions) $\varphi_k(\cdot)$ $(k = 1, ..., K)$

$$Y_t = A_o X_t + B_o \Delta X_t + u_{y\cdot x,t}, \tag{5.5}$$

where $B_o = \Omega_{yx}\Omega_{xx}^{-1}$ and $u_{y\cdot x,t} = u_{y,t} - B_o u_{x,t}$. For this model, it is easy to show that the LS estimate of A_o continues to suffer from second order bias effects and the LS estimate of B_o is not generally consistent. On the other hand, the IV estimate of A_o in the augmented equation has optimal properties. It can be written in projection form as

$$\widehat{A}_{IV} = \left(Y' R_{\Delta X,K} X\right)\left(X' R_{\Delta X,K} X\right)^{-1}$$

where $Y' = [Y_1, ..., Y_n]$ with similar definitions for the observation matrices X' and ΔX, the projector $P_K = \Phi_K \left(\Phi_K' \Phi_K\right)^{-1} \Phi_K'$, $\Phi_K = [\Phi_K'(\frac{1}{n}), ..., \Phi_K'(1)]'$, $\Phi_K(\cdot) = [\varphi_1(\cdot), ..., \varphi_K(\cdot)]$ and the composite projector $R_{\Delta X,K} = P_K - P_K \Delta X \left(\Delta X' P_K \Delta X\right)^{-1} \Delta X' P_K$. Similarly, the IV estimate of B_o can be written as

$$\widehat{B}_{IV} = \left(Y' R_{X,K} \Delta X\right)\left(\Delta X' R_{X,K} \Delta X\right)^{-1}$$

where $R_{X,K} = P_K - P_K X \left(X' P_K X\right)^{-1} X' P_K$.[5]

The following Lemma gives the asymptotic distributions of the IV estimates $\widehat{A}_{IV_K,n}$

[5] The trend IV estimate is related to the spectral regression estimates proposed in Phillips (1991b), although those estimates are formulated in the frequency domain. Spectral regression first transfers the cointegration system (5.1) and (5.2) to frequency domain ordinates and then estimates A_o by GLS regression. The spectral transformation projects the whole model on the deterministic function $\exp(i\lambda t)$ at different frequencies $\lambda \in R$, which helps to orthogonalize the projections at different frequencies. However, optimal weights constructed using the empirical spectral density are used in this procedure. Phillips (1991b) also gives a narrow band spectral estimation procedure which uses frequency ordinates in the neighborhood of the origin. Trend IV estimation only projects the (endogenous) regressors on the deterministic functions (trend IVs) and does not need optimal weighting to achieve efficiency. It is more closely related to the narrow band procedure but does not involve frequency domain techniques.

and $\widehat{B}_{IV_K,n}$ when the number of the trend basis functions K is fixed.

Lemma 5.1 *Under the Assumption (5.3), we have*

$$n(\widehat{A}_{IV} - A_o) \to_d \frac{\sum_{k=1}^K \eta_{x,k}^2 \sum_{k=1}^K \xi_{x,k}\eta_{y\cdot x,k} - \sum_{k=1}^K \eta_{x,k}\eta_{y\cdot x,k} \sum_{k=1}^K \xi_{x,k}\eta_{x,k}}{\sum_{k=1}^K \eta_{x,k}^2 \sum_{k=1}^K \xi_{x,k}^2 - \left[\sum_{k=1}^K \xi_{x,k}\eta_{x,k}\right]^2} \tag{5.6}$$

and

$$\widehat{B}_{IV} \to_d B_o + \frac{\sum_{k=1}^K \xi_{x,k}^2 \sum_{k=1}^K \eta_{x,k}\eta_{y\cdot x,k} - \sum_{k=1}^K \xi_{x,k}\eta_{y\cdot x,k} \sum_{k=1}^K \xi_{x,k}\eta_{x,k}}{\sum_{k=1}^K \xi_{x,k}^2 \sum_{k=1}^K \eta_{x,k}^2 - \left[\sum_{k=1}^K \xi_{x,k}\eta_{x,k}\right]^2}, \tag{5.7}$$

where $\eta_{y\cdot x,k} = \int_0^1 \varphi_k(r)dB_{y\cdot x}(r)$, *and* $\xi_{x,k}$, $\eta_{x,k}$, $\eta_{y,k}$ *are defined by*

$$\xi_{x,k} = \int_0^1 \varphi_k(t)B_x(t)dt, \ \eta_{x,k} = \int_0^1 \varphi_k(t)dB_x(t), \ \text{and } \eta_{y,k} = \int_0^1 \varphi_k(t)dB_y(t), \tag{5.8}$$

for all k.

From Lemma 5.1, we see that the IV estimate $\widehat{A}_{IV}$ of A_o in the augmented equation 5.1 is consistent, but it suffers second order bias when the number of the trend basis functions K is fixed. Moreover, the IV estimate $\widehat{B}_{IV}$ of B_o, is not consistent when K is fixed. By Corollary 2.2, we get

$$\xi_{x,k}^2 = \left[\int_0^1 \varphi_k(r)dB_x(r)\right]^2 \stackrel{d}{=} \Omega_{xx}\chi_k^2(1) \text{ for all } k \in \mathbb{Z}_+$$

where Ω_{xx} is the long-run variance of $u_{x,t}$ and $\chi_k^2(1)$ denotes a chi-square random variable with degree of freedom 1. Moreover, $\chi_k^2(1)$ is independent of $\chi_{k'}^2(1)$ for any $k \neq k'$ and $k, k' \in \mathbb{Z}_+$. Using the law of large numbers, we have

$$\frac{1}{K}\sum_{k=1}^K \left[\int_0^1 \varphi_k(r)dB_x(r)\right]^2 \to_{a.s.} \Omega_{xx}. \tag{5.9}$$

Under sequential asymptotics, we see that

$$n(\widehat{A}_{IV} - A_o) = \frac{\sum_{k=1}^K \xi_{x,k}\eta_{y\cdot x,k} + O_p(K^{-1})}{\sum_{k=1}^K \xi_{x,k}^2 + O_p(K^{-1})} \tag{5.10}$$

and

$$\widehat{B}_{IV} = B_o + O_p(K^{-1}). \tag{5.11}$$

Results in (5.10) and (5.11) indicate that when the number of trend IVs diverges to infinity, the IV estimate $\widehat{A}_{IV}$ of A_o may be as efficient as the maximum likelihood (ML) estimate under Gaussianity (Phillips (1991a)) and the IV estimate $\widehat{B}_{IV}$ of B_o may be consistent. These conjectures are justified in Phillips (2012) and shown to hold under joint asymptotics.

Let $\widehat{\Omega}_{K,n} = K^{-1}\left(Y' - \widehat{A}_{IV}X' - \widehat{B}_{IV}\Delta X'\right)P_K\left(Y' - \widehat{A}_{IV}X' - \widehat{B}_{IV}\Delta X'\right)'$ and define $B_{y\cdot x}(t) = B_y(t) - B_oB_x(t)$. The following theorem is from Phillips (2012).

Theorem 5.1 *Under the Assumption (5.3) and the following rate condition,*

$$\frac{1}{K} + \frac{K}{n^{(1-2/p)\wedge(5/6-1/3p)}} + \frac{K^5}{n^4} \to 0 \tag{5.12}$$

as $n \to \infty$, we have

(a) $n(\widehat{A}_{IV} - A_o) \to_d \left[\int_0^1 B_x(t)dB'_{y\cdot x}(t)\right]'\left[\int_0^1 B_x(t)B'_x(t)dr\right]^{-1}$;

(b) $\widehat{B}_{IV} \to_p B_o$;

(c) $\widehat{\Omega}_{K,n} \to_p \Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}$.

Theorem 5.1 implies that the IV estimate $\widehat{A}_{IV}$ is consistent and as efficient as the ML estimate under Gaussian errors (see Phillips, 1991, for the latter). Moreover, the IV estimates of the long-run coefficients are also consistent. It is easy to see that

$$E\left[\varphi_k(t)X_t\right] = \varphi_k(t)E\left[X_t\right] = 0$$

for any $k \in \mathbb{Z}_+$, which implies that trend IVs do not satisfy the relevance condition in the IV estimation literature. As a result, the fact that efficient estimation using trend IVs is possible may appear somewhat magical, especially in view of existing results on IV estimation in stationary systems where relevance of the instruments is critical to asymptotic efficiency and can even jeopardize consistency when the instruments are weak (Phillips, 1989; Staiger and Stock, 1997). Furthermore, the results in Theorem 5.1 make it clear that what is often regarded as potentially dangerous spurious correlation among trending variables can itself be used in a systematic way to produce rather startling positive results.

6 Automated Efficient Estimation of Cointegrated Systems

As illustrated in the previous section, the trend IV approach is very effective in efficient estimation of the cointegration systems. In reality, when the cointegration systems have the triangle representation (5.1) and (5.2), this method is very straightforward and easy to be implemented. However, when the cointegration rank of the cointegrated system is unknown, it is not clear how the trend IV approach can be applied to achieve optimal estimation. Determination of the cointegration rank is important for estimation and inference of cointegrated systems, because under-selected cointegration rank produces inconsistent estimation, while over-selected cointegration rank leads to second order bias and inefficient estimation (c.f., Liao and Phillips, 2010). More recently, Liao and Phillips (2012) proposes an automated efficient estimation method for the cointegrated systems. The new method not only consistently selects the cointegration rank and the lagged differences in general vector error correction models (VECMs) in one-step, but also performs efficient estimation of the cointegration matrix and nonzero transient dynamics simultaneously.

