NEW TOOLS FOR UNDERSTANDING SPURIOUS REGRESSIONS

BY

PETER C.B. PHILLIPS

COWLES FOUNDATION PAPER NO. 966

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
1998

Econometrica, Vol. 66, No. 6 (November, 1998), 1299–1325

NEW TOOLS FOR UNDERSTANDING SPURIOUS REGRESSIONS[1]

BY PETER C. B. PHILLIPS

Some new tools for analyzing spurious regressions are presented. The theory utilizes the general representation of a stochastic process in terms of an orthonormal system and provides an extension of the Weierstrass theorem to include the approximation of continuous functions and stochastic processes by Wiener processes. The theory is applied to two classic examples of spurious regressions: regression of stochastic trends on time polynomials, and regressions among independent random walks. It is shown that such regressions reproduce in part and in whole the underlying orthonormal representations.

KEYWORDS: Loève Karhunen representation, nonsense correlation, orthonormal systems, spurious regression, Weierstrass theorem.

1. INTRODUCTION

SPURIOUS REGRESSIONS, OR NONSENSE CORRELATIONS as they were originally called, have a long history in statistics, dating back at least to Yule (1926). Textbooks and the literature of statistics and econometrics abound with interesting examples, many of them quite humorous. One is the high correlation between the number of ordained ministers and the rate of alcoholism in Britain in the nineteenth century. Another is that of Yule (1926), reporting a correlation of 0.95 between the proportion of Church of England marriages to all marriages and the mortality rate over the period 1866–1911. Yet another is the econometric example of alchemy reported by Hendry (1980) between the price level and cumulative rainfall in the U.K. The latter "relation" proved resilient to many econometric diagnostic tests and was humorously advanced by its author as a new "theory" of inflation. With so many well known examples like these, the pitfalls of regression and correlational studies are now common knowledge, even to nonspecialists. The situation is especially difficult in cases where the data are trending—as indeed they are in the examples above—because "third" factors that drive the trends come into play in the behavior of the regression, although these factors may not be at all evident in the data. Moreover, as we have come to understand in recent years (although the essence of the problem was evidently understood by Yule in his original article), it is the commonality of trending mechanisms in data that often leads to spurious regression relations.

[1]The original version of this paper, entitled, "Spurious Regression Unmasked," was delivered as an Invited Lecture at the XIV Latin American Meetings of the Econometric Society, Rio de Janeiro, August 5–9, 1996. That version of the paper is available as Cowles Foundation Discussion Paper No. 1135 and can be obtained on request. Some of the ideas that appear in Section 3 of the paper were first suggested by the author while presenting an overview at a conference on Unit Roots and Cointegration at INSEE/ENSAE in June, 1991. Computations in the paper were performed by the author in GAUSS and the paper was typed by the author in Scientific Word 2.5. The author's thanks go to the co-editor and three referees for comments on the original version of the paper, and to the NSF for research support under Grant No. SBR 94-22922.

What makes the phenomenon dramatic is that it occurs even when the data are otherwise independent.

In a prototypical spurious regression, the fitted coefficients are statistically significant when there is no "true relationship" between the dependent variable and the regressors. The statistical significance is deemed spurious and misleading because there is no meaningful relationship between the variables. Using Monte Carlo simulations, Granger and Newbold (1974) showed that this phenomenon occurs when independent random walks are regressed on one another. Phillips (1986) gave an analytic theory of regressions of this type for quite general stochastic trends, showing, inter alia, that the t- and F-ratio significance tests have divergent asymptotic behavior in such regressions. Therefore, such outcomes are inevitable in large samples. Similar phenomena occur in regressions of stochastic trends on deterministic polynomial regressors, as shown in Durlauf and Phillips (1988). The simple heuristic explanation for phenomena of this type is that conventional statistical tests do nothing more than reveal the presence of a trend in the dependent variable by making the fitted coefficients significant for all regressors that themselves have trends. Thus, the commonality of trending mechanisms in data is the source of these spurious regressions.

The purpose of this paper is to develop some new tools for analyzing and understanding such regressions. These tools help us explain why significant regression coefficients occur in what seem to be manifestly incorrect regression specifications relating variables that are statistically independent. The common theme, of course, is that all the variables share the common feature of a trending mechanism, even though they may otherwise be unrelated and even though the trending mechanisms themselves may be very different. We develop an asymptotic theory to explain this phenomena. A fascinating feature of the theory is that, just as we may model a continuous function by Fourier series in terms of different orthonormal system coordinates, so too we may validly represent a trending stochastic process in various ways, including the use of trending regressors that are independent of the time series being modelled. The fact that the fitted regression coefficients are significant in such cases is shown to be a statistical manifestation of the existence of this underlying representation.

It is important to recognize that such representations as we will discuss in this paper do not take the place of temporal predictive models. Nor do they serve as mechanisms for understanding temporal causal relationships between time series. In an important respect, the limit theory we present is a limit theory of a "sample period fit," in which the sample period can be viewed as a snapshot of an infinite time series. Such asymptotic analysis is already used, albeit implicitly, in econometrics. One example, for instance, is the derivation of trend break limit theory, wherein the breaks are considered to occur at some fraction of the sample that turns out, in the limit, to be the same fraction of the infinite trajectory. In this respect, therefore, the "snapshot of infinity" asymptotic theory of this paper is not a radical departure from some established lines of asymptotic analysis in econometrics. It will turn out to be an important mode of analysis in the general development of misspecification-robust asymptotics for trending time series.

The starting point in the approach that we adopt is a general orthonormal representation theory of a continuous stochastic process, and the theory that we use here is outlined in Section 2 of the paper. Our theoretical development is primarily focussed on stochastic trends and their associated Brownian motion limits, but many of our results hold for other limiting stochastic processes (such as diffusions) that are amenable to an orthonormal representation, and to deterministic functions of time other than polynomials and trigonometric functions. Section 3 shows how the orthonormal representation of a stochastic process is accurately reproduced by a fitted regression, and is completely captured when the number of regressors grows with the sample size. Section 4 shows that the Weierstrass approximation theorem can be extended to give a theory of approximation of continuous functions by independent Wiener processes, gives some illustrations, and applies the theory to the case of the classic spurious regression of independent random walks. Section 5 concludes the paper. Proofs are collected together and notation is listed in an Appendix.

2. SOME PRELIMINARY REPRESENTATION THEORY

We start by making use of the general representation theory of a stochastic process in terms of an orthonormal system. Several forms are available, the most common of which is the Loève-Karhunen representation, which is given in Lemma 2.1 below. This result ensures that any random function that is continuous in quadratic mean has a decomposition into a countable linear combination of orthogonal functions. The representation is analogous to the Fourier series expansion of a continuous function. Thus, suppose $X(t)$ is a zero mean stochastic process that is continuous in quadratic mean on the interval $[0,1]$ and has covariance function $\gamma(r,s)$. Let $\{\varphi_k\}_{k=1}^{\infty}$ be a complete orthonormal system in $L_2[0,1]$ with the property that these functions serve as the eigenvectors of the covariance operator, i.e., $\lambda_k \varphi_k(r) = \int_0^1 \gamma(r,s)\varphi_k(s)\,ds$, where λ_k is the eigenvalue of $\gamma(r,s)$ corresponding to the eigenfunction φ_k. Mercer's theorem (e.g., Shorack and Wellner (1986, p. 208)) ensures that the covariance function can be decomposed as

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

where the series converges absolutely and uniformly on $[0,1]$.[2] The corresponding decomposition for the stochastic process $X(t)$ is most often called the Loève-Karhunen expansion, although the stationary Gaussian case is sometimes attributed to Kac and Siegert (1947). The following statement of the expansion is given in Loève (1963, p. 478):

2.1 LEMMA: *A random function* $X(t)$ *that is continuous in quadratic mean on the interval* $[0,1]$ *has on this interval the orthogonal expansion*

$$(2) \qquad X(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, \varphi_k(t)\xi_k,$$

with

$$E(\xi_k \xi_j) = \delta_{kj}, \qquad \int_0^1 \varphi_k(s)\varphi_j(s)\,ds = \delta_{kj},$$

iff the λ_k *are the eigenvalues and the* φ_k *are the orthonormalized eigenfunctions of the autocovariance function* $\gamma(r,s)$. *The series* (2) *converges in quadratic mean uniformly on* $[0,1]$. *The orthogonal random quantities* ξ_k *that appear in* (2) *can be represented in the form* $\xi_k = \lambda_k^{-1/2}\int_0^1 x(s)\varphi_k(s)\,ds$. *The* δ_{kj} *above is Kronecker's delta.*

Just as Fourier series of continuous functions do not always converge pointwise (but do converge in mean), the representation (2) of the stochastic process $X(t)$ converges in quadratic mean but not necessarily pointwise. For this reason, the equivalence in (2) is sometimes represented by the symbol " $\sim$ ", signifying that the series is convergent in the L_2 sense and that distributional equivalence applies.

There are many different representations of standard Brownian motion $W(r)$ that originate in the general form (1). The simplest is the Loève-Karhunen expansion itself, which is obtained by using the eigenvalues and eigenfunctions of the covariance kernel $\gamma(r,x) = r \wedge s$, viz.

$$(3) \qquad \lambda_k = \frac{4}{(2k-1)^2\pi^2}, \qquad \varphi_k(r) = \sqrt{2}\sin[(k-1/2)\pi r]$$

directly in (2), giving the following L_2-representation

$$(4) \qquad W(r) = \sqrt{2}\sum_{k=1}^{\infty} \frac{\sin[(k-1/2)\pi r]}{(k-1/2)\pi}\xi_k,$$

where the components ξ_k are independently and identically distributed (iid) as $N(0,1)$. It is easily seen by applying the Martingale convergence theorem (for square integrable martingales) that the series representation (4) of $W(r)$ is convergent almost surely and uniformly in $r \in [0,1]$, so this series does converge pointwise.

Another commonly used representation is developed as follows. Let $V(r) = W(r) - rW(1)$ be the Brownian bridge process corresponding to the Brownian motion $W(r)$. The covariance function of $V(r)$ is $\gamma(r,s) = r \wedge s - rs$, which can be decomposed as in (1) above with eigenfunctions given by the orthonormal system $\{\sqrt{2}\sin(k\pi r)\}_{k=1}^{\infty}$ and corresponding eigenvalues $\lambda_k = (k\pi)^{-2}$—e.g., Shorack and Wellner (1986, pp. 213–214). This leads to the following L_2-representation of $V(r)$:

$$(5) \qquad V(r) = \sqrt{2}\sum_{k=1}^{\infty}\frac{\sin(k\pi r)}{k\pi}\xi_k, \qquad \text{with} \quad \xi_k = \sqrt{2}\int_0^1 \frac{\sin(k\pi s)}{k\pi}V(s)\,ds.$$

The components ξ_k in this decomposition are also iid $N(0,1)$, as can be verified by direct calculation. The representation (5) gives rise to a corresponding expansion for the Brownian motion $W(r)$, viz.

$$(6) \qquad W(r) = r\xi_0 + \sqrt{2}\sum_{k=1}^{\infty}\frac{\sin(k\pi r)}{k\pi}\xi_k,$$

with

$$\xi_0 = W(1), \quad \xi_k = \sqrt{2}\int_0^1 \frac{\sin(k\pi s)}{k\pi}(W(s) - sW(1))\,ds.$$

Again, the ξ_k are iid $N(0,1)$. The series (6) is known to converge almost surely and uniformly for $r \in [0,1]$—e.g., Hida (1980, p. 73, Remark 2), and Brieman (1992, p. 261), where the series are defined over the intervals $[0, 2\pi]$, and $[0, \pi]$.

The representation (6) has a linear trend component with the random coefficient ξ_0, and shows that $W(r)$ can be written in terms of both polynomial and sinusoidal functions. In fact, (6) is one of many alternative functional representations of Brownian motion. For instance, we may write each of the trigonometric functions in the orthonormal system $\{\sqrt{2}\sin(k\pi r)\}_{k=1}^{\infty}$ in terms of another orthonormal basis, such as orthonormal polynomials in r, and then substitute these orthonormal series for the sinusoidal functions in (6), giving a new representation of $W(r)$ in terms of the new basis. The coefficients in this new representation are still random and normally distributed, but no longer necessarily independent. Another popular representation of $W(r)$ is in terms of Schauder functions (orthogonal tent functions) and here again the convergence is uniform in $r \in [0,1]$ almost surely—see Karatzas and Shreve (1991, Lemma 3.1, p. 57). In all of these different representations to the continuous stochastic process $W(r)$ is written as an infinite linear combination of deterministic functions with random coefficients. What distinguishes the Loève-Karhunen expansion, is that the random coefficients as well as the deterministic functions form an orthonormal sequence.

These expansions may be used directly to create representations for linear diffusion processes like $J_c(r)$ which satisfy the stochastic differential equation $dJ_c(r) = cJ_c(r)\,dr + dW(r)$ for some constant c. With initial condition $J_c(r) = 0$, the solution of this equation has the form:

$$(7) \qquad J_c(r) = \int_0^r e^{(r-s)c}\,dW(s) = W(r) + c\int_0^r e^{(r-s)c}W(s)\,ds.$$

Substituting (4) into (7) we find

$$J_c(r) = \sqrt{2} \sum_{k=1}^{\infty} \frac{1}{(k-1/2)\pi} \times \left[\sin[(k-1/2)\pi r] + c \int_0^r e^{(r-s)c} \sin[(k-1/2)\pi s] ds \right] \xi_k$$
$$= \sqrt{2} \sum_{k=1}^{\infty} \frac{1}{(k-1/2)^2\pi^2 + c^2} [ce^{cr} - c\cos[(k-1/2)\pi r] + (k-1/2)\pi \sin[(k-1/2)\pi r]].$$

The substitution is valid because the series (4) is uniformly convergent almost surely and can be integrated term by term. Another representation can be obtained by using (6) instead of (4) in (7), and yet another is the Loève-Karhunen expansion (2) itself, based on the eigenvalues and eigenfunctions of the covariance kernel $\gamma(r,s) = e^{(r+s)c}/2c[1 - e^{-2(r \wedge s)c}]$ of $J_c(r)$.

3. REPRODUCTION OF THE ORTHOGONAL REPRESENTATION BY SPURIOUS REGRESSION

The existence of expansions like (4)–(6) indicates that continuous processes such as Brownian motion can be represented and, indeed, generated by deterministic functions of time with random coefficients. To the extent that standardized discrete time series with a unit root converge weakly to Brownian motion processes, we infer that deterministic functions of the same type may be used to model such time series. This brings us to the study of prototypical spurious regressions in which unit root nonstationary time series are regressed on deterministic functions, a topic first studied analytically in Durlauf and Phillips (1988) for the case of a linear trend.

We are concerned to ask the following question. Consider the time series $y_t = \Sigma_1^t u_s$, where u_t is a stationary time series with zero mean and finite absolute moments to order $p > 2$. What are the properties of a regression of the form

$$y_t = \sum_{k=1}^{K} \hat{b}_k \varphi_k\left(\frac{t}{n}\right) + \hat{u}_t \tag{8}$$

or, equivalently (with $\hat{a}_k = n^{-1/2}\hat{b}_k$),

$$\frac{y_t}{\sqrt{n}} = \sum_{k=1}^{K} \hat{a}_k \varphi_k\left(\frac{t}{n}\right) + \frac{\hat{u}_t}{\sqrt{n}}, \tag{9}$$

when the limiting behavior of the dependent variable is a Brownian motion, i.e.,

$$\frac{y_{[n\cdot]}}{\sqrt{n}} \Rightarrow B(\cdot) \equiv BM(\sigma^2), \tag{10}$$

and the regressors φ_k form a complete orthonormal system in $L_2[0,1]$. In what follows, we assume that the functional central limit theorem (10) holds (see Phillips and Solo (1992), for primitive conditions); and, to be specific and to relate outcomes directly to those of the previous section, we take φ_k and λ_k to be the eigenfunctions and eigenvalues of the covariance kernel $\sigma^2 r \wedge s$ of the limiting Brownian motion B, which are obtained from (3) above by scaling λ_k by σ^2.

In view of (2) and (10), we may very well expect that the regressors in (9) take on the role of the deterministic functions in the associated orthonormal representation of the limiting Brownian motion $B(\cdot)$. Perhaps, we can go even further than this. If $K \to \infty$ as $n \to \infty$, could (9) succeed in reproducing the entire L_2 orthonormal representation of $B(\cdot)$? We now proceed to examine whether these heuristic notions can be made more precise.

Let $\hat{a}_K = (\hat{a}_k)$ be the coefficients and $\varphi_K = (\varphi_k)$ be the K-vector of regressors in (9). Let $c_K \in \mathbb{R}^K$ be any vector with $c_K' c_K = 1$, $t_{c_K'\hat{a}_K}$ be the usual least squares regression t-ratio for the linear combination of coefficients $c_K' a_K$, and let R^2 and DW be the regression coefficient of determination and Durbin Watson statistics, respectively. The following two theorems give the asymptotic properties of these statistics when K is fixed and when $K \to \infty$.

3.1 THEOREM: *For fixed K, as $n \to \infty$ we have*:

(a) $$c_K'\hat{a}_K \Rightarrow c_K'\left[\int_0^1 \varphi_K B\right] \overset{d}{=} N\left(0, \sigma^2 c_K' \int_0^1\int_0^1 \varphi_K(r)(r\wedge s)\varphi_K(s)'\,ds\,dr\,c_K\right) = N(0, c_K' \Lambda_K c_K),$$

(b) $$n^{-2}\sum_{t=1}^{n} \hat{u}_t^2 \Rightarrow \int_0^1 B_{\varphi_K}^2,$$

(c) $$n^{-1/2} t_{c_K'\hat{a}_K} \Rightarrow c_K'\left[\int_0^1 \varphi_K B\right] \Big/ \left(\int_0^1 B_{\varphi_K}^2\right)^{1/2},$$

(d) $$R^2 \Rightarrow 1 - \int_0^1 B_{\varphi_K}^2 \Big/ \int_0^1 B^2, \qquad DW \overset{p}{\to} 0,$$

where $B_{\varphi_K}(\cdot) = B(\cdot) - (\int_0^1 B\varphi_K')(\int_0^1 \varphi_K\varphi_K')^{-1}\varphi_K(\cdot)$ is the L_2-projection residual of B on φ_K, $\Lambda_K = diag(\lambda_1, \ldots, \lambda_K)$, and λ_k is the eigenvalue of the covariance function $\sigma^2 r \wedge s$ corresponding to φ_k.

3.2 REMARKS: (a) Theorem 3.1(a) shows that the fitted coefficients in the regression (9) tend to random variables in the limit as $n \to \infty$. Moreover, the random limits are equivalent in distribution to the corresponding random elements in the Loève-Karhunen representation of the limit process $B(\cdot)$. Thus, (9) reproduces accurately in the limit the appropriate elements in the orthogonal representation of the limiting form of the dependent variable process. In this sense, we can interpret (9) as a partial but nonetheless correctly specified empirical version of an orthogonal representation of Brownian motion. We use the word "partial" here because (9) has only K regressors, i.e., $\varphi_K = (\varphi_j)_{j=1}^K$. The model is correctly specified because the regressors that are omitted from (9), viz. $\varphi_\perp = (\varphi_{K+j})_{j=1}^\infty$, are all orthogonal to the included variables. Hence, (9) is indeed well suited to least squares regression. All of the above holds in spite of the fact that the Durbin Watson statistic $DW \xrightarrow{p} 0$, indicating that the residuals in the fitted model are serially dependent. Thus, conventional wisdom that the regression model (8) is spurious and that the low DW statistic signals that inference is hazardous is mistaken here. On the other hand, conventional wisdom that the low DW statistic may signal poor predictive performance of the model may well be appropriate.

(b) Part (c) of Theorem 3.1 shows that the usual regression t ratios of the fitted coefficients diverge at the rate $O_p(n^{1/2})$, and, therefore, ultimately exceed any finite critical values as n increases. Hence, such tests indicate statistically significant regression coefficients with probability that goes to one as $n \to \infty$. The fitted coefficients in (9) are not spuriously significant because the significant t ratios correctly indicate the presence of the orthonormal representation

$$(11)\qquad B(r) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, \varphi_k(r)\xi_k, \quad \text{where} \quad \xi_k \equiv \text{iid } N(0,1),$$
$$= \sum_{k=1}^{\infty} \varphi_k(r)\eta_k, \quad \text{where} \quad \eta_k \equiv \text{iid } N(0,\lambda_k).$$

In effect, the fitted regression (9) is an empirical model for (11). Setting $\eta_K = (\eta_K)_1^K$, we have

$$(12)\qquad c_K'\hat{a}_K \Rightarrow N(0, c_K'\Lambda_K c_K) \stackrel{d}{=} c_K'\eta_K.$$

The significant t ratios signal that the regressors play an important role in representing the dependent variable—or its limiting version, the stochastic process $B(r)$.