Liao and Phillips (2012) first study the following simple VECM system

$$\Delta Y_t = \Pi_o Y_{t-1} + u_t = \alpha_o \beta_o' Y_{t-1} + u_t \tag{6.1}$$

where $\Pi_o = \alpha_o \beta_o'$ has rank $0 \leq r_o \leq m$, α_o and β_o are $m \times r_o$ matrices with full rank and $\{u_t\}$ is an m-dimensional *iid* process with zero mean and nonsingular covariance matrix Ω_u. The following assumption is imposed on Π_o.

Assumption 6.1 (RR) *(i) The determinantal equation $|I - (I + \Pi_o)\lambda| = 0$ has roots on or outside the unit circle; (ii) the matrix Π_o has rank r_o, with $0 \leq r_o \leq m$; (iii) if $r_o > 0$, then the matrix $R = I_{r_o} + \beta_o' \alpha_o$ has eigenvalues within the unit circle.*

The unknown parameter matrix Π_o is estimated in the following penalized GLS estimation

$$\widehat{\Pi}_{g,n} = \underset{\Pi \in R^{m \times m}}{\arg\min} \left\{ \sum_{t=1}^{n} \|\Delta Y_t - \Pi Y_{t-1}\|^2_{\widehat{\Omega}^{-1}_{u,n}} + \sum_{k=1}^{m} \frac{n\lambda_{r,k,n}}{||\phi_k(\widehat{\Pi}_{1st})||^{\omega}} \|\Phi_{n,k}(\Pi)\| \right\}, \tag{6.2}$$

where $\|A\|^2_B = A'BA$ for any $m \times 1$ vector A and $m \times m$ matrix B, $\widehat{\Omega}_{u,n}$ is some first-step consistent estimator of Ω_u, $\omega > 0$ is some constant, $\lambda_{r,k,n}$ $(k = 1, ..., m)$ are tuning

parameters that directly control the penalization, $||\phi_k(\Pi)||$ denotes the k-th largest modulus of the eigenvalues $\{\phi_k(\Pi)\}_{k=1}^m$ of the matrix Π [6], $\Phi_{n,k}(\Pi)$ is the k-th row vector of $Q_n\Pi$, and Q_n denotes the normalized left eigenvector matrix of $\widehat{\Pi}_{1st}$. The matrix $\widehat{\Pi}_{1st}$ is a first-step (OLS) estimate of Π_o. The penalty functions in (6.2) are constructed based on the so called adaptive Lasso penalty (Zou, 2006) and they play the role of selecting the cointegrating rank in the penalized estimation. More importantly, if the cointegration rank is simultaneously determined in the estimation of Π_o, the selected rank structure will be automatically imposed on the penalized GLS estimate $\widehat{\Pi}_{g,n}$. As a result, $\widehat{\Pi}_{g,n}$ would be automatically efficient if the true cointegration rank could be consistently selected in the penalized GLS estimation (6.2).

The asymptotic properties of the penalized GLS estimate are given in the following theorem from Liao and Phillips (2012).

Theorem 6.1 (Oracle Properties) *Suppose Assumption 6.1 hold. If $\widehat{\Omega}_{u,n} \to_p \Omega_u$ and the tuning parameter satisfies $n^{\frac{1}{2}}\lambda_{r,k,n} = o(1)$ and $n^{\omega}\lambda_{r,k,n} \to \infty$ for $k = 1, ..., m$, then as $n \to \infty$,*

$$\Pr\left(rank(\widehat{\Pi}_{g,n}) = r_o\right) \to 1 \tag{6.3}$$

where $rank(\widehat{\Pi}_{g,n})$ denotes the rank of $\widehat{\Pi}_{g,n}$. Moreover $\widehat{\Pi}_{g,n}$ has the same limit distribution as the reduced rank regression (RRR) estimator which assumes the true rank r_o is known.

Theorem 6.1 shows that if the tuning parameters $\lambda_{r,k,n}$ $(k = 1, ..., m)$ converge to zero at certain rate, then the consistent cointegration selection and the efficient estimation can be simultaneously achieved in the penalized GLS estimation (6.2). Specifically, the tuning parameter $\lambda_{r,k,n}$ $(k = 1, ..., m)$ should converge to zero faster than $\sqrt{n}$ so that when $\Pi_o \neq 0$, the convergence rate of $\widehat{\Pi}_{g,n}$ is not slower than root-n. On the other hand, $\lambda_{r,k,n}$ should converge to zero slower than $n^{-\omega}$ so that the cointegration rank r_o is selected with probability approaching one.

The *iid* assumption on u_t ensures that Π_o is consistently estimated, which is usually required for consistent model selection in the Lasso model selection literature. But Cheng and Phillips (2009, 2012) showed that the cointegration rank r_o can be consistently selected by information criteria even when u_t is weakly dependent, in particular when u_t satisfies

[6] For any $m \times m$ matrix Π, we order the eigenvalues of Π in decreasing order by their moduli, i.e. $|\phi_1(\Pi)| \geq |\phi_2(\Pi)| \geq ... \geq |\phi_m(\Pi)|$. For complex conjugate eigenvalues, we order the eigenvalue a positive imaginary part before the other.

conditions such as LP below. We therefore anticipate that similar properties hold for Lasso estimation.

Assumption 6.2 (LP) *Let $D(L) = \sum_{j=0}^{\infty} D_j L^j$, where $D_0 = I_m$ and $D(1)$ has full rank. Let u_t have the Wold representation*

$$u_t = D(L)\varepsilon_t = \sum_{j=0}^{\infty} D_j \varepsilon_{t-j}, \text{ with } \sum_{j=0}^{\infty} j^{\frac{1}{2}} ||D_j|| < \infty, \tag{6.4}$$

where ε_t is iid $(0, \Sigma_{\varepsilon\varepsilon})$ with $\Sigma_{\varepsilon\varepsilon}$ positive definite and finite fourth moments.

It is clear that under Assumption 6.2, Π_o can not be consistently estimated in general. As a result, the probability limit of the GLS estimate of Π_o may have rank smaller or larger than r_o. However, Liao and Phillips (2012) show that the cointegration rank r_o can be consistently selected by penalized estimation as in (6.2) even when u_t is weakly dependent and Π_o is not consistently estimated, thereby extending the consistent rank selection result of Cheng and Phillips (2009) to Lasso estimation.

Theorem 6.2 *Under Assumption LP, if $n^{\frac{1+\omega}{2}} \lambda_{r,k,n} = o(1)$ and $n^{\frac{1}{2}} \lambda_{r,k,n} = o(1)$ for $k = 1, ..., m$, then we have*

$$\Pr\left(rank(\widehat{\Pi}_{g,n}) = r_o\right) \to 1 \text{ as } n \to \infty. \tag{6.5}$$

Theorem 6.2 states that the true cointegration rank r_o can be consistently selected, even though the matrix Π_o is not consistently estimated. Moreover, even when the probability limit Π_1 of the penalized GLS estimator has rank less than r_o, Theorem 6.2 ensures that the correct rank r_o is selected in the penalized estimation. This result is new in the Lasso model selection literature as Lasso techniques are usually advocated because of their ability to shrink small estimates (in magnitude) to zero in penalized estimation. However, Theorem 6.2 shows that penalized estimation here does not shrink the estimates of the extra $r_o - r_1$ zero eigenvalues of Π_1 to zero.

Liao and Phillips (2012) also study the general VECM model

$$\Delta Y_t = \Pi_o Y_{t-1} + \sum_{j=1}^{p} B_{o,j} \Delta Y_{t-j} + u_t \tag{6.6}$$

with simultaneous cointegration rank selection and lag-order selection. To achieve consistent lag-order selection, the model in (6.6) has to be consistently estimable. Thus, we assume that given p in (6.6), the error term u_t is an m-dimensional *iid* process with zero mean and nonsingular covariance matrix Ω_u. Define

$$C(\phi) = \Pi_o + \sum_{j=0}^{p} B_{o,j}(1-\phi)\phi^j, \text{ where } B_{o,0} = -I_m.$$

The following assumption extends Assumption 6.1 to accommodate the general structure in (6.6).

Assumption 6.3 (RR) *(i) The determinantal equation $|C(\phi)| = 0$ has roots on or outside the unit circle; (ii) the matrix Π_o has rank r_o, with $0 \le r_o \le m$; (iii) the $(m-r_o)\times(m-r_o)$ matrix*

$$\alpha'_{o,\perp}\left(I_m - \sum_{j=1}^{p} B_{o,j}\right)\beta_{o,\perp} \tag{6.7}$$

is nonsingular, where $\alpha_{o,\perp}$ and $\beta_{o,\perp}$ are the orthonormal complements of α_o and β_o respectively.

The unknown parameters (Π_o, B_o) are estimated by penalized GLS estimation

$$\begin{aligned}(\widehat{\Pi}_{g,n}, \widehat{B}_{g,n}) \quad = \quad & \underset{\Pi, B_1, \ldots, B_p \in R^{m\times m}}{\arg\min} \left\{ \sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j=1}^{p} B_j \Delta Y_{t-j} \right\|^2_{\widehat{\Omega}^{-1}_{u,n}} \right. \\ & \left. + \sum_{j=1}^{p} \frac{n\lambda_{b,j,n}}{||\widehat{B}_{j,1st}||^{\omega}} \|B_j\| + \sum_{k=1}^{m} \frac{n\lambda_{r,k,n}}{||\phi_k(\widehat{\Pi}_{1st})||^{\omega}} \|\Phi_{n,k}(\Pi)\| \right\}\end{aligned} \tag{6.8}$$

where $\lambda_{b,j,n}$ and $\lambda_{r,k,n}$ ($j = 1, ..., p$ and $k = 1, ..., m$) are tuning parameters, $\widehat{B}_{j,1st}$ and $\widehat{\Pi}_{1st}$ are some first step (OLS) estimates of $B_{o,j}$ and Π_o ($j = 1, ..., p$) respectively. Denote the index set of the zero components in B_o as $\mathcal{S}^c_B$ such that $\|B_{o,j}\| = 0$ for all $j \in \mathcal{S}^c_B$ and $\|B_{o,j}\| \neq 0$ otherwise. The asymptotic properties of the penalized GLS estimates $(\widehat{\Pi}_{g,n}, \widehat{B}_{g,n})$ are presented in the following theorem from Liao and Phillips (2012).

Theorem 6.3 *Suppose that Assumption 6.3 is satisfied and $\widehat{\Omega}_{u,n} \to_p \Omega_u$. If $n^{\frac{1}{2}}(\lambda_{r,k,n} +$*

$\lambda_{b,j,n}) = O(1)$, $n^{\omega}\lambda_{r,k,n} \to \infty$ *and* $n^{\frac{1+\omega}{2}}\lambda_{b,j,n} \to \infty$ *(*$k = 1, ..., m$ *and* $j = 1, ..., p$*), then*

$$\Pr\left(r(\widehat{\Pi}_{g,n}) = r_o\right) \to 1 \text{ and } \Pr\left(\widehat{B}_{g,j,n} = 0\right) \to 1 \tag{6.9}$$

for $j \in \mathcal{S}_B^c$ *as* $n \to \infty$*; moreover* $\widehat{\Pi}_{g,n}$ *and the penalized GLS estimate of the nonzero components in* B_o *have the same joint limiting distribution as that of the general RRR estimate which assumes the true rank* r_o *and true zero components in* B_o *are known.*

From Theorem 6.1 and Theorem 6.3, we see that the tuning parameter plays an important role in ensuring that the penalized estimate is efficient and the true model is consistently selected in penalized GLS estimation. In empirical applications, the conditions stated in these two theorems do not provide a clear suggestion of how to select the tuning parameters. In the Lasso literature the tuning parameters are usually selected by cross-validation or information criteria methods. However, such methods of selecting the tuning parameter are computationally intensive and they do not take the finite sample properties of the penalized estimates into account. Liao and Phillips (2012) provide a simple data-driven tuning parameter selection procedure based on balancing first order conditions that takes both model selection and finite sample properties of the penalized estimates into account. The new method is applied to model GNP, consumption and investment using US data, where there is obvious co-movement in the series. The results reveal the effect of this co-movement through the presence of two cointegrating vectors, whereas traditional information criteria fail to find co-movement and set the cointegrating rank to zero for these data.

7 Series Estimation of the Long-Run Variance

Previous sections have shown how the long-run behavior of integrated processes can be fully reproduced in the limit by simple linear projections on trend basis functions. Motivated by this result, we are concerned to ask the following questions. First, let $\{u_t\}$ be a stationary process and $\{\varphi_k(\cdot)\}_k$ be a set of trend basis functions. What are the asymptotic properties of the projection of $\{u_t\}_{t=1}^n$ on $\varphi_k(\cdot)$ with a fixed number K of basis functions? Further, what are the asymptotic properties of this projection when the number of basis functions goes to infinity?

As first observed in Phillips (2005b), such projections produce consistent estimates

of the long-run variance (LRV) of the process $\{u_t\}$, when K goes to infinity with the sample size. This large K asymptotic theory justifies the Gaussian approximation of t-ratio statistics and Chi-square approximations of Wald statistics in finite samples. More recently, Sun (2011, 2012) showed that when K is fixed, t-ratio statistics have an asymptotic student-t distribution and Wald statistics have asymptotic F distributions. The fixed-K asymptotic theory is argued in Sun (2012) to provide more accurate size properties for both t-ratio and Wald statistics in finite samples.

Formally, suppose that the process $\{u_t\}$ satisfies the following assumption.

Assumption 7.1 *For all $t \geq 0$, u_t has Wold representation*

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \ \sum_{j=0}^{\infty} j^a \left|c_j\right| < \infty \ , \ C(1) \neq 0 \ \textit{and} \ a > 3 \tag{7.1}$$

with $\varepsilon_t = iid(0, \sigma_\varepsilon^2)$ with $E\left(|\varepsilon_t|^p\right) < \infty$ for some $p > 2$.

Under Assumption 7.1.(i), the scaled partial sum $n^{-\frac{1}{2}} \sum_{i=1}^{t} u_i$ satisfies the following FCLT

$$B_n(\cdot) \equiv \frac{\sum_{i=1}^{[n\cdot]} u_i}{\sqrt{n}} \rightarrow_d B_\omega(\cdot) \text{ as } n \rightarrow \infty \tag{7.2}$$

where $B_\omega(\cdot)$ is a BM with variance $\omega^2 = \sigma_\varepsilon^2 C^2(1)$. Note that ω^2 is the LRV of the process $\{u_t\}$.

The projection of $\{u_t\}_{t=1}^n$ on $\varphi_k(\frac{t}{n})$ for some $k \in \mathbb{Z}_+$ can be written as

$$\left[\sum_{t=1}^{n} \varphi_k^2(\frac{t}{n})\right]^{-1} \sum_{t=1}^{n} \varphi_k(\frac{t}{n}) u_t,$$

where

$$\sum_{t=1}^{n} \varphi_k(\frac{t}{n}) u_t \rightarrow_d \int_0^1 \varphi_k(r) dB_\omega(r) \text{ as } n \rightarrow \infty \tag{7.3}$$

by standard functional limit arguments and Wiener integration, and

$$\frac{1}{n} \sum_{t=1}^{n} \varphi_k^2(\frac{t}{n}) \rightarrow \int_0^1 \varphi_k^2(r) dr = 1 \text{ as } n \rightarrow \infty \tag{7.4}$$

by the integrability and normalization of $\varphi_k(\cdot)$. From the results in (7.3) and (7.4), we

deduce that

$$\sqrt{n}\frac{\sum_{t=1}^{n}\varphi_k(\frac{t}{n})u_t}{\sum_{t=1}^{n}\varphi_k^2(\frac{t}{n})} \to_d \int_0^1 \varphi_k(r)dB_\omega(r) \text{ as } n \to \infty.$$

By Corollary 2.2, $\int_0^1 \varphi_k(r)dB_\omega(r) \stackrel{d}{=} N(0,\omega^2)$ and for any $k \neq k'$, the two random variables $\int_0^1 \varphi_k(r)dB_\omega(r)$ and $\int_0^1 \varphi_{k'}(r)dB_\omega(r)$ are independent with each other. These results motivate us to define the following orthonormal series estimate of the LRV

$$\omega_{K,n}^2 = \frac{1}{K}\sum_{k=1}^{K}\left[n^{-\frac{1}{2}}\sum_{t=1}^{n}\varphi_k(\frac{t}{n})u_t\right]^2, \tag{7.5}$$

which leads to the following t-ratio test statistic

$$t_{K,n} = B_n(1)/\sqrt{\omega_{K,n}^2}. \tag{7.6}$$

Lemma 7.1 *Suppose that Assumption 7.1 is satisfied and the number K of trend basis functions are fixed. Then the series LRV estimate defined in (7.5) satisfies*

$$\omega_{K,n}^2 \to_d \frac{\omega^2}{K}\chi^2(K) \tag{7.7}$$

where $\chi^2(K)$ is a chi-square random variable with degrees of freedom K. Moreover, the t-ratio test statistic defined in (7.6) satisfies

$$t_{K,n} \to_d t_K \tag{7.8}$$

where t_K is a student-t random variable with degree of freedom K.

While Lemma 7.1 applies to univariate processes, it is readily extended to the case where $\{u_t\}$ is a multiple time series. In that case, the series LRV estimate is defined as

$$\omega_{K,n}^2 = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{t=1}^{n}\varphi_k(\frac{t}{n})u_t\sum_{t=1}^{n}\varphi_k(\frac{t}{n})u_t'}{n}$$

and the Wald-type test is defined as

$$W_{K,n} = B_n(1)'\left(\omega_{K,n}^2\right)^{-1}B_n(1).$$

Then using similar arguments to those in the proof of Lemma 7.1, we obtain

$$\frac{K-d_u+1}{Kd_u}W_{K,n} \to_d \mathcal{F}_{d_u,K-d_u+1},$$

where $\mathcal{F}_{d_u,K-d_u+1}$ is a F random variable with degrees of freedom $(d_u, K-d_u+1)$ and d_u denotes the dimensionality of the vector u_t.

The weak convergence in (7.7) implies that when the number of the trend basis functions is fixed, the series LRV estimate $\omega^2_{K,n}$ is not a consistent estimate of ω^2. However, the weak convergence in (7.8) indicates that the t-ratio test statistic is asymptotically pivotal. Using sequential asymptotic arguments, we see from (7.7) that when K goes to infinity, $\chi^2(K)/K$ converges to 1, which implies that $\omega^2_{K,n}$ may be a consistent estimate of ω^2 with large K. Similarly, from (7.7), we see that $t_{K,n}$ has an asymptotic Gaussian distribution under sequential asymptotics. These sequential asymptotic results provide intuition about the consistency of $\omega^2_{K,n}$ when K goes to infinity, as well as intuition concerning the improved size properties of the fixed K asymptotics in finite samples.