(c) The t ratios, $t_{c_K'\hat{a}_K}$, that are analyzed in Theorem 3.1 are computed using the conventional least squares regression formulae. In their place, robust t ratios which accommodate serial dependence in the residuals could be computed and it is of interest to examine whether the remarks made above in (b) continue to apply. Serial correlation robust t ratios for the coefficient $c_K'\hat{a}_K$ in (9) are based on the formula $\tilde{t}_{c_K'\hat{a}_K} = c_K'\hat{a}_K/\tilde{s}_{c_K'\hat{a}_K}$, where

$$\tilde{s}^2_{c_K'\hat{a}_K} = c_K'(\Phi_K'\Phi_K)^{-1}\left(n\,\widehat{\text{lrvar}}\left(n^{-1/2}\hat{u}_t\varphi_K\left(\frac{t}{n}\right)\right)\right)(\Phi_K'\Phi_K)^{-1}c_K,$$

and

$$\widehat{\text{lrvar}}(X_t) = \sum_{j=-M}^{M} k\left(\frac{j}{M}\right)c(j,X), \quad c(j,X) = n^{-1}\sum_{1\le t,t+j\le n} X_tX_{t+j}.$$

Here, $\widehat{\text{lrvar}}(\cdot)$ signifies a kernel estimate of the long run variance of its argument, $k(\cdot)$ is a lag kernel, and M is a bandwidth parameter for which $M \to \infty$, and $M/n \to 0$ as $n \to \infty$. Using the same approach as that developed in Phillips (1991) for analyzing the asymptotic properties of kernel estimates based on $I(1)$ data, it can be shown that

$$\frac{1}{M}\widehat{\text{lrvar}}\left(n^{-1/2}\hat{u}_t \varphi_K\left(\frac{t}{n}\right)\right) = \frac{1}{M}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right) c\left(j, n^{-1/2}\hat{u}_t \varphi_K\left(\frac{t}{n}\right)\right)$$
$$\Rightarrow \left(\int_{-1}^{1} k(s)\, ds\right)\left(\int_0^1 B_{\varphi_K}^2 \varphi_K \varphi_K'\right).$$

It follows that

$$\frac{n}{M}\tilde{s}_{c_K'\hat{a}_K}^2 \Rightarrow \left(\int_{-1}^{1} k(s)\, ds\right)\left(\int_0^1 \varphi_K \varphi_K'\right)^{-1}\left(\int_0^1 B_{\varphi_K}^2 \varphi_K \varphi_K'\right)\left(\int_0^1 \varphi_K \varphi_K'\right)^{-1}$$
$$= \left(\int_{-1}^{1} k(s)\, ds\right)\left(\int_0^1 B_{\varphi_K}^2 \varphi_K \varphi_K'\right),$$

and thus the serial correlation robust t ratio $\tilde{t}_{c_K'\hat{a}_K}$ has the asymptotic behavior

$$\tilde{t}_{c_K'\hat{a}_K} = \frac{c_K'\hat{a}_K}{\tilde{s}_{c_K'\hat{a}_K}} = \frac{O_p(1)}{O_p\left(\dfrac{M^{1/2}}{n^{1/2}}\right)} = O_p\left(\frac{n^{1/2}}{M^{1/2}}\right).$$

We deduce that the robust t ratios of the coefficients also diverge as $n \to \infty$, but at the rate $(n/M)^{1/2}$, which is slower than the conventional t ratio $t_{c_K'\hat{a}_K}$ by a factor, $M^{1/2}$, which depends on the bandwidth M. Hence, the conclusion of Theorem 3.1 regarding the inevitable statistical significance of the coefficients applies even when serial correlation corrections are made to the standard errors of the estimated coefficients.

(d) An important feature of the true model (11) is that the coefficients η_k are random variables, whereas the variables $\varphi_k(r)$ are deterministic. The empirical regression (9) correctly reproduces this feature of the true model as $n \to \infty$, as is clear from (12).

(e) With some changes in notation, Theorem 3.1 holds if the limiting behavior of the dependent variable is a general continuous stochastic process $X(r)$ rather than Brownian motion. Suppose that for some $\alpha > 0, n^{-\alpha} y_{[n\cdot]} \Rightarrow X(\cdot)$, a continuous stochastic process on $[0,1]$ with continuous covariance function $\gamma(r,s)$ whose eigenfunctions and eigenvalues are given by φ_k and λ_k. Instead of (9), we run the empirical regression

$$\frac{y_t}{n^\alpha} = \sum_{k=1}^{K} \hat{a}_k \varphi_k\left(\frac{t}{n}\right) + \frac{\hat{u}_t}{n^\alpha}.$$

Then, in place of (a), (b), and (c) of Theorem 3.1, we have the following limiting behavior:

$$\text{(a}'\text{)} \qquad c_K'\hat{a}_K \Rightarrow c_K'\left[\int_0^1 \varphi_K X\right] \stackrel{d}{=} N\left(0, c_K'\left[\int_0^1\int_0^1 \varphi_K(r)\gamma(r,s)\varphi_K(s)'\right]c_K\, ds\, dr\right)$$
$$= N(0, c_K'\Lambda_K c_K),$$

$$\text{(b}'\text{)} \qquad n^{-(1+2\alpha)}\sum_{t=1}^{n}\hat{u}_t^2 \Rightarrow \int_0^1 X_{\varphi_K}^2,$$

where $X_{\varphi_K}(\cdot) = X(\cdot) - (\int_0^1 X\varphi_K')(\int_0^1 \varphi_K\varphi_K')^{-1}\varphi_K(\cdot)$, and

$$\text{(c}'\text{)} \qquad n^{-1/2}t_{c_K'\hat{a}_K} \Rightarrow c_K'\left[\int_0^1 \varphi_K X\right] \Big/ \left(\int_0^1 X_{\varphi_K}^2\right)^{1/2}.$$

Thus, the empirical regression asymptotics correctly reproduce the form of the random coefficients in the general Loève-Karhunen representation of $X(\cdot)$ given by (2) and correctly signal their significance. These results apply, for example, to the linear diffusion process $J_c(r) = \int_0^r e^{(r-s)c}\, dW(s)$ in (7) for some constant c, and thereby (i)–(iii) above cover the important case of near integrated time series y_t (i.e., time series with a root, $1 + c/n$, that is near to unity) for which we have $n^{-1/2}y_{[n.]} \Rightarrow J_c(\cdot)$.

(f) As mentioned in the Introduction, the type of limit theory we are using in Theorem 3.1 can be characterized as a limit theory of the sample period fit or "snapshot asymptotics." The terminology can be explained as follows. The dependent variable y_t in the empirical regression (8) is transformed into a standardized random element $n^{-1/2}y_{[n\cdot]}$ in the function space $C[0,1]$. According to (10) the sample behavior of $n^{-1/2}y_{[n\cdot]}$ is approximated by the limit Brownian motion process $B(\cdot)$. Indeed, as discussed in the Section 6 (see (22) below), the probability space can be expanded so that the sample path can be approximated by the Brownian motion up to an error of $o_{a.s.}(1)$. In effect, the sample trajectory of $n^{-1/2}y_{[n\cdot]}$ on $C[0,1]$ is a snapshot of the full limiting trajectory of $B(\cdot)$ on the same space. From the Loève-Karhunen representation of $B(\cdot)$, we know that there is a representation of $B(\cdot)$ in terms of the deterministic functions φ_k with random coefficients. Likewise, the regression (9) gives us an empirical "snapshot" of this limiting representation.

We now proceed to consider what happens when the number of regressors in (9) tends to infinity with n.

3.3 Theorem: *As $K \to \infty$, $c_K'\Lambda_K c_K$ tends to a positive constant $\sigma_c^2 = c'\Lambda c$, where $c = (c_k)$, $\Lambda = diag(\lambda_1, \lambda_2, \ldots)$ and $c'c = 1$. Moreover, if $K \to \infty$ and $K/n \to 0$ as $n \to \infty$, we have*:

$$\text{(a)} \qquad c_K'\hat{a}_K \Rightarrow N(0, \sigma_c^2),$$

$$\text{(b)} \qquad n^{-2}\sum_{t=1}^{n}\hat{u}_t^2 \xrightarrow{p} 0,$$

(c) $n^{-1/2} t_{c'_K \hat{a}}$ *diverges*,

(d) $R^2 \xrightarrow{p} 1$.

3.4 Remarks: (a) Part (a) of Theorem 3.3 gives the limiting distribution of the coefficients of both K and $n \to \infty$. In this case, c_K becomes infinite dimensional and $c'_K \hat{a}_K$ becomes an l_2 inner product. As in the finite dimensional case, $c'_K \hat{a}_K$ converges weakly to a random variable, but in place of (12) we now have

$$c'_K \hat{a}_K \Rightarrow N(0, c'\Lambda c) \overset{d}{=} c'\eta,$$

and the limit distribution is the same as that of the variate $c'\eta = \Sigma_1^\infty c_k \eta_k$ from the orthonormal representation (11).

(b) Part (c) of Theorem 3.3 shows that the t ratio $t_{c'_K \hat{a}}$ diverges as both K and $n \to \infty$. As in the fixed K regressor case, all of the fitted coefficients are statistically significant as $n \to \infty$, according to the usual regression t tests. However, the rate of divergence of the t ratio is greater in the case where $K \to \infty$ than it is when K is fixed. In other words, the regression coefficients appear more significant, not less significant, with the addition of regressors as $n \to \infty$. This is explained by the fact that the residual variance in the regression (9) tends in probability to zero when both K and $n \to \infty$, i.e., there is no residual variance from this regression in the limit, as indicated in part (b) of the Theorem. In effect, as $K, n \to \infty$, the regression (9) succeeds in reproducing the entire Loève-Karhunen representation of the limit process $B(\cdot)$ and thereby fully represents the dependent variable in the limit. The fact that the empirical regression fully captures the series representation in the limit is confirmed by the limiting regression R^2 of unity.