The following theorem from Phillips (2005b), which was proved using trend basis functions of the form (2.15) but which holds more generally, shows that $\omega^2_{K,n}$ is indeed a consistent estimate of ω^2 under the joint asymptotics framework.

Theorem 7.1 *Let $\gamma_u(\cdot)$ denote the autocovariance function of the process $\{u_t\}$. Suppose that Assumption 7.1 holds and the number of trend basis functions K satisfies*

$$\frac{n}{K^2}+\frac{K}{n}\to 0. \tag{7.9}$$

Then

(a) $\lim_{n\to\infty}\frac{n^2}{K^2}E\left(\omega^2_{K,n}-\omega^2\right)=-\frac{\pi^2}{6}\sum_{h=-\infty}^{\infty}h^2\gamma_u(h)$;

(b) if $K=o(n^{4/5})$, *then* $\sqrt{K}\left(\omega^2_{K,n}-\omega^2\right)\to_d N(0,2\omega^4)$;

(c) if $K^5/n^4\to 1$, *then* $\frac{n^4}{K^4}E\left(\omega^2_{K,n}-\omega^2\right)^2=\frac{\pi^4}{36}\left[\sum_{h=-\infty}^{\infty}h^2\gamma_u(h)\right]^2+2\omega^4$.

Theorem 7.1.(a) implies that $\omega^2_{K,n}$ has bias of order K^2/n^2 as shown in

$$E\left[\omega^2_{K,n}\right]=\omega^2-\frac{K^2}{n^2}\left[\frac{\pi^2}{6}\sum_{h=-\infty}^{\infty}h^2\gamma_u(h)+o(1)\right].$$

From (b), the variance of $\omega^2_{K,n}$ is of $O(K^{-1})$. Thus, given the sample size n, increases

in the number of the trend basis functions K increases bias and reduces variance. The situation is analogous to bandwidth choice in kernel estimation.

The process $\{u_t\}$ studied above is assumed to be known. For example, u_t could be a function of data Z_t and some known parameter θ_o, i.e. $u_t = f(Z_t, \theta_o)$. However, in applications, usually we have to estimate the LRV of the process $\{f(Z_t, \theta_o)\}_t$, where θ_o is unknown but for which a consistent estimate $\widehat{\theta}_n$ may be available. As an illustration, in the rest of this section we use Z-estimation with weakly dependent data to show how the series LRV estimate can be used to conduct auto-correlation robust inference.

The Z-estimate $\widehat{\theta}_n$ can be defined as

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} m(Z_t, \widehat{\theta}_n) = o_p(\varepsilon_n)$$

where $m(\cdot,\cdot) : R^{d_z} \times R^{d_\theta} \to R^{d_\theta}$ is a measurable function and ε_n is a $o(1)$ sequence. Let $M(\theta) = E\left[m(Z, \theta)\right]$. The following assumptions are convenient for the following development and exposition.

Assumption 7.2 *(i) $M(\theta)$ is continuous differentiable in the local neighborhood of θ_o and $\frac{\partial M(\theta_o)}{\partial \theta'}$ has full rank; (ii) the Z-estimate $\widehat{\theta}_n$ is root-n normal, i.e.*

$$\sqrt{n}(\widehat{\theta}_n - \theta_o) \to_d N\left(0, M_-(\theta_o)V(\theta_o)M'_-(\theta_o)\right)$$

where $M_-(\theta_o) = \left[\frac{\partial M(\theta_o)}{\partial \theta'}\right]^{-1}$ and $V(\theta_o) = \lim_{n\to\infty} Var\left[n^{-\frac{1}{2}} \sum_{t=1}^{n} m(Z_t, \theta_o)\right]$; (iii) let N_n denote some shrinking neighborhood of θ_o, then

$$\sup_{\theta \in \mathcal{N}_n} n^{-\frac{1}{2}} \sum_{t=1}^{n} \phi_k(\frac{t}{n}) \left\{m(Z_t, \theta) - m(Z_t, \theta_0) - E\left[m(Z_t, \theta) - m(Z_t, \theta_0)\right]\right\} = o_p(1);$$

(iv) the following FCLT holds

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \phi_k(\frac{t}{n}) m(Z_t, \theta_0) \to_d \int_0^1 \phi_k(r) dB_m(r) \text{ for } k = 1, ..., K,$$

where $B_m(\cdot)$ denotes a vector BM with variance-covariance matrix $V(\theta_o)$; (v) we have

$$M_{+,n}(\widehat{\theta}_n) \equiv n^{-1} \sum_{t=1}^{n} \frac{\partial m(Z_t, \widehat{\theta}_n)}{\partial \theta'} \to_p M_-^{-1}(\theta_o).$$

The conditions in Assumption 7.2 are mild and easy to verify. The series LRV estimate is defined as

$$V_{K,n}(\widehat{\theta}_n) = \frac{1}{K}\sum_{k=1}^{K}\Lambda_{k,n}\Lambda_{k,n}', \tag{7.10}$$

where $\Lambda_{k,n} \equiv \sum_{t=1}^{n}\phi_k(\frac{t}{n})m(Z_t,\widehat{\theta}_n)$ $(k = 1, ..., K)$. Under Assumption 7.2, we have the following lemma, which generalizes Lemma 7.1 to vector stochastic processes with unknown parameters.

Lemma 7.2 *Suppose that the number of the trend basis functions K is fixed and the basis functions satisfy $\int_0^1 \phi_k(r)dr = 0$ $(k = 1, ..., K)$. Then under Assumption 7.1 and Assumption 7.2, we have*

$$\begin{aligned} \mathcal{F}_n &\equiv (\widehat{\theta}_n - \theta_o)' M_{+,n}(\widehat{\theta}_n) V_{K,n}^{-1}(\widehat{\theta}_n) M_{+,n}(\widehat{\theta}_n)(\widehat{\theta}_n - \theta_o)/d_\theta \\ &\to_d \frac{K}{K - d_\theta + 1}\mathcal{F}_{d_\theta, K-d_\theta+1}, \end{aligned}$$

where $\mathcal{F}_{d_\theta,K-d_\theta+1}$ is a F random variable with degree of freedom $(d_\theta, K - d_\theta + 1)$ and d_θ denotes the dimensionality of θ_o.

Lemma 7.2 shows that when the number of the trend basis functions K is fixed, the series LRV estimate $V_{K,n}(\widehat{\theta}_n)$ is inconsistent, but the Wald-type test statistic $\mathcal{F}_n$ is asymptotically pivotal. Autocorrelation robust inference about θ_o can be conducted using the statistic $\mathcal{F}_n^* \equiv (K - d_\theta + 1)\mathcal{F}_n/K$ and the asymptotic $\mathcal{F}_{d_\theta,K-d_\theta+1}$ distribution. As noted in Sun (2012), the restriction $\int_0^1 \phi_k(r)dr = 0$ $(k = 1, ..., K)$ helps to remove the estimation effect in $\widehat{\theta}_n$ from the asymptotic distribution of $V_{K,n}(\widehat{\theta}_n)$. As a result, the statistic $\mathcal{F}_n^*$ enjoys an exact asymptotic F-distribution. Using similar arguments to those in Phillips (2005b), it can be shown that under some suitable rate condition on K the series LRV estimate $V_{K,n}(\widehat{\theta}_n)$ is consistent, i.e.

$$V_{K,n}(\widehat{\theta}_n) = \frac{1}{K}\sum_{k=1}^{K}\Lambda_{k,n}\Lambda_{k,n}' \to_p V(\theta_o),$$

as $n, K \to \infty$ jointly. In that case, the test statistic $\mathcal{F}_n$ has an asymptotic chi-square distribution with d_θ degrees of freedom.

Orthonormal series LRV estimates are becoming increasingly popular for autocorrelation robust inference in econometric models. Sun (2011) proposed a new testing procedure

for hypotheses on deterministic trends in a multivariate trend stationary model, where the LRV is estimated by the series method. For empirical applications, the paper provides an optimal procedure for selecting K in the sense that the type II error is minimized while controlling for the type I error. Sun (2012) uses a series LRV estimate for autocorrelation robust inference in parametric M-estimation. This paper also shows that critical values from the fixed-K limit distribution of the Wald-type test statistic are second-order correct under conventional increased-smoothing asymptotics. Sun and Kim (2012a,b) use the series LRV estimate for inference and specification testing in a generalized method of moments (GMM) setting. The series LRV estimate has also been used in inference for semi/nonparametric econometric models with dependent data. In particular, recent work of Chen, Hahn and Liao (2011) uses the series method to estimate the LRV of a two-step GMM estimate when there are some infinite dimensional parameters estimated by first-step sieve M-estimation. In related work, Chen, Liao and Sun (2012) use series methods to estimate the LRVs of sieve estimates of finite dimensional and infinite dimensional parameters in semi/nonparametric models with weakly dependent data.

8 Concluding Remarks

As explained in previous sections, the KL representation of stochastic processes can be very useful in modelling, estimation, and inference in econometrics. This chapter has outlined the theory behind the KL representation and some of its properties. The applications of the KL representation that we have reviewed belong to three categories:

(i) The link between stochastic trends and their deterministic trend representations. This link is a powerful tool for understanding the relationships between the two forms of trend and the implications of these relationships for practical work. As we have discussed, the KL representation provides new insights that help explain spurious regressions as a natural phenomena when an integrated or near integrated process is regressed on a set of deterministic trend variables. And the representation helps to demonstrate the effect of adding deterministic trends or trend breaks to regressions in which unit root tests are conducted;

(ii) The KL representation reveals that traditional warnings of spurious regressions as uniformly harmful is unjustified. For example, as recovered in its KL representation,

an integrated process can be perfectly modelled by trend basis functions. This relation, which in traditional theory is viewed as a spurious regression, turns out to be extremely useful in the efficient estimation of the cointegrated systems as discussed in section 5;

(iii) Trend basis functions may be used to fit stationary processes, leading to a novel LRV estimation method that is simple and effective because of the natural focus on long run behavior in the trend basis. The resulting series LRV estimate is automatically positive definite and is extremely easy to compute. Moreover, t-ratio and Wald-type test statistics constructed using the series LRV estimate are found to have standard limit distributions under both fixed-K and large-K asymptotics. These features make the use of series LRV estimation attractive for practical work in econometrics, as discussed in section 7.

There are many potential research directions that seem worthy of future research. We mention some of these possibilities in what follows.

First, KL representations of non-degenerate or full rank stochastic processes[7] are discussed in this chapter. It would be interesting to study KL forms of vector processes which are of deficient rank, such as multiple time series that are cointegrated. Phillips (2005a) gives some discussion of this idea and introduces the concept of coordinate cointegration in this context, which subsumes the usual cointegration concept. In this context trend basis functions may be useful in testing for co-movement and efficient estimation of co-moving systems when system rank is unknown.

Second, trend basis representations of different stochastic processes differ. Such differences may be used to test if observed data are compatible with a certain class of stochastic processes. For example, one may be interested in testing a BM null against an O-U process alternative. From section 2, we known that BM has the following KL representation

$$B(t) = \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin\left[(k-1/2)\,\pi t\right]}{(k-1/2)\,\pi} \xi_{\omega,k} \tag{8.1}$$

where $\xi_{\omega,j}$ are *iid* $N(0,\omega^2)$ and ω^2 is the variance of $B(\cdot)$. Using the above representation and the expression in (2.22), we obtain the following alternate representation of an O-U

[7] A full rank or non-degenerate process refers to a random sequence which upon scaling satisfies a functional law with a non-degenerate limit process, such as a Brownian motion with positive definite variance matrix.

process (c.f. Phillips, 1998)

$$
\begin{aligned}
J_c(t) &= \sqrt{2}\sum_{k=1}^{\infty}\frac{\xi_{\omega,k}}{(k-1/2)\,\pi}\left\{\sin\left[(k-1/2)\,\pi t\right]+c\int_0^t e^{(t-s)c}\sin\left[(k-1/2)\,\pi s\right]ds\right\} \\
&= \sqrt{2}\sum_{k=1}^{\infty}\frac{\xi_{\omega,k}}{(k-1/2)^2\pi^2+c^2}\left\{ce^{ct}-c\cos\left[(k-1/2)\,\pi t\right]\right. \\
&\quad \left.+(k-1/2)\,\pi\sin\left[(k-1/2)\,\pi t\right]\right\}. \qquad (8.2)
\end{aligned}
$$

If the data $\{X_t\}$ are projected on the trend IVs $\left\{\sin\left[\left(k-\frac{1}{2}\right)\frac{\pi t}{n}\right],\ \cos\left[\left(k-\frac{1}{2}\right)\frac{\pi t}{n}\right]:k\leq K\right\}$, then under the null, the projection will reproduce the representation in (8.1) when $K\to\infty$. However, under the alternative, as is apparent from (8.2), the projection has an asymptotic form that is very different from (8.1) and includes the cosine and exponential functions. It is of interest to see if significance tests on the coefficients in this regression can usefully discriminate integrated and locally integrated processes which have BM and O-U process limits after standardization.

Third, although trend basis functions are effective in modeling integrated processes and can be used to efficiently estimate cointegration systems, in finite samples it is not clear how many trend basis functions should be used. From the KL representation of BM in (8.1), it is apparent that the trend IVs $\{\sqrt{2}\sin\left[(k-1/2)\,\pi t\right]\}_k$ have a natural ordering according to the variances of their random coefficients $\{\frac{\xi_{\omega,k}}{(k-1/2)\pi}\}_{k=1}^{\infty}$. This ordering is useful in itself for selecting trend IVs, but it would also be useful to calculate the asymptotic mean square error (AMSE) of the trend IV estimate. Then an optimal IV selection criterion could be based on minimizing the empirical AMSE. However, calculation of the AMSE is complicated by the mixed normal limit theory of trend IV estimates and the presence of functional limits in the first order asymptotics, so explicit formulae are not presently available.

In other recent work Liao and Phillips (2011) propose to select trend IVs using Lasso penalized estimation. In particular, in the notation of section 6 of the present paper, trend IVs can be selected by means of the following penalized LS regression

$$
\min_{\Pi\in R^{K\times 2m_x}}\left\|Z_n-\Phi_K\Pi\right\|^2+n\lambda_n\sum_{k=1}^{K}\left\|\Pi_k\right\|, \qquad (8.3)
$$

where $Z_n' = [n^{-\frac{1}{2}}X_1, ..., n^{-\frac{1}{2}}X_n]$, Π_k denotes the k-th row ($k = 1, ..., K$) of the $K \times m_x$ coefficient matrix Π and λ_n is a tuning parameter. The coefficient vector Π_k is related to the k-th trend IV $\varphi_k(\cdot)$ and if Π_k is estimated as zero, then the k-th trend IV $\varphi_k(\cdot)$ would not be used as an instrument for the "endogenous" variable Z. The tuning parameter λ_n determines the magnitude of the shrinkage effect on the estimator of Π_k. The larger the tuning parameter λ_n is, the larger the shrinkage effect will be, leading to more zero coefficient estimates in Π_k. In consequence, the problem of trend IV selection becomes a problem of selecting the tuning parameter λ_n. Liao and Phillips (2011) provide data-driven tuning parameters in the penalty function, making Lasso IV selection fully adaptive for empirical implementation.

Fourth, as noted in Phillips (2005a), the KL representation, when restricted to a subinterval of $[0, 1]$ such as $[0, r]$ ($r \in (0, 1)$), is useful in studying the evolution of a trend process over time. For example, the KL representation of BM on $[0, r]$ has the following form

$$B(s) = \sum_{k=1}^{\infty} \varphi_k\left(\frac{s}{r}\right)\eta_k(r) \text{ for any } s \in [0, r], \tag{8.4}$$

where $\eta_k(r) = r^{-1}\int_0^r B(s)\varphi_k\left(\frac{s}{r}\right)ds$. It follows that $B(r) = \sum_{k=1}^{\infty}\varphi_k(1)\eta_k(r)$, where $B(r)$ and $\eta_k(r)$ are both measurable with respect to the natural filtration $\mathcal{F}_r$ of the BM $B(\cdot)$. The process $\eta_k(r)$ describes the evolution over time of the coefficient of the coordinate basis $\varphi_k(\cdot)$. The evolution of these trend coordinates can be estimated by recursively regressing the sample data on the functions $\varphi_k(\cdot)$ and the resulting estimates deliver direct information on how individual trend coordinates have evolved over time.

The restricted KL representation in (8.4) may also be used for forecasting. In particular, setting $s = r$ in (8.4), the optimal predictor of $B(r)$ given $\mathcal{F}_p$ and coordinates up to K is

$$E\left[B(r)|\,\mathcal{F}_p, K\right] = \sum_{k=1}^{K}\varphi_k(1)E\left[\eta_k(r)|\,\mathcal{F}_p\right]. \tag{8.5}$$

By the definition of $\eta_k(\cdot)$ and using explicit formulae for φ_k, the conditional expectation in (8.5) can be written as

$$E\left[\eta_k(r)|\,\mathcal{F}_p\right] = \frac{1}{r}\int_0^p B(s)\varphi_k\left(\frac{s}{r}\right)ds + B(p)\frac{\sqrt{2}\cos\left[(k-1/2)\frac{\pi p}{r}\right]}{(k-1/2)\pi}. \tag{8.6}$$

Summing over $k = 1, ..., K$, we get

$$E\left[B(r)|\,\mathcal{F}_p, K\right] = \sum_{k=1}^{K} \varphi_k(1) \left[\frac{1}{r}\int_0^p B(s)\varphi_k\left(\frac{s}{r}\right) ds + \frac{\sqrt{2}\cos\left[\left(k-\frac{1}{2}\right)\frac{\pi p}{r}\right] B(p)}{\left(k-\frac{1}{2}\right)\pi}\right]. \quad (8.7)$$

Let $N = [np]$ and $N + h = [nr]$ so that (8.6) and (8.7) effectively provide h-step ahead optimal predictors of these components. $E\left[\eta_k(r)|\,\mathcal{F}_p\right]$ may be estimated from sample data by

$$\widehat{\eta}_k(r|p) = \sum_{t=1}^{N} \frac{X_t/\sqrt{n}}{N+h}\varphi_k(\frac{t}{N+h}) + \frac{\sqrt{2}\cos\left[(k-1/2)\,\frac{\pi N}{N+h}\right] X_N/\sqrt{n}}{(k-1/2)\,\pi}$$

which leads to the following h-step ahead predictor of the trend in the data

$$\widehat{X}_{N+h,N} = \sum_{k=1}^{K} \varphi_k(1) \left[\sum_{t=1}^{N} \frac{X_t}{N+h}\varphi_k(\frac{t}{N+h}) + \frac{\sqrt{2}\cos\left[(k-1/2)\,\frac{\pi N}{N+h}\right] X_N}{(k-1/2)\,\pi}\right].$$

As pointed out in Phillips (2005a), this forecasting approach can be pursued further to construct formulae for trend components and trend predictors corresponding to a variety of long run models for the data. Such formulae enable trend analysis and prediction in a way that captures the main features of the trend for K small and which can be related back to specific long term predictive models for large K. The approach therefore helps to provide a foundation for studying trends in a general way, covering most of the trend models that are presently used for economic data.