(c) In the same way as for Theorem 3.1—see Remark 3.2(d)—Theorem 3.3 can be extended to apply to more general stochastic processes than Brownian motion. The proof of Theorem 3.3 relies on the use of an extended probability space in which a strong invariance principle applies—see (22) in the Appendix. Strong invariance principles like (22) have been proved in the literature for standardized partial sums that converge to Brownian motion. These results can be extended to apply to linear diffusions, like $J_c(r)$ in (7) above, as shown in Lemma 6.3 below. It follows that Theorem 3.3 is valid for both integrated and near integrated time series with the appropriate changes in notation.

(d) Theorems 3.1 and 3.3 give results for empirical regressions like (9) where the regressors are the eigenfunctions that appear in the Loève-Karhunen representation. As discussed in Section 2, there are other representations of limiting processes like Brownian motion and diffusions that are of a similar form, but for which the coefficients may not be independent and/or the functions may not be linearly independent. In such cases, it is possible to develop a limit theory for the empirical regressions, but effects such as the possible collinearity of the regressors in the limit as $K \to \infty$ need to be taken into account. The earlier version of this paper (1996) explored such consequences for the case of an empirical version of equation (6) and can be obtained from the author on request.

4. WIENER PROCESS APPROXIMATION THEORY

The above analysis uses series of deterministic functions with random coefficients to represent stochastic processes like Brownian motion. It is of some interest to ask if the reverse is possible, viz. can we represent an arbitrary deterministic function on a certain interval in terms of stochastic processes? To deal with this question we will take a slightly different approach and try to approximate an arbitrary continuous function on the $[0,1]$ interval in terms of independent Brownian motion processes. The idea is analogous to that of the uniform approximation of a continuous function by polynomials or trigonometric functions. The following shows that there is, in fact, a Wiener process version of the famous Weierstrass approximation theorem.

4.1 THEOREM: *Let $f(\cdot)$ be any continuous function on the interval $[0,1]$, and let $\varepsilon > 0$ be arbitrarily small. Then we can find a sequence of independent standard Brownian motions $\{W_i\}_{i=1}^N$, and a sequence of random variables $\{d_i\}_{i=1}^N$ such that as $N \to \infty$,*

(a) $$\sup_{r \in [0,1]} \left| f(r) - \sum_{i=1}^{N} d_i W_i(r) \right| < \varepsilon \quad a.s.,$$

(b) $$\int_0^1 \left[f(r) - \sum_{i=1}^{N} d_i W_i(r) \right]^2 dr < \varepsilon \quad a.s.$$

4.2 REMARKS: (a) The Weierstrass approximation theorem tells us that any continuous function $f(r)$ can be uniformly approximated on the interval $[0,1]$ by a trigonometric polynomial of the form

$$\alpha_0 + \sum_{k=1}^{K} (\alpha_k \sin(2\pi k r) + \beta_k \cos(2\pi k r)). \tag{13}$$

In this series approximation, the coefficients $\{\alpha_k, \beta_k\}$ are nonrandom and the functions are deterministic continuous functions. In an analogous way, Part (a) of Theorem 4.1 shows that we can find a set of N independent Wiener processes on $C[0,1]$ and a sequence of N random variables such that, with probability one as $N \to \infty$, the function $f(r)$ can be uniformly approximated on the interval $[0,1]$ by the linear combination $\sum_{i=1}^{d} d_i W_i(r)$ of Wiener processes.

(b) Part (b) of Theorem 4.1 is sufficient to ensure that the system of Wiener processes $\{W_i\}_{i=1}^{\infty}$ is complete in $L_2[0,1]$ with probability one (e.g., see Tolstov (1976, p. 58)). It follows that, given any continuous function $f(r)$, we can find a sequence $\{W_i(r), d_i\}_{i=1}^{\infty}$ such that with probability one

$$\lim_{N \to \infty} \int_0^1 \left[f(r) - \sum_{i=1}^{N} d_i W_i(r) \right]^2 dr = 0, \tag{14}$$

and thus

$$f(r) \sim \sum_{i=1}^{\infty} d_i W_i(r) \tag{15}$$

in L_2. We may replace the Wiener processes $W_i(r)$ by orthogonal functions $V_i(r)$ in $L_2[0,1]$ using the Gram-Schmidt process, i.e.,

$$\begin{cases} V_1 = W_1, \\ V_2 = W_2 - \left(\int_0^1 W_2 V_1\right)\left(\int_0^1 V_1^2\right)^{-1} V_1, \\ V_3 = W_3 - \left(\int_0^1 W_3 V_a\right)\left(\int_0^1 V_a V_a'\right)^{-1} V_a, \quad V_a' = [V_1, V_2]. \end{cases} \tag{16}$$

In place of (14), we then have

$$\lim_{N \to \infty} \int_0^1 \left[f(r) - \sum_{i=1}^{N} e_i V_i(r) \right]^2 dr = 0$$

with probability one. By virtue of the orthogonality of the functions $\{V_i(r)\}$ in $L_2[0,1]$, we get the following stochastic Fourier representation in L_2:

$$f(r) \sim \sum_{i=1}^{\infty} e_i V_i(r), \quad \text{with} \quad e_i = \left(\int_0^1 f V_i\right)\left(\int_0^1 V_i^2\right)^{-1}, \tag{17}$$

and, with probability one, we have Parseval's equality,

$$\int_0^1 f^2 = \sum_{i=1}^{\infty} e_i^2 \left(\int_0^1 V_i^2\right),$$

holding, but now with random coefficients.

(c) We can apply the approximation theory of Theorem 4.1 to the sample path of an arbitrary Brownian motion $B(\cdot)$ on the interval $[0,1]$. Since the sample path of B is continuous, we can find a probability space such that Theorem 4.1 applies and then we have $B(r) \sim \sum_{i=1}^{\infty} d_i W_i(r)$ in the $L_2[0,1]$ sense. We formalize this as follows.

4.3 THEOREM: *Let $B(\cdot)$ be a Brownian motion on the interval $[0,1]$, and let $\varepsilon > 0$ be arbitrarily small. Then we can find a sequence of independent standard Brownian motions $\{W_i\}_{i=1}^{N}$ that are independent of B, and a sequence of random* *variables* $\{d_i\}_{i=1}^N$ *defined on an augmented probability space* $(\Omega, \mathcal{F}, P)$ *such that, as* $N \to \infty$,

$$\text{(a)} \qquad \sup_{r \in [0,1]} \left| B(r) - \sum_{i=1}^{N} d_i W_i(r) \right| < \varepsilon \quad a.s.(P),$$

$$\text{(b)} \qquad \int_0^1 \left[B(r) - \sum_{i=1}^{N} d_i W_i(r) \right]^2 dr < \varepsilon \quad a.s.(P),$$

$$\text{(c)} \qquad B(r) \sim \sum_{i=1}^{\infty} d_i W_i(r) \text{ in } L_2[0,1] \quad a.s.(P).$$

4.4 Remarks: (a) Part (c) of Theorem 4.3 shows that an arbitrary Brownian motion $B(\cdot)$ has an L_2 representation in terms of independent standard Brownian motions with random coefficients. As is clear from the proof of this theorem, the coefficients d_i are statistically dependent on $B(\cdot)$.

(b) Part (c) of Theorem 4.3 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 $B(r)$. The variables $W_i(r)$ are functions that take the form of Brownian motion 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 Brownian motion in terms of a collection of other independent Brownian motions. The coefficients in this model provide the connective tissue that relates these random functions.

(c) Let us now replace $\{W_i(r)\}$ by the orthogonal system $\{V_i(r)\}$ defined in (16). Then, in place of Part (c) we have, as in (17),

$$\text{(18)} \qquad B(r) \sim \sum_{i=1}^{\infty} e_i V_i(r), \qquad \text{with} \quad e_i = \left(\int_0^1 B V_i \right) \left(\int_0^1 V_i^2 \right)^{-1}.$$

(d) When we run an empirical regression of one random walk on a set of independent random walks, we reproduce a finite sample version of the model given in Part (c) of Theorem 6.3. Or, equivalently, if we transform the regressors so that they are orthogonal, then we reproduce a finite sample version of the representation (18).

4.5 Example: As an illustration, consider the quadratic function $f_1(r) = r^2$, for $-\pi \le r \le \pi$, combined with its periodic extension outside this interval. The Fourier series for this function is (c.f. Tolstov (1976, pp. 24–25))

$$r^2 \sim \frac{\pi^2}{3} - 4\left(\cos r - \frac{\cos 2r}{2^2} + \frac{\cos 3r}{3^2} - \cdots \right),$$

and this series converges to $f(r) = r^2$ in the interval $[-\pi, \pi]$ and to its periodic extension outside of this interval.

The function together with four terms of its Fourier series are shown in Figure 1. Figure 2 shows the same function with its approximation in terms of N independent Wiener processes with $N = 150$. The coefficients in the approximation are calculated using least squares regression of $f_1(r)$ on 1,000 observations generated from 125 independent random walks. With this number of terms, the Wiener process series captures the shape of the periodic quadratic function f_1 comparably well.

The purpose of this example is simply to illustrate the feasibility of the approximation by Wiener processes and to give some idea of the number of terms that are needed to achieve a level of approximation comparable to that of a Fourier series for a simple continuous function like $f_1(r)$. Needless to say, empirical regressions of the type (15) are not being recommended for practical use, nor do we develop a theory for the selection of the number of Wiener processes in such regressions.

4.6 Example: Finally, we consider the standard Gaussian random walk $y_t = \Sigma_{j=1}^{t} u_{0j}$, where $u_{0j} \equiv \text{iid } N(0,1)$. Let $x_t = (x_{kt}) = (\Sigma_{j=1}^{t} u_{kj})_{k=1}^{K}$ be K independent Gaussian random walks, all of which are independent of y_t. Consider the linear regression $y_t = \hat{b}_x' x_t + \hat{u}_t$, based on $n(>K)$ observations of these series. The large n asymptotic behavior of $\hat{b}_x$ is given by (Phillips (1986))

$$\hat{b}_x \Rightarrow \left[\int_0^1 W_x W_x'\right]^{-1}\left[\int_0^1 W_x W_y\right],$$

where W_x and W_y are the standard Brownian motion weak limits of the standardized partial sum processes $n^{-1/2}x_{[n\cdot]}$ and $n^{-1/2}y_{[n\cdot]}$, respectively.

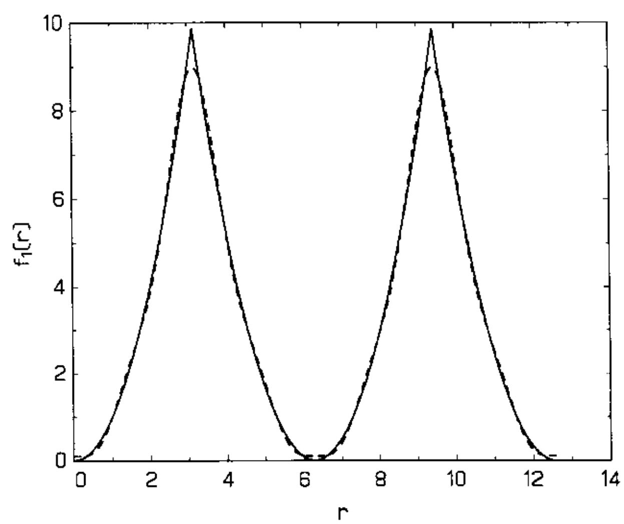

Figure 1.—Fourier series $f_1(r)$: 3 terms.

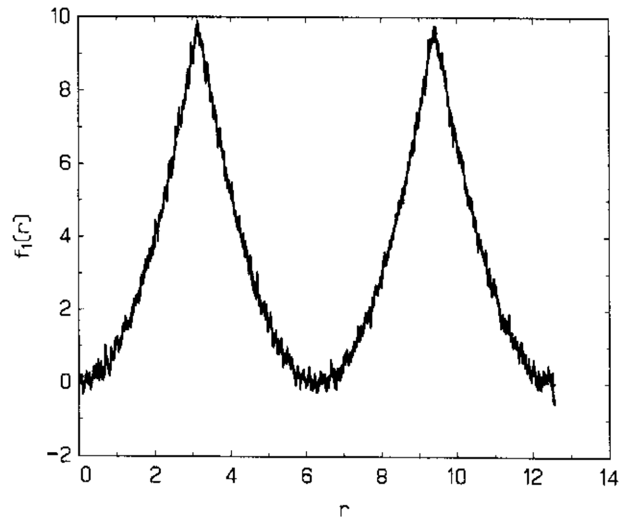

FIGURE 2.—$f_1(r)$: 125 Wiener terms.

Suppose we orthogonalize the regressors $\{x_k = (x_{kt})_1^n: k = 1, \ldots, K\}$ using the Gram Schmidt process

$$z_{1t} = x_{1t},$$

$$z_{2t} = x_{2t} - (x'_{2\cdot}x_{1\cdot})(x'_{1\cdot}x_{1\cdot})^{-1}x_{1t},$$

$$z_{3t} = x_{3t} - (x_{3\cdot}X_a)(X'_aX_a)^{-1}x_{at}, \qquad X_a := [x_{1\cdot}, x_{2\cdot}] := [x'_{a\cdot}], \qquad \text{etc.}$$

By standard weak convergence arguments, we find

$$n^{-1/2}z_{1[n\cdot]} \Rightarrow V_1(\cdot), \qquad n^{-1/2}z_{2[n\cdot]} \Rightarrow V_2(\cdot), \quad n^{-1/2}z_{3[n\cdot]} \Rightarrow V_3(\cdot), \quad \text{etc.}$$

Now let $z_t = (z_{kt})_1^K$, and consider the regression $y_t = \hat{b}'_{zK}z_t + \hat{u}_t$. In this case, writing $\hat{b}_{zK} = (\hat{b}_{zK})_1^K$, we have the limit

$$\hat{b}_{zk} \Rightarrow \left[\int_0^1 V_k^2\right]^{-1}\left[\int_0^1 V_k W_y\right] \overset{d}{=} e_k$$

as in (18). Thus, the empirical regression of y_t on z_t reproduces the first K terms in the orthonormal representation of the limit Brownian motion W_y in terms of an orthogonalized coordinate system formed from K independent standard Brownian motions. The regression t ratios are $t_{b_k} = \hat{b}_{zk}/s_{\hat{b}_{zk}}$ and these have the limiting behavior

$$n^{-1/2}t_{b_k} \Rightarrow \frac{e_k}{\left[\int_0^1 W_{VK}^2 / \int_0^1 V_k^2\right]^{1/2}},$$

where $W_{VK}(\cdot) = W_y(\cdot) - (\int_0^1 W_y V_K)(\int_0^1 V_K V_K')^{-1} V_K(\cdot)$, and $V_K(\cdot) = (V_k(\cdot))_{k=1}^K$. As in the case of deterministic regressors (cf. Theorem 4.1), the regression t ratios diverge at the rate $n^{1/2}$ (shown in Phillips (1986)), indicating certain significance of the regressors in the limit. Moreover, in view of (18), $\int_0^1 W_{VK}^2 \to 0$ a.s. as $K \to \infty$, and we can expect the divergence rate of these t ratios to increase when both $K, n \to \infty$. Figure 3 shows the sampling densities of the t ratio, t_{b_1} with $K = 1, 10, 20$ and $n = 100$ based on 30,000 simulations. The increase in the divergence rate of the t ratio as K increases is apparent in these graphs.

Finally, the behavior of the R^2 in the regression $y_t = \hat{b}_{zK}' z_t + \hat{u}_t$ is

$$R^2 \Rightarrow 1 - \int_0^1 W_{VK}^2 \Big/ \int_0^1 W_y^2, \quad \text{for fixed } K,$$

$$R^2 \xrightarrow{p} 1 \quad \text{when} \quad K \to \infty \quad \text{as} \quad n \to \infty.$$

It follows that the empirical spurious regression fully explains y_t in the limit when the number of independent random walk regressors goes to infinity.

5. CONCLUSION

This paper shows that there are mathematical models underlying the classic spurious regressions of a random walk on deterministic trends and the regression of a random walk on random walks. The empirical regressions just pick off the first few terms in the series representation of the stochastic process that is the weak limit of a suitably standardized version of the dependent variable in the regression. Moreover, it is shown that, if the number of regressors in such regressions is allowed to grow with the sample size (n), these empirical regres-

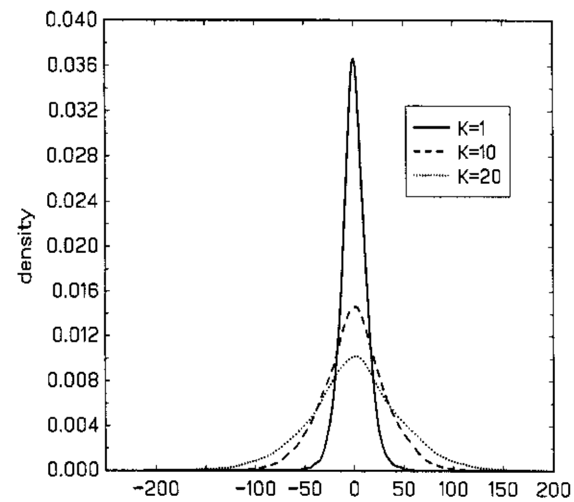

FIGURE 3.—Densities of t ratio t_{b_1}: $\eta = 100$.

sions succeed in accurately reproducing the full series representation in the limit as $n \to \infty$ and that the regression R^2 tends to unity. The theory also explains why it is natural in these regressions for the fitted coefficients to be random variables in the limit—they are exactly this in the underlying model! Thus, not only is there a valid mathematical model underlying such regressions, but the complete model is consistently estimable in the limit as $n \to \infty$.

While these results are definitive in that they fully explain what happens in regressions of this sort and why it happens, there is room for considerable debate about the implications of these results for empirical research. One viewpoint was clearly stated by a referee of the paper in the following way:

> "We use the term spurious regression in contrast to say the concept of cointegrated regressions, i.e., the possibility that certain sets of variables explain the trend of the dependent process in an economically sensible way. The fact that trending time series have valid representations in terms of other independent processes or deterministic functions of time is not of much interest from an economic viewpoint, unless it helps separate the wheat from the chaff."

In commenting on this orthodox view, I will make only two points here and leave it to future debate to take the discussion further. First, it needs to be emphasized that cointegrating regressions do not explain trends. Instead, they relate trends in multiple time series and thereby pass the trending behavior along to secondary variables that are usually also endogenous, leaving the trends themselves to be explained by unit roots, time polynomials, and trend breaks. As this paper shows, the trends themselves can be validly modelled in a variety of ways. Thus, the central issue addressed in this paper remains present in modern cointegration-based models of nonstationary time series. The second point is that the nature of trending mechanisms in economics is little understood and econometricians have little guidance from economic theory models about meaningful economic specifications. Were this not so, we would not be as heavily dependent as we presently are on unit root models, time polynomials, trend breaks, kernel regression fits and such like in capturing trends in empirical research. Against this background and with the current class of nonstationary models used in econometrics, it is virtually inevitable that the trending processes that appear in econometric models have little intrinsic economic meaning, even though the trends themselves may be of considerable economic interest. This paper shows that, even in the impoverished class of trending mechanism that we currently employ in empirical research, a limit theory of the trending process is possible and that it will often be based, in part at least, on a "limit theory of the sample period fit." This limit theory brings with it attendant qualifications such as those in the Introduction about the use of these mechanisms in a predictive context.

The results presented here have some implications for unit root modelling and testing. In recent years much of that literature has emphasized the importance of setting up a general maintained hypothesis that includes "alternative" specifications to a unit root model, such as deterministic trends and trend breaks. The results of this paper show that such specifications are not necessarily alternatives to a unit root model at all. Since unit root processes have limiting representations entirely in terms of these functions, it is apparent that we can mistakenly "reject" a unit root model in favor of a trend "alternative" when in fact that alternative model is nothing other than an alternate representation of the unit process itself. A development of the asymptotic theory in this case and a study of the impact of such considerations on empirical work are left for a future paper.

Cowles Foundation for Research in Economics, Department of Economics, P.O. Box 208281, New Haven, CT 06520, U.S.A.; phillips@econ.yale.edu; http://korora.econ.yale.edu

Manuscript received July, 1996; final revision received November, 1997.