Finally, in general semi-parametric and nonparametric models, the series-based LRV estimation method described earlier also requires a selection procedure to determine the number of the trend basis functions. The test-optimal procedures proposed in Sun (2011, 2012) may be generalized to semi-parametric and nonparametric models. Moreover, current applications of series LRV estimation methods involve semi-parametric or nonparametric models of stationary data. It is of interest to extend this work on series LRV estimation and associated inference procedures to econometric models with nonstationary data.

9 Appendix

Proof of Lemma 2.1. The proof of this lemma is included for completeness. The symmetry of $\gamma(\cdot,\cdot)$ follows by its definition. To show continuity, note that for any $t_o, s_o, t_1, s_1 \in$

$[a,b]$, by the triangle and Hölder inequalities

$$\begin{aligned} |\gamma(t_1,s_1)-\gamma(t_o,s_o)| &= |E\left[X(s_1)X(t_1)\right]-E\left[X(s_o)X(t_o)\right]| \\ &\leq \|X(t_1)\|\,\|X(s_1)-X(s_o)\|+\|X(s_o)\|\,\|X(t_1)-X(t_o)\| \end{aligned}$$

which together with the q.m. continuity of $X(\cdot)$ implies that

$$|\gamma(t_1,s_1)-\gamma(t_o,s_o)|\to 0 \tag{9.1}$$

for any $t_o,s_o,t_1,s_1\in[a,b]$ such that $t_1\to t_o$ and $s_1\to s_o$. The convergence in (9.1) implies that $\gamma(\cdot,\cdot)$ is a continuous function on $[a,b]\times[a,b]$ with $|\gamma(a,a)|<\infty$ and $|\gamma(b,b)|<\infty$. As a result, we get the following condition

$$\max_{t\in[a,b]}|\gamma(t,t)|<\infty. \tag{9.2}$$

Furthermore, we see that for any $g\in L^2[a,b]$

$$\begin{aligned} \int_a^b\int_a^b g(t)\gamma(t,s)g(s)dsdt &= \int_a^b\int_a^b E\left[g(t)X(t)g(s)X(s)\right]dsdt \\ &= E\left[\int_a^b g(t)X(t)\int_a^b g(s)X(s)dsdt\right] \\ &= E\left[\left(\int_a^b g(t)X(t)dt\right)^2\right]\geq 0 \end{aligned} \tag{9.3}$$

where the second equality is by (9.2) and Fubini's Theorem. ■

Proof of Theorem 2.1. The proof of this Theorem is included for completeness. Let $Z_k\equiv\int_a^b X(t)\varphi_k(t)dt$. Then it is clear that

$$E\left[Z_k\right]=E\left[\int_a^b X(t)\varphi_k(t)dt\right]=\int_a^b E\left[X(t)\right]\varphi_k(t)dt=0 \tag{9.4}$$

and

$$\begin{aligned} E\left[Z_k Z_{k'}\right] &= E\left[\int_a^b \int_a^b X(s)X(t)\varphi_k(t)\varphi_{k'}(s)dsdt\right] \\ &= \int_a^b \int_a^b \gamma(s,t)\varphi_k(t)\varphi_{k'}(s)dsdt \\ &= \lambda_k \int_a^b \varphi_k(t)\varphi_{k'}(t)dt = \lambda_k \delta_{kk'}, \end{aligned} \tag{9.5}$$

and moreover

$$E\left[Z_k X(t)\right] = E\left[X(t)\int_a^b X(t)\varphi_k(t)dt\right] = \int_a^b \gamma(t,s)\varphi_k(s)dt = \lambda_k \varphi_k(t), \tag{9.6}$$

for any $k, k' \in \mathbb{Z}_+$. Note that the uniform bound of $\gamma(\cdot,\cdot)$ and Fubini's theorem ensure that we can exchange the integration and expectation in (9.4)-(9.6). Let M be some positive integer, then by definition, (9.6) and uniform convergence in (2.9), we deduce that

$$\begin{aligned} \left\| X(t) - \sum\nolimits_{k=1}^{M} Z_k \varphi_k(t) \right\|^2 &= \gamma(t,t) - 2\sum\nolimits_{k=1}^{M} \varphi_k(t) E\left[Z_k X(t)\right] + \sum\nolimits_{k=1}^{M} \lambda_k \varphi_k^2(t) \\ &= \gamma(t,t) - \sum\nolimits_{k=1}^{M} \lambda_k \varphi_k^2(t) \\ &= \sum\nolimits_{k=M+1}^{\theta} \lambda_k \varphi_k^2(t) \to 0, \text{ as } M \to \infty \end{aligned} \tag{9.7}$$

uniformly over $t \in [a,b]$, which proves sufficiency. Next suppose that $X(t)$ has the following representation

$$X(t) = \sum_{k=1}^{\infty} \alpha_k^{\frac{1}{2}} \xi_k^* g_k(t) \text{ with } E\left[\xi_k^* \xi_{k'}^*\right] = \int_a^b g_k(t) g_{k'}(t) dt = \delta_{kk'}.$$

Then by definition

$$\begin{aligned}\gamma(s,t) &= E\left[\sum_{j=1}^{\infty}\alpha_k^{\frac{1}{2}}\xi_k^* g_k(s)\sum_{j=1}^{\infty}\alpha_k^{\frac{1}{2}}\xi_k^* g_k(t)\right] \\ &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\alpha_j^{\frac{1}{2}}\alpha_k^{\frac{1}{2}} g_j(s)g_k(t)\delta_{jk} \\ &= \sum_{k=1}^{\infty}\alpha_k g_k(s)g_k(t).\end{aligned}$$

Hence for any $k \in \mathbb{Z}_+$

$$\int_a^b \gamma(t,s)g_k(s)dt = \int_a^b \left[\sum\nolimits_{j=1}^{\infty}\alpha_j g_j(t)g_j(s)g_k(s)\right]ds = \sum_{j=1}^{\infty}\alpha_j g_j(t)\delta_{jk} = \alpha_k g_k(t),$$

which implies that $\{(\alpha_k, g_k)\}_{k=1}^{\infty}$ are the eigenvalues and orthonormal eigenfunctions of the kernel function $\gamma(\cdot,\cdot)$. This proves necessity. ■

Proof of Lemma 5.1. First, note that

$$n(\widehat{A}_{K,n} - A_o) = \frac{\frac{1}{n}U'_{y\cdot x}R_{\Delta X,K}X}{\frac{1}{n^2}X'R_{\Delta X,K}X}.$$

We next establish the asymptotic distributions of related quantities in the above expression.

$$\begin{aligned}\frac{X'P_K X}{n^2} &= \frac{X'\Phi_K\left(\Phi'_K\Phi_K\right)^{-1}\Phi'_K X}{n^2} \\ &= \frac{\sum_{t=1}^n X_t\Phi_K(\frac{t}{n})}{n^{\frac{3}{2}}}\left(\frac{\sum_{t=1}^n \Phi'_K(\frac{t}{n})\Phi_K(\frac{t}{n})}{n}\right)^{-1}\frac{\sum_{t=1}^n X_t\Phi'_K(\frac{t}{n})}{n^{\frac{3}{2}}} \\ &\to_d \left[\int_0^1 B_x(r)\Phi_K(r)dr\right]\left[\int_0^1 B_x(r)\Phi'_K(r)dr\right] \\ &\stackrel{d}{=} \sum_{k=1}^K\left[\int_0^1 B_x(r)\varphi_k(r)dr\right]^2 \stackrel{d}{=} \sum_{k=1}^K \xi_{x,k}^2. \qquad (9.8)\end{aligned}$$

$$
\begin{aligned}
\Delta X' P_K \Delta X &= \Delta X' \Phi_K \left(\Phi_K' \Phi_K\right)^{-1} \Phi_K' \Delta X \\
&= \frac{\sum_{t=1}^n \Delta X_t \Phi_K(\frac{t}{n})}{n^{\frac{1}{2}}} \left(\frac{\sum_{t=1}^n \Phi_K'(\frac{t}{n})\Phi_K(\frac{t}{n})}{n}\right)^{-1} \frac{\sum_{t=1}^n \Delta X_t \Phi_K'(\frac{t}{n})}{n^{\frac{1}{2}}} \\
&\to_d \left[\int_0^1 \Phi_K(r) dB_x(r)\right] \left[\int_0^1 \Phi_K'(r) dB_x(r)\right] \\
&\stackrel{d}{=} \sum_{k=1}^K \left[\int_0^1 \varphi_k(r) dB_x(r)\right]^2 \stackrel{d}{=} \sum_{k=1}^K \eta_{x,k}^2. \qquad (9.9)
\end{aligned}
$$

$$
\begin{aligned}
\frac{X' P_K \Delta X}{n} &= \frac{X' \Phi_K \left(\Phi_K' \Phi_K\right)^{-1} \Phi_K' \Delta X}{n} \\
&= \frac{\sum_{t=1}^n X_t \Phi_K(\frac{t}{n})}{n^{\frac{3}{2}}} \left(\frac{\sum_{t=1}^n \Phi_K'(\frac{t}{n})\Phi_K(\frac{t}{n})}{n}\right)^{-1} \frac{\sum_{t=1}^n \Delta X_t \Phi_K'(\frac{t}{n})}{n^{\frac{1}{2}}} \\
&\to_d \left[\int_0^1 B_x(r) \Phi_K(r) dr\right] \left[\int_0^1 \Phi_K'(r) dB_x(r)\right] \\
&\stackrel{d}{=} \sum_{k=1}^K \int_0^1 B_x(r) \varphi_k(r) dr \int_0^1 \varphi_k(r) dB_x(r) \stackrel{d}{=} \sum_{k=1}^K \xi_{x,k} \eta_{x,k}. \qquad (9.10)
\end{aligned}
$$