APPENDIX

PROOFS

A.1 PROOF OF THEOREM 3.1: Since $\varphi_K([n\cdot]/n) \to \varphi_K(\cdot)$, we have

$$n^{-1}\sum_{t=1}^{n} \varphi_K(t/n)\varphi_K(t/n)' \to \int_0^1 \varphi_K \varphi_K' = I_K.$$

Then, using (10), we obtain $n^{-1}\Sigma_1^n \varphi_K(t/n) y_t/\sqrt{n} \Rightarrow \int_0^1 \varphi_K B$. Let Φ_K be the observation matrix of the regressors and let $y = (y_t)_1^n$ in (8). Then, we have

$$\text{(19)} \qquad c_K'\hat{a}_K = c_K'\left(\frac{\Phi_K'\Phi_K}{n}\right)^{-1}\left(\frac{1}{n}\Phi_K'\frac{y}{\sqrt{n}}\right) \Rightarrow c_K'\int_0^1 \varphi_K B$$

$$\stackrel{d}{=} N\left(0, c_K'\int_0^1\int_0^1 \varphi_K(r)(r\wedge s)\varphi_K(s)'\,ds\,dr\,c_K\right),$$

giving the stated result. Now let the orthonormal representation of the Brownian motion $B(\cdot)$ be given by $B(\cdot) = \Sigma_1^\infty \sqrt{\lambda_k}\,\varphi_k(\cdot)\xi_k$, where the ξ_k are iid $N(0,1)$ and λ_k is the eigenvalue of the covariance function $\sigma^2 r\wedge s$ corresponding to φ_k. Write this representation in the form

$$\text{(20)} \qquad B(\cdot) = \varphi_K(\cdot)'\Lambda_K^{1/2}\xi_K + \varphi_\perp(\cdot)'\Lambda_\perp^{1/2}\xi_\perp,$$

where the functions in $\varphi_\perp$ are all orthonormal and orthogonal to those in the vector φ_K, the elements of $\xi_\perp$ are all iid $N(0,1)$ and $\Lambda_K = \text{diag}(\lambda_1, \ldots, \lambda_K)$, $\Lambda_\perp = \text{diag}(\lambda_K, \lambda_K, \ldots)$. Using this representation of the Brownian motion $B(\cdot)$, we get

$$c_K'\int_0^1 \varphi_K B \stackrel{d}{=} c_K'\left(\int_0^1 \varphi_K\varphi_{K'}\right)\Lambda_K^{1/2}\xi_K = c_K'\Lambda_K^{1/2}\xi_K \stackrel{d}{=} N(0, c_K'\Lambda_K c_K)$$

as required for part (a). Note that the limiting form of the distribution also follows from a direct reduction of covariance matrix, viz.

$$\sigma^2\int_0^1\int_0^1 \varphi_K(r)(r\wedge S)\varphi_K(s)'\,ds\,dr = \int_0^1 \varphi_K(r)\varphi_K(r)'\,dr\,\Lambda_K = \Lambda_K.$$

For parts (b) and (c), define $t_{c_K'\hat{a}_K}=c_K'\hat{a}_K/s_{c_K'\hat{a}_K}$, where

(21) $$s^2_{c_K'\hat{a}_K}=\left(n^{-1}\sum_{t=1}^{n}(n^{-1/2}\hat{u}_t)^2\right)c_K'(\Phi_K'\Phi_K)^{-1}c_K.$$

A simple calculation reveals that

$$n^{-2}\sum_{t=1}^{n}\hat{u}_t^2\Rightarrow\int_0^1 B^2_{\varphi K},$$

where $B_{\varphi_K}(\cdot)=B(\cdot)-(\int_0^1 B\varphi_K')(\int_0^1\varphi_K\varphi_K')^{-1}\varphi_K(\cdot)$ is the L_2-projection residual of B on φ_K, giving part (b). Further,

$$ns^2_{c_K'\hat{a}_K}=\left(n^{-2}\sum_{t=1}^{n}\hat{u}_t^2\right)c_K'(n^{-1}\Phi_K'\Phi_K)^{-1}c_K\Rightarrow\int_0^1 B^2_{\varphi_K},$$

and we deduce that

$$n^{-1/2}t_{c_K'\hat{a}_K}=\frac{c_K'\hat{a}_K}{n^{1/2}s_{c_K'\hat{a}_K}}\Rightarrow\frac{c_K'\left[\int_0^1\varphi_K B\right]}{\left[\int_0^1 B^2_{\varphi_K}\right]^{1/2}},$$

as required for (c). The first half of part (d) follows immediately from (b) and the usual formula for the regression R^2. The second half of part (d) follows from the fact that

$$DW=\frac{\Sigma(n^{-1/2}\,\Delta\hat{u}_t)^2}{\Sigma(n^{-1/2}\hat{u}_t)^2}=\frac{(1/n^2)\Sigma(\Delta\hat{u}_t)^2}{(1/n)\Sigma(n^{-1/2}\hat{u}_t)^2}$$

$$=\frac{(1/n^2)\Sigma[u_t-\hat{b}_K'\,\Delta\varphi_{Kt}]^2}{(1/n)\Sigma(n^{-1/2}\hat{u}_t)^2}=O_p(n^{-1}).$$

A.2 PROOF OF THEOREM 3.3: First, note that $\Sigma_1^\infty c_k^2=1$, and $\Sigma_1^\infty\lambda_k=\int_0^1\gamma(r,r)\,dr=\int_0^1 r\,dr$. Hence, $\Sigma_1^\infty c_k^4<\infty$, and $\Sigma_1^\infty\lambda_k^2<\infty$. It follows that

$$c_K'\Lambda_K c_K=\sum_1^K c_k^2\lambda_k\le\left(\sum_1^K c_k^4\right)^{1/2}\left(\sum_1^K\lambda_k^2\right)^{1/2}\le\left(\sum_1^\infty c_k^4\right)^{1/2}\left(\sum_1^\infty\lambda_k^2\right)^{1/2}<\infty.$$

Thus, $c_K'\Lambda_K c_K$ is an increasing sequence that is bounded above and is therefore convergent. We write $\lim_{K\to\infty}c_K'\Lambda_K c_K=\sigma_c^2=c'\Lambda c$, say, where $c=(c_k)$, $\Lambda=\text{diag}(\lambda_1,\lambda_2,\dots)$ and $c'c=1$.

To prove part (a) we write, as in (19), $c_K'\hat{a}_K=c_K'(n^{-1}\Phi_K'\Phi_K)^{-1}(n^{-1}\Phi_K'y_K/n^{1/2})$. Using the Hungarian strong approximation (e.g., Csörgõ and Horváth (1993)) to the partial sum process $y_k=\Sigma_{i=1}^k u_j$, we can construct an expanded probability space with a Brownian motion $B(\cdot)$ for which

(22) $$\sup_{0\le k\le n}|y_k-B(k)|=o_{a.s.}(n^{1/p}),\quad\text{or}$$

$$\sup_{0\le k\le n}\left|\frac{y_k}{\sqrt{n}}-B\left(\frac{k}{n}\right)\right|=o_{a.s.}(1).$$

This gives the representation

$$\frac{y_{t-1}}{\sqrt{n}} = B\left(\frac{[nr]}{n}\right) + o_{a.s.}(1),$$

for $(t-1)/n \le r < t/n, t \ge 1$. It follows that we may write, as $n \to \infty$,

$$n^{-1}\sum_1^n \varphi_K\left(\frac{t}{n}\right)\left(\frac{y_t}{\sqrt{n}}\right) = \int_0^1 \varphi_K(r)B(r)\,dr + o_{a.s.}(1).$$

Also, since $K/n \to 0$ as $n \to \infty$, we have

$$n^{-1}\sum_1^n \varphi_K\left(\frac{t}{n}\right)\varphi_K\left(\frac{t}{n}\right)' = \int_0^1 \varphi_K(r)\varphi_K(r)'\,dr + o(1) = I_K + o(1),$$

leading to

$$\text{(23)} \qquad c_K'\hat{a}_K = c_K'[I_K + o(1)]^{-1}\left[\int_0^1 \varphi_K(r)B(r)\,dr + o_{a.s.}(1)\right]$$

$$= c_K'\int_0^1 \varphi_K(r)B(r)\,dr + o_{a.s.}(1).$$

Now use the orthonormal representation (20) of the Brownian motion $B(\cdot)$ in (23), and since the series converges uniformly we may integrate term by term, leading to

$$c_K'\hat{a}_K \overset{d}{=} c_K'\int_0^1 \varphi_K(r)[\varphi_K(r)'\Lambda_K^{1/2}\xi_K + \varphi_\perp(r)'\Lambda_\perp^{1/2}\xi_\perp]dr + o_{a.s.}(1)$$

$$= c_K'\Lambda_K^{1/2}\xi_K + c_K'\int_0^1 \varphi_K(r)\varphi_\perp(r)'\,dr\Lambda_\perp^{1/2}\xi_\perp\,dr + o_{a.s.}(1)$$

$$= c_K'\Lambda_K^{1/2}\xi_K + o_{a.s.}(1),$$

by virtue of the orthogonality of φ_K and the elements of $\varphi_\perp$. Now

$$c_K'\Lambda_K^{1/2}\xi_K \overset{d}{=} N(0, c_K'\Lambda_K c_K) \Rightarrow N(0, c'\Lambda c),$$

as $K \to \infty$. Thus, in the original probability space, when $K \to \infty$ as $n \to \infty$ with $K/n \to 0$, we have

$$\text{(24)} \qquad c_K'\hat{a}_K \overset{d}{=} c_K'\Lambda_K^{1/2}\xi_K + o_{a.s.}(1) \overset{d}{=} N(0, c_K'\Lambda_K c_K) + o_{a.s.}(1) \Rightarrow N(0, c'\Lambda c),$$

as required for part (a).