The results in (9.8), (9.9) and (9.10) imply that

$$
\begin{aligned}
\frac{X' R_{\Delta X,K} X}{n^2} &= \frac{X' P_K X}{n^2} - \frac{X' P_K \Delta X}{n} \left(\Delta X' P_K \Delta X\right)^{-1} \frac{\Delta X' P_K X}{n} \\
&\to_d \sum_{k=1}^K \lambda_k \xi_{x,k}^2 - \frac{\left[\sum_{k=1}^K \xi_{x,k} \eta_{x,k}\right]^2}{\sum_{k=1}^K \eta_{x,k}^2}. \qquad (9.11)
\end{aligned}
$$

Next, note that

$$
\begin{aligned}
\frac{U_{y\cdot x}' P_K X}{n} &= \frac{U_{y\cdot x}' \Phi_K \left(\Phi_K' \Phi_K\right)^{-1} \Phi_K' X}{n} \\
&= \frac{\sum_{t=1}^n u_{y\cdot x,t} \Phi_K(\frac{t}{n})}{n^{\frac{1}{2}}} \left(\frac{\sum_{t=1}^n \Phi_K'(\frac{t}{n})\Phi_K(\frac{t}{n})}{n}\right)^{-1} \frac{\sum_{t=1}^n X_t \Phi_K'(\frac{t}{n})}{n^{\frac{3}{2}}} \\
&\to_d \left[\int_0^1 \Phi_K(r) dB_{y\cdot x}(r)\right]' \left[\int_0^1 \Phi_K'(r) B_x(r) dr\right] \\
&\stackrel{d}{=} \sum_{k=1}^K \xi_{x,k} \eta_{y\cdot x,k}, \qquad (9.12)
\end{aligned}
$$

and

$$
\begin{aligned}
U'_{y\cdot x}P_K\Delta X &= U'_{y\cdot x}\Phi_K\left(\Phi'_K\Phi_K\right)^{-1}\Phi'_K\Delta X \\
&= \frac{\sum_{t=1}^n u_{y\cdot x,t}\Phi_K(\frac{t}{n})}{n^{\frac{1}{2}}}\left(\frac{\sum_{t=1}^n \Phi'_K(\frac{t}{n})\Phi_K(\frac{t}{n})}{n}\right)^{-1}\frac{\sum_{t=1}^n \Delta X_t\Phi'_K(\frac{t}{n})}{n^{\frac{1}{2}}} \\
&\to_d \left[\int_0^1 \Phi_K(r)dB_{y\cdot x}(r)\right]\left[\int_0^1 \Phi'_K(r)dB_x(r)\right] \\
&\stackrel{d}{=} \sum_{k=1}^K \eta_{x,k}\eta_{y\cdot x,k}. \qquad (9.13)
\end{aligned}
$$

The results in (9.9), (9.10), (9.11), (9.12) and (9.13) imply that

$$
\begin{aligned}
\frac{U'_{y\cdot x}R_{\Delta X,K}X}{n} &= \frac{U'_{y\cdot x}P_KX}{n} - U'_{y\cdot x}P_K\Delta X\left(\Delta X'P_K\Delta X\right)^{-1}\frac{\Delta X'P_KX}{n} \\
&\to_d \sum_{k=1}^K \xi_{x,k}\eta_{y\cdot x,k} - \frac{\sum_{k=1}^K \eta_{x,k}\eta_{y\cdot x,k}\sum_{k=1}^K \xi_{x,k}\eta_{x,k}}{\sum_{k=1}^K \eta_{x,k}^2}. \qquad (9.14)
\end{aligned}
$$

The result in (5.6) follows directly by (9.11) and (9.14).

For the second result, note that

$$
\widehat{B}_{K,n} = B_o + \frac{U'_{y\cdot x}R_{X,K}\Delta X}{\Delta X'R_{X,K}\Delta X}.
$$

The asymptotic distributions of the quantities in the above expression are obtained as follows. Under (9.8), (9.9) and (9.10), we have

$$
\begin{aligned}
\Delta X'R_{X,K}\Delta X &= \Delta X'P_K\Delta X - \frac{\Delta X'P_KX}{n}\left(\frac{X'P_KX}{n^2}\right)^{-1}\frac{X'P_K\Delta X}{n} \\
&\to_d \sum_{k=1}^K \eta_{x,k}^2 - \frac{\left[\sum_{k=1}^K \xi_{x,k}\eta_{x,k}\right]^2}{\sum_{k=1}^K \xi_{x,k}^2}. \qquad (9.15)
\end{aligned}
$$

Similarly, under (9.8), (9.12) and (9.13), we have

$$\begin{aligned} U'_{y\cdot x}R_{X,K}\Delta X &= U'_{y\cdot x}P_K\Delta X - \frac{U'_{y\cdot x}P_K X}{n}\left(\frac{X'P_K X}{n^2}\right)^{-1}\frac{X'P_K\Delta X}{n} \\ &\to_d \sum_{k=1}^{K}\eta_{x,k}\eta_{y\cdot x,k} - \frac{\sum_{k=1}^{K}\xi_{x,k}\eta_{y\cdot x,k}\sum_{k=1}^{K}\xi_{x,k}\eta_{x,k}}{\sum_{k=1}^{K}\xi_{x,k}^2}. \end{aligned} \tag{9.16}$$

The result in (5.6) follows directly by (9.15) and (9.16). ■

Proof of Lemma 7.1. By (7.3) and the continuous mapping theorem (CMT), we obtain

$$\omega_{K,n}^2 \to_d \frac{\omega^2\sum_{k=1}^{K}\left[\frac{1}{\omega}\int_0^1\phi_k(r)dB_\omega(r)\right]^2}{K} \stackrel{d}{=} \frac{\omega^2}{K}\chi^2(K), \tag{9.17}$$

where the equivalence in distribution follows from the fact that $\frac{1}{\omega}\int_0^1\phi_k(r)dB_\omega(r)$ is a standard normal random variable for any k and is independent of $\frac{1}{\omega}\int_0^1\phi_{k'}(r)dB_\omega(r)$ for any $k\neq k'$. From (7.2), (9.17) and the CMT, we deduce that

$$t_{K,n} = \frac{B_n(1)}{\sqrt{\omega_{K,n}^2}} \to_d \frac{B_\omega(1)/\omega}{\sqrt{\chi^2(K)/K}} \stackrel{d}{=} t_K, \tag{9.18}$$

where the equivalence in distribution follows by definition of the student-t and the fact that $B_\omega(1)$ is independent of $\int_0^1\phi_{k'}(r)dB_\omega(r)$ for any k. ■

Proof of Lemma 7.2. First note that we can rewrite

$$\begin{aligned} & n^{-\frac{1}{2}}\sum_{t=1}^{n}\phi_k(\frac{t}{n})m(Z_t,\widehat{\theta}_n) \\ = \; & n^{-\frac{1}{2}}\sum_{t=1}^{n}\phi_k(\frac{t}{n})m(Z_t,\theta_0) + n^{-\frac{1}{2}}\sum_{t=1}^{n}\phi_k(\frac{t}{n})E\left[m(Z_t,\widehat{\theta}_n)-m(Z_t,\theta_0)\right] \\ & +n^{-\frac{1}{2}}\sum_{t=1}^{n}\phi_k(\frac{t}{n})\left\{m(Z_t,\widehat{\theta}_n)-m(Z_t,\theta_0)-E\left[m(Z_t,\widehat{\theta}_n)-m(Z_t,\theta_0)\right]\right\}. \end{aligned} \tag{9.19}$$

By Assumption 7.2.(i), (ii) and $\int_0^1 \phi_k(r)dr = 0$, we have

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \phi_k(\frac{t}{n}) E\left[m(Z_t, \widehat{\theta}_n) - m(Z_t, \theta_0)\right] = \frac{1}{n} \sum_{t=1}^{n} \phi_k(\frac{t}{n}) O_p(1) = o_p(1). \tag{9.20}$$

Hence, using the results in (9.19), (9.20) and Assumption 7.2.(iii)-(iv), we deduce that

$$\begin{aligned} n^{-\frac{1}{2}} \sum_{t=1}^{n} \phi_k(\frac{t}{n}) m(Z_t, \widehat{\theta}_n) &= n^{-\frac{1}{2}} \sum_{t=1}^{n} \phi_k(\frac{t}{n}) m(Z_t, \theta_0) + o_p(1) \\ &\to_d \int \phi_k(r) dB_m(r) \equiv \xi_k. \end{aligned} \tag{9.21}$$