For parts (b) and (c), we have $n^{-1/2}t_{c_K'\hat{a}_K} = c_K'\hat{a}_K/(n^{1/2}s_{c_K'\hat{a}_K})$. The behavior of the numerator is given in (24). The square of the denominator is

$$ns^2_{c_K'\hat{a}_K} = \left(n^{-2}\sum_{t=1}^n \hat{u}_t^2\right)c_K'(n^{-1}\Phi_K'\Phi_K)^{-1}c_K.$$

Now

$$c_K'(n^{-1}\Phi_K'\Phi_K)^{-1}c_K = c_K'\left[\int_0^1 \varphi_K(r)\varphi_K(r)'\,dr + o(1)\right]^{-1}c_K = 1 + o(1),$$

as $n \to \infty$ for all K such that $K/n \to 0$. Next

$$\begin{aligned}\frac{1}{n^2}\sum_{t=1}^{n}\hat{u}_t^2 &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t}{\sqrt{n}}\right)^2 - \left(\frac{1}{n}\sum_{t=1}^{n}\frac{y_t}{\sqrt{n}}\varphi_K\left(\frac{t}{n}\right)'\right)\\ &\quad\times\left(\frac{1}{n}\sum_{t=1}^{n}\varphi_K\left(\frac{t}{n}\right)\varphi_K\left(\frac{t}{n}\right)'\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}\varphi_K\left(\frac{t}{n}\right)\frac{y_t}{\sqrt{n}}\right)\\ &= \left(\int_0^1 B(r)^2\,dr + o_{a.s.}(1)\right) - \left(\int_0^1 B(r)\varphi_K(r)'\,dr + o_{a.s.}(1)\right)\\ &\quad\times\left(\int_0^1 \varphi_K(r)\varphi_K(r)'\,dr + o(1)\right)^{-1}\left(\int_0^1 \varphi_K(r)B(r)\,dr + o_{a.s.}(1)\right)\\ &= \int_0^1 B_{\varphi_K}(r)^2\,dr + o_{a.s.}(1),\end{aligned}$$

where

$$\begin{aligned}(25)\qquad B_{\varphi_K}(r) &= B(r) - \left(\int_0^1 B\varphi_K'\right)\left(\int_0^1 \varphi_K\varphi_K'\right)^{-1}\varphi_K(r)\\ &= B(r) - \left(\int_0^1 B\varphi_K'\right)\varphi_K(r)\\ &= B(r) - \sum_{k=1}^{K}\left(\int_0^1 B(s)\varphi_k(s)\,ds\right)\varphi_k(r).\end{aligned}$$

But, $(\varphi_k)_1^\infty$ is a complete orthonormal system in $L_2[0,1]$ and, by virtue of Lemma 2.1, we have

$$(26)\qquad B(r) = \sum_1^\infty \varphi_k(r)\left[\int_0^1 \varphi_k(s)B(s)\,ds\right]$$

in quadratic means. It follows from (25) and (26) that, as $K \to \infty$, $B_{\varphi_K} \to 0$ in quadratic mean. Hence, as $K \to \infty$,

$$E\left[\int_0^1 B_{\varphi_K}(r)^2\,dr\right] \to 0,$$

and it follows that

$$n^{-2}\sum_{t=1}^{n}\hat{u}_t^2,\ ns^2_{c_K'\hat{a}_K} \xrightarrow{p} 0,$$

giving part (b). In consequence,

$$n^{-1/2}t_{c_K'\hat{a}_K} = \frac{c_K'\hat{a}_K}{n^{1/2}s_{c_K'\hat{a}_K}}$$

diverges as $n \to \infty$ when $K \to \infty$ and $K/n \to 0$, thereby establishing part (c). Part (d) follows directly from (b).

A.3 LEMMA: *Let $y_t = \sum_1^t e^{(t-j)c/n}u_j$ be a near integrated time series for some constant c, where u_t is stationary with zero mean, finite absolute moments to order $p > 2$ and has partial sums that satisfy the invariance principle $n^{-1/2}\sum_1^{[nr]}u_j \Rightarrow W(\cdot) \equiv BM(1)$. Then, there exists a probability space containing $\{y_t\}$ and a diffusion process $J_c(r) = \int_0^r e^{(r-s)c}\,dW(s)$ in which y_t satisfies the strong approximation*

$$\sup_{r\in[0,1]}\left|\frac{y_{[nr]}}{\sqrt{n}} - J_c(r)\right| = o_{a.s.}(1).$$

A.4 PROOF OF LEMMA 6.3: It is known (cf. Phillips (1987, Lemma 11)) that y_t satisfies the invariance principle $n^{-1/2}y_{[nr]} \Rightarrow J_c(r)$, and from (7) we have

$$J_c(r) = W(r) + c\int_0^r e^{(r-s)c}W(s)\,ds. \tag{27}$$

As in the proof of Theorem 3.2, use the Hungarian strong approximation (e.g., Csörgõ and Horváth (1993)) to the partial sum process $\sum_{i=1}^{k} u_j$ and construct an expanded probability space that contains $\{u_t, y_t\}$ and the Brownian motion $W(\cdot)$ and for which the following strong approximation holds:

$$\sup_{r\in[0,1]}\left|\frac{\sum_{i=1}^{[nr]}u_j}{\sqrt{n}} - W(r)\right| = o_{a.s.}(1).$$

Set $X_n(\cdot) = n^{-1/2}\sum_1^{[n\cdot]}u_j$, and write

$$\begin{aligned} n^{-1/2}y_{[nr]} &= n^{-1/2}\sum_1^{[nr]} e^{([nr]-j)c/n}u_j = n^{-1/2}\sum_1^{[nr]} e^{([nr]-j)c/n}\int_{(j-1)/n}^{j/n} dX_n(s) \\ &= \sum_1^{[nr]} e^{([nr]-j)c/n}\int_{(j-1)/n}^{j/n} dX_n(s) \\ &= \sum_1^{[nr]}\int_{(j-1)/n}^{j/n} e^{(r-s)c+\{([nr]/n-r)+(j/n-s)\}c}\,dX_n(s) \\ &= \int_0^r e^{(r-s)/c}\,dX_n(s)[1+o_{a.s.}(1)] \\ &= \int_0^r e^{(r-s)c}\,dX_n(s) + o_{a.s.}(1), \end{aligned} \tag{28}$$

since $e^{\{([nr]/n-r)+(j/n-s)\}c} = e^{O(n^{-1})} = 1+o(1)$ uniformly in $r\in[0,1]$, and in $s\in[(j-1)/n, j/n]$ and uniformly over $j=1,\dots,n$.

Next, apply integration by parts to the first term of (28) which is justified because $e^{(r-s)c}$ is continuous and $X_n(s)$ is of bounded variation for finite n. Hence

$$n^{-1/2}y_{[nr]} = X_n(r) + c\int_0^r e^{(r-s)c}X_n(s)\,ds + o_{a.s.}(1). \tag{29}$$

It follows from (27) and (29) that

$$\begin{aligned} \sup_{r\in[0,1]}\left|\frac{y_{[nr]}}{\sqrt{n}} - J_c(r)\right| &\le \sup_{r\in[0,1]}|X_n(r)-W(r)| \\ &\quad + \sup_{r\in[0,1]}\left[c\int_0^r e^{(r-s)c}\,ds\right]\sup_{s\in[0,1]}|X_n(s)-W(s)| + o_{a.s.}(1) \\ &= o_{a.s.}(1), \end{aligned}$$

giving a strong approximation for $n^{-1/2}y_{[nr]}$ in terms of the diffusion $J_c(r)$.

A.5 PROOF OF THEOREM 4.1: Let $\{W_i(r)\}$ be any sequence of independent Wiener processes in the $[0,1]$ interval. Using the series representation (4) for each process $W_i(r)$ in the sequence we may write

$$W_i(r) = \sqrt{2}\sum_{k=1}^{\infty}\frac{\sin[(k-1/2)\pi r]}{(k-1/2)\pi}\xi_{ik}, \tag{30}$$

where the ξ_{ij} are independent $N(0,1)$ variates. It is well known (cf. Tolstov (1976)) that the continuous function $f(r)$ can be uniformly approximated on the interval $[0,1]$ by a trigonometric polynomial of the form

$$\alpha_0 + \sum_{k=1}^{K} (\alpha_k \sin(2\pi kr) + \beta_k \cos(2\pi kr)).$$

Since $\{\sqrt{2}\sin[(k-1/2)\pi r]\}_{k=1}^{\infty}$ is a complete orthonormal system for $L_2[0,1]$, a slight modification to the proof of this approximation theorem (using a piecewise linear approximation to $f(r)$ and the fact that the Fourier series of a continuous, piecewise smooth, and arbitrarily close approximation to $f(r)$ is convergent uniformly—see, e.g., Tolstov (1976, Theorem 2, p. 81)) shows that the function $f(r)$ can also be uniformly approximated by a trigonometric polynomial of the form

$$\sum_{k=1}^{K} a_k \left(\frac{\sqrt{2}\sin[(k-1/2)\pi r]}{(k-1/2)\pi} \right) = a_K' \psi_K(r), \quad \text{say},$$

for some K; i.e., given $\varepsilon > 0$, there exist coefficients $(a_k)_{k=1}^{K}$ and some K for which

$$\text{(31)} \qquad \sup_{r\in[0,1]} \left| f(r) - \sum_{k=1}^{K} a_k \left(\frac{\sqrt{2}\sin[(k-1/2)\pi r]}{(k-1/2)\pi} \right) \right| < \frac{\varepsilon}{2}.$$

We now seek to combine (30) and (31) to produce an arbitrarily close approximation to $f(r)$ by Wiener processes. Given a fixed K for which (31) holds, we take a probability space on which the sequence $\{W_i(r)\}$ of Wiener processes and the random variables $\{\xi_{ij}\}$ are defined and we employ the representations

$$\text{(32)} \qquad W_i(r) = \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin[(k-1/2)\pi r]}{(k-1/2)\pi} \xi_{ik} = \psi_K(r)' \xi_{iK} + \psi_\perp(r)' \xi_{i\perp},$$

where $\xi_{iK}' = [\xi_{i1}, \dots, \xi_{iK}]$, $\xi_{i\perp}' = [\xi_{iK+1}, \xi_{iK+2}, \dots]$, and

$$\psi_\perp(r)' = \sqrt{2} \left[\frac{\sin[(K+1/2)\pi r]}{(K+1/2)\pi}, \frac{\sin[(K+3/2)\pi r]}{(K+3/2)\pi}, \dots \right].$$

As discussed earlier in connection with (4), the series (32) converge almost surely and uniformly in $r \in [0,1]$.