Under Assumption 7.2.(i), (ii) and (v), we get

$$\sqrt{n} V^{-\frac{1}{2}}(\theta_o) M_{+,n}(\widehat{\theta}_n)(\widehat{\theta}_n - \theta_o) \to_d N(0, I_{d_\theta}) \stackrel{d}{=} \xi_0. \tag{9.22}$$

Using the results in (9.21), (9.22) and the CMT, we deduce that

$$\begin{aligned} d_\theta \mathcal{F}_n &= \left[V^{-\frac{1}{2}}(\theta_o) M_{+,n}(\widehat{\theta}_n) \sqrt{n}(\widehat{\theta}_n - \theta_o)\right]' \\ &\quad \times \left\{ \frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{n} V^{-\frac{1}{2}}(\theta_o) \Lambda_{k,n} \Lambda_{k,n}' V^{-\frac{1}{2}}(\theta_o)\right] \right\}^{-1} \\ &\quad \times \left[V^{-\frac{1}{2}}(\theta_o) M_{+,n}(\widehat{\theta}_n) \sqrt{n}(\widehat{\theta}_n - \theta_o)\right] \\ &\to_d \xi_0' \left(\frac{1}{K} \sum_{k=1}^{K} \xi_k \xi_k' \right)^{-1} \xi_0, \end{aligned}$$

which has Hotelling's T^2-distribution. Using the relation between the T^2-distribution and $\mathcal{F}$-distribution, we get

$$\frac{K - d_\theta + 1}{K} \mathcal{F}_n \to_d \mathcal{F}_{d_\theta, K - d_\theta + 1},$$

which finishes the argument. ■

References

[1] Anderson, T.W., and D.A. Darling (1952): "Asymptotic Theory of Certain 'Goodness-of-fit' Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, Vol. 23, 193–212.

[2] Berlinet, A., and Thomas-Agnan, C. (2003): *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, Boston

[3] Bosq, D. (2000): *Linear Processes in Function Spaces*, Lecture Notes in Statistics. Springer, Berlin.

[4] Chen X., Liao Z. and Sun Y. (2012): "Sieve Inference on Semi-nonparametric Time Series Models," *Cowles Foundation Discussion Paper*, No 1849.

[5] Chen X., Hahn J. and Liao Z. (2012): "Simple Estimation of Asymptotic Variance for Semiparametric Two-step Estimators with Weakly Dependent Data," *Working Paper*, Yale University and UCLA.

[6] Cheng X. and P.C.B. Phillips (2009): "Semiparametric Cointegrating Rank Selection," *Econometrics Journal*, Vol. 12, 83-104,

[7] Cheng X. and P.C.B. Phillips (2012): "Cointegrating Rank Selection in Models with Time-Varying Variance," *Journal of Econometrics*, Forthcoming.

[8] Durlauf, S. N. and Phillips P.C.B. (1988): "Trends versus Random Walks in Time Series Analysis," *Econometrica*, Vol. 56; 1333-1354.

[9] Granger, C.W.J. and Newbold P. (1974): "Spurious Regression in Econometrics", *Journal of Econometrics*, Vol. 2, 111-120.

[10] Hannan, E. J. (1970): *Multiple Time Series*, New York: John Wiley and Sons.

[11] Johansen S. (1998): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, Vol. 12, No. 2-3, 231-254.

[12] Johansen S. (1995): *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, USA.

[13] Kac, M. and Siegert, A. J. F. (1947): "An Explicit Representation of a Stationary Gaussian Process," *Annals of Mathematical Statistics,* Vol. 18, 438–442.

[14] Karhunen, K. (1946): "Zur Spektraltheorie Stochastischer Prozesse," *Annales Academiae Scientiarum Fennicae*, Vol. 37.

[15] Liao Z. and P.C.B. Phillips (2012): "Automated Estimation of Vector Error Correction Models," *Working Paper*, Yale University and UCLA.

[16] Liao Z. and P.C.B. Phillips (2010): "Reduced Rank Regression of Partially Nonstationary Vector Autoregressive Processes under Misspecification," *Working Paper,* Yale University and UCLA.

[17] Liao Z. and P.C.B. Phillips (2011): "A Lasso-type of IV Selection Approach in the Efficient Estimation of Cointegrated Systems," *Working Paper,* Yale University and UCLA.

[18] Loéve, M.M. (1955): *Probability Theory*, Princeton, N.J. Van Nostrand.

[19] Ouliaris, S., J. Y. Park and P. C. B. Phillips (1989), "Testing for a unit root in the presence of a maintained trend." In B. Raj (ed.), *Advances in Econometrics and Modelling*, Norwell, MA: Kluwer Academic Publishers, Chapter 1.

[20] Ploberger, W. and Phillips, P.C.B. (2003): "Empirical Limits for Time Series Econometric Models," *Econometrica*, Vol. 71, No. 2, 627-673.

[21] Parzen, E. (1959): "Statistical Inference on Time Series by Hilbert Space Methods," *Technical Report*, 23, Statistics Department, Stanford University.

[22] Parzen, E. (1961a): "An Approach to Time Series Analysis," *Annals of Mathematical Statistics*, Vol. 32, No. 4, 951–989.

[23] Parzen, E. (1961b): "Regression Analysis of Continuous Parameter Time Series," *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley.

[24] Parzen, E. (1963). "Probability Density Functionals and Reproducing Kernel Hilbert Spaces," In *Proc. Sympos. Time Series Analysis*, Rosenblatt ed., Wiley, New York.

[25] Phillips, P. C. B. (1987). “Time Series Regression with a Unit Root,” *Econometrica,* 55, 277–302.

[26] Phillips, P.C.B. (1989): “Partially Identified Econometric Models,” *Econometric Theory*, Vol. 5, 181-240.

[27] Phillips, P.C.B. (1991a): "Optimal Inference in Cointegrated Systems," *Econometrica*, Vol. 59, No. 2, 283-306.

[28] Phillips, P.C.B. (1991b): "Spectral Regression for Cointegrated Time Series," In W. Barnett, J. Powell and G. Tauchen (ed.), *Nonparametric And Semiparametric Methods in Econometrics And Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, Cambridge University Press.

[29] Phillips, P.C.B. (1995): "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, Vol. 63, No. 5, 1023-1078.

[30] Phillips, P.C.B. (1998): "New Tools for Understanding Spurious Regressions," *Econometrica*, Vol. 66, No. 6, 1299-1326.

[31] Phillips, P.C.B. (2001): "New Unit Root Asymptotics in the Presence of Deterministic Trends," *Journal of Econometrics*, Vol. 11, 323-353.

[32] Phillips, P.C.B. (2005a): "Challenges of Trending Time Series Econometrics," *Mathematics and Computers in Simulation*, Vol. 68, No. 5-6, 401-416.

[33] Phillips, P.C.B. (2005b): "HAC Estimation By Automated Regression," *Econometric Theory*, Vol. 21, No. 1, 116-142.

[34] Phillips, P.C.B. (2007): “Unit Root Log Periodogram Regression,” *Journal of Econometrics*, Vol. 138, 104-124.

[35] Phillips, P.C.B. (2012): "Optimal Estimation of Cointegrated Systems with Irrelevant Instruments," *Journal of Econometrics*, Forthcoming.

[36] Phillips, P.C.B. and B.E. Hansen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economics Studies*, Vol. 57, 99-125.

[37] Phillips, P. C. B. and P. Perron (1988). “Testing for a unit root in time series regression,” *Biometrika* 75, 335–346.

[38] Phillips, P.C.B. and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, Vol. 20, No. 2, 971-1001.

[39] Pugachev V.S. (1959): "A Method for Solving the Basic Integral Equation of Statistical Theory of Optimum Systems in Finite Form," *Prikl. Math. Mekh.*, Vol. 23, 3-14 (English translation pp. 1-16).

[40] Staiger, D. and J.H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, Vol. 65, 557-586.

[41] Shorack, G.R., and Wellner, J.A. (1986): *Empirical Processes With Applications to Statistics*, Wiley, NY.

[42] Stephens, M.A. (1976): "Asymptotic Results for Goodness-of-fit Statistics with Unknown Parameters," *Annals of Statistics*, Vol. 4, 357–369.

[43] Sun, Y. and Kim, M.S. (2012a), "Simple and Powerful GMM Over-identification Tests," *Working Paper*, Department of Economics, UC San Diego.

[44] Sun, Y. and Kim, M.S. (2012b), "Asymptotic F Test in a GMM Framework with Cross Sectional Dependence," *Working Paper*, Department of Economics, UC San Diego.

[45] Sun, Y. (2011), "Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter," *Journal of Econometrics,* Vol. 164. No. 2, 345-366.

[46] Sun, Y. (2012), "Heteroscedasticity and Auto-correlation Robust F Test Using Orthonormal Series Variance Estimator." *Working Paper*, Department of Economics, UC San Diego.

[47] Watson, G.S. (1962): "Goodness-of-fit Tests on a circle II," *Biometrika*, Vol 49, 57-63

[48] Yule, U. (1926): "Why Do We Sometimes Get Nonsense-correlations Between Time Series? A Study in Sampling and the Nature of Time Series." *Journal of the Royal Statistical Society* Vol. 89, 1-69.

[49] Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, Vol. 101, No. 476, 1418-1429.