Taking the linear least squares approximation to the first term of (32) based on N observations $(i = 1, \dots, N)$, we obtain $\hat{\psi}_K = (\Xi_{KN}' \Xi_{KN})^{-1} \Xi_{KN}' W_N$, and

$$\hat{\psi}_K - \psi_K = \left(\frac{\Xi_{KN}' \Xi_{KN}}{N} \right)^{-1} \left(\frac{\Xi_{KN}' \Xi_{\perp N}}{N} \right) \psi_\perp = X_N' \psi_\perp, \quad \text{say},$$

where $\Xi_{KN}' = [\xi_{1K}, \dots, \xi_{NK}]$, $\Xi_{\perp N}' = [\xi_{1\perp}, \dots, \xi_{N\perp}]$, and $W_N = (W_i)_{N\times 1}$. The random variables ξ_{ij} in $[\Xi_{KN}, \Xi_{\perp N}]$ are iid $N(0,1)$. Hence, by the strong law of large numbers, as $N \to \infty$, we have $N^{-1}\Xi_{KN}'\Xi_{KN} \overset{a.s.}{\to} I_K$, and $N^{-1}\Xi_{KN}'\Xi_{\perp N} \overset{a.s.}{\to} 0$, so that $X_N \overset{a.s.}{\to} 0$, and $\hat{\psi}_K - \psi_K \overset{a.s.}{\to} 0$. Moreover, the strong convergence of $\hat{\psi}_K$ to ψ_K is uniform in $r \in [0,1]$. To see this, write

$$\begin{aligned}
|\hat{\psi}_K - \psi_K| &= [(\hat{\psi}_K - \psi_K)'(\hat{\psi}_K - \psi_K)]^{1/2} \\
&= [\psi_\perp'(X_N X_N')\psi_\perp]^{1/2} \le [\psi_\perp' \psi_\perp]^{1/2} [\lambda_{\max}(X_N X_N')]^{1/2} \\
&= \left[2 \sum_{k=K}^{\infty} \left(\frac{\sin[(k+\frac{1}{2})\pi r]}{(k+\frac{1}{2})\pi} \right)^2 \right]^{1/2} [\lambda_{\max}(X_N' X_N)]^{1/2} \\
&\le \left[\frac{2}{\pi^2} \sum_{k=K}^{\infty} \frac{1}{k^2} \right]^{1/2} [\lambda_{\max}(X_N' X_N)]^{1/2},
\end{aligned}$$

where $\lambda_{\max}(\cdot)$ signifies the largest eigenvalue of its argument matrix. Since $X_N \overset{a.s.}{\to} 0$ and $\lambda_{\max}(X_N' X_N)$ is a continuous function of the elements of X_N, we have $\lambda_{\max}(X_N' X_N) \overset{a.s.}{\to} 0$. It follows that

$$\sup_{r\in[0,1]} |\hat{\psi}_K - \psi_K| \le \left[\frac{2}{\pi^2} \sum_{k=K+1}^{\infty} \frac{1}{k^2} \right]^{1/2} [\lambda_{\max}(X_N' X_N)]^{1/2}$$
$$\le \left(\frac{1}{3}\right)^{1/2} [\lambda_{\max}(X_N' X_N)]^{1/2} \overset{a.s.}{\to} 0,$$

as $N \to \infty$. Hence, given $\delta > 0$ there exists (by Egoroff's theorem) a set C_δ with $P(C_\delta) > 1 - \delta$ and a number $N_\delta > 0$ for which

$$\sup_{r\in[0,1]} |\hat{\psi}_K - \psi_K| < \frac{\varepsilon}{2\sum_{k=1}^{K} |a_K|}$$

for all $N > N_\delta$. Then, we have

$$|f(r) - a_K' \hat{\psi}_K(r)| \le |f(r) - a_K' \psi_K(r)| + |a_K' \psi_K(r) - a_K' \hat{\psi}_K(r)|$$

and

$$\sup_{r\in[0,1]} |f(r) - a_K' \hat{\psi}_K(r)| \le \sup_{r\in[0,1]} |f(r) - a_K' \psi_K(r)| + \sup_{r\in[0,1]} |a_K' \psi_K(r) - a_K' \hat{\psi}_K(r)|$$
$$\le \frac{\varepsilon}{2} + \sum_{k=1}^{K} |a_k| \sup_{r\in[0,1]} |\psi_K(r) - \hat{\psi}_K(r)| \le \varepsilon, \tag{33}$$

for all $\omega \in C_\delta$.

Now note that we can write

$$a_K' \hat{\psi}_K(r) = a_K' (\Xi_{KN}' \Xi_{KN})^{-1} \Xi_{KN}' W_N = \sum_{i=1}^{N} d_i W_i(r) \tag{34}$$

with $d_i = a_K' (\Xi_{KN}' \Xi_{KN})^{-1} \xi_{iK}$. It follows from (33) and (34) that

$$\sup_{r\in[0,1]} \left| f(r) - \sum_{i=1}^{N} d_i W_i(r) \right| \le \varepsilon \quad \text{a.s.}$$

as $N \to \infty$, giving part (a) of the required result. Replacing ε by $\varepsilon^{1/2}$ in the above, part (b) follows immediately.

A.6 Proof of Theorem 4.3: Let $(\Omega_b = C[0,1], \mathscr{F}_b, P_b)$ be the probability space on which the Brownian motion $B(\cdot)$ is defined. Let $B(\cdot, \omega_b)$ be a sample path of B. There exists a C with $P_b(C) = 1$ such that, for all $\omega_b \in C$, the sample path $B(r, \omega_b)$ is continuous. Further, there exists a compact set C_K of $C[0,1]$ (under the sup norm) such that with arbitrarily large probability the sample paths $B(r, \omega_b)$ lie in C_K. Take any such $\omega_b \in C_K$. We can apply Theorem 4.1 to $B(r, \omega_b)$ noting that the theorem holds uniformly for continuous functions in a compact set like C_K.

We expand the probability space to the product space

$$(\Omega, \mathscr{F}, P) = (\Omega_b \times \Omega_W, \mathscr{F}_b \times \mathscr{F}_W, P_b \times P_W)$$

to include a sequence of independent standard Brownian motions $\{W_i\}_{i=1}^N$ (defined on $(\Omega_W, \mathscr{F}_W, P_W)$ and independent of B) and a sequence of random variables $\{d_i\}_{i=1}^N$ (defined on $(\Omega, \mathscr{F}, P)$) for which

$$\sup_{r\in[0,1]} \left| B(r, \omega_b) - \sum_{i=1}^{N} d_i W_i(r) \right| < \varepsilon, \tag{35}$$
$$\int_0^1 \left[B(r, \omega_b) - \sum_{i=1}^{N} d_i W_i(r) \right]^2 dr < \varepsilon \qquad \text{a.s.} \ (P_W)$$

as $N \to \infty$. This is possible for all $\omega_b \in C_K$ and, as is clear from the construction of the coefficients d_i in the proof of Theorem 4.1, we have the dependence $d_i = d_i(\omega_b, \omega_W)$ on the sample path $B(\cdot, \omega_b)$ as well as $\omega_W \in \Omega_W$, but the functions $\{W_i(r)\}$ are invariant to ω_b. We also have the dependence $N = N(\omega_b, \omega_W)$ on the sample path ω_b of the Brownian motion B and $\omega_W \in \Omega_W$. But, since Theorem 4.1(a) holds uniformly for f in a compact set, (35) holds for all $\omega_b \in C_K$ and N large enough. Since $P_b(C_K)$ is arbitrarily close to one, we deduce that given the Brownian motion $B(\cdot)$, there exist independent Brownian motions $\{W_i(r)\}$ and random coefficients $\{d_i\}$ that are defined on the augmented space $(\Omega, \mathcal{F}, P)$ for which, as $N \to \infty$, we have

$$\sup_{r \in [0,1]} \left| B(r) - \sum_{i=1}^{N} d_i W_i(r) \right| < \varepsilon, \qquad \int_0^1 \left[B(r) - \sum_{i=1}^{N} d_i W_i(r) \right]^2 dr < \varepsilon \qquad \text{a.s.} (P),$$

giving (a). Parts (b) and (c) follow directly.

Notation

$C[0,1]$	space of continuous functions on $[0,1]$.
$L_2[0,1]$	space of square integrable functions on $[0,1]$.
$\Rightarrow$	weak convergence.
$[\cdot]$	integer part of.
$r \wedge s$	$\min(r, s)$.
$\overset{a.s.}{\to}$	almost sure convergence.
$\overset{d}{=}$	distributional equivalence.
$:=$	definitional equality.
$o_{a.s.}(1)$	tends to zero almost surely.
$\overset{p}{\to}$	convergence in probability.

REFERENCES

Brieman, L. (1992): *Probability*. Philadelphia: SIAM.

Csörgő, M., and L. Horváth (1993): *Weighted Approximations in Probability and Statistics*. New York: Wiley.

Durlauf, S. N., and P. C. B. Phillips (1988): "Trends Versus Random Walks in Time Series Analysis," *Econometrica*, 56, 1333–1354.

Granger, C. W. J., and P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 74, 111–120.

Grenander, U., and M. Rosenblatt (1957): *Statistical Analysis of Stationary Time Series*. New York: Wiley.

Hendry, D. F. (1980): "Econometrics: Alchemy or Science?" *Economica*, 47, 387–406.

Hida, T. (1980): *Brownian Motion*. New York: Springer-Verlag.

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

Karatzas, I., and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, Second Edition. New York: Springer-Verlag.

Loève, M. (1963): *Probability Theory*. Third Edition. New York: Van Nostrand.

Phillips, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.

——— (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (1988): "Multiple Regression with Integrated Processes," in *Statistical Inference from Stochastic Processes, Contemporary Mathematics*, ed. by N. U. Prabhu, 80, 79–106.

——— (1991): "Spectral Regression for Cointegrated Time Series," in *Nonparametric and Semiparametric Methods in Economics and Statistics*, ed. by W. Barnett, J. Powell and G. Tauchen. New York: Cambridge University Press.

PHILLIPS, P. C. B., AND VICTOR SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

SHORACK, G. R., AND J. A. WELLNER (1986): *Empirical Processes with Applications to Statistics*. New York: Wiley.

TOLSTOV, G. P. (1976): *Fourier Series*. New York: Dover.

YULE, G. 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*, 89, 1–69.