COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1135

SPURIOUS REGRESSION UNMASKED

Peter C. B. Phillips

October 1996

SPURIOUS REGRESSION UNMASKED[1]

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

10 July 1996

[1] Invited Lecture presented at the XIV Latin American Meetings of the Econometric Society, Rio de Janeiro, August 5-9, 1996. 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, 1989. The computations were performed by the author in GAUSS and the paper was typed by the author in Scientific Word 2.5. Thanks go to the NSF for research support under Grant No. SES 94–22922.

Abstract

This paper argues that trending time series can admit valid regression representations even when the dependent variable and the regressors are statistically independent, i.e. in situations that are presently characterized in the literature as "spurious regressions". Our theory is directed mainly at the two classic examples of regressions of stochastic trends on time polynomials and regressions among independent random walks. But it has more general applicability and, we think, wider implications. Contrary to established wisdom, our theory justifies regressions of this type as valid models for the data. The radical conclusion that emerges from this study is that there are no spurious regressions for trending time series, just alternative valid representations of the limiting dependent variable process in terms of other stochastic processes and deterministic functions of time. We find statistical inference in such cases to be valid, not spurious, a conclusion that is in direct contrast to universal thinking about this subject since Yule (1926) first wrote about nonsense correlations.

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 UK. The latter 'relation' proved resilient to many econometric diagnostic tests and was humourously 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 non-specialists. The situation is especially difficult in cases where the data is trending - as indeed it is in both the examples above - because "third" factors that drive the trends come into play in the behaviour 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. 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. 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 that involve general stochastic trends, showing that the t- and F-ratio significance tests have divergent asymptotic behaviour in such regressions. Therefore, such outcomes are inevitable in large samples. Similar phenomena occur in regressions of stochastic trends on determistic 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.

From a contrarian perspective, one can argue that such regression outcomes are quite reasonable, given the shortcomings of the model specifications. For example, in the regression of a stochastic trend on deterministic time polynomials, it seems quite reasonable for conventional methods of statistical inference to signal that there is a trend in the dependent variable by casting the deterministic trends as 'significant' regressors, even though the fitted coefficients may be very small. Moreover, trends are an overriding characteristic of most economic time series and, to the extent that these trends are almost certainly imperfectly captured by empirical formulations, such outcomes seem very likely to be inevitable in applied econometric research. In this sense, trending mechanisms in regression can do good service as proxies for one another in empirical specifications that have endemic shortcomings, and they may not therefore deserve the pejorative connotation of a 'spurious' regression.

This paper puts forward a new perspective on spurious regressions that develops this line of argument to some logical conclusions. We seek to explain why significant regression coefficients occur in what seem to be manifestly incorrect regression specifications that 'spuriously' relate variables that may be 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. The radical conclusion that we reach is that, in contrast to conventional wisdom, there are no spurious regressions for trending time series. Our theory unmasks 'spurious' regressions as valid empirical regressions that capture different mathematical representations of the limiting form of the dependent process. A fascinating feature of this theory is that, just as we may model a continuous function by Fourier series in terms of different orthonormal system coordinates, so too may we validly model a trending process in various ways, including the use of 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 nothing other than the correct statistical manifestation of the existence of this underlying model. Thus, we find inference in such cases to be valid and not spurious, in direct contrast to universal thinking about this subject since Yule's original work.

The starting point in the approach that we adopt is the 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. An illustration for the important case of the regression of an integrated time series on a trigonometric polynomial is given in Section 4, and some associated issues of efficient regression are studied in Section 5. Section 6 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 7 concludes the paper. Proofs are collected together in Section 8 and notation is listed in Section 9.

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 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 $L_2[0,1]$, and let λ_k be the eigenvalues of $\gamma(r,s)$ corresponding to the functions φ_k, i.e. $\lambda_k\varphi_k(r) = \int_0^1 \gamma(r,s)\varphi_k(s)ds$. 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 on the interval $[0,1]$ has on this interval the orthogonal expansion*

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

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 its 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 that originate in the general form (1). The one we will work with is most easily developed as follows. Suppose $W(r)$ is a standard Brownian motion on $[0,1]$, and $V(r) = W(r) - rW(1)$ is the corresponding Brownian bridge process. 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 $V(r)$:

$$V(r)=\sqrt{2}\sum_{k=1}^{\infty}\frac{\sin(k\pi r)}{k\pi}\xi_k, \qquad \text{with } \xi_k=\sqrt{2}\int_0^1\frac{\sin(k\pi s)}{k\pi}V(s)ds. \tag{3}$$

The components ξ_k in this decomposition are independently and indentically distributed (iid) as $N(0,1)$, as can be verified by direct calculation. The representation (3) gives rise to a corresponding expansion for the Brownian motion $W(r)$, viz.

$$W(r)=r\xi_0+\sqrt{2}\sum_{k=1}^{\infty}\frac{\sin(k\pi r)}{k\pi}\xi_k, \tag{4}$$

with

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

The series (4) is known to converge almost surely and uniformly for $r\in[0,1]$ - e.g. Hida(1980, pp. 73, Remark 2), and Brieman (1992, p. 261), where the series are defined over the intervals $[0,2\pi]$, and $[0,\pi]$. The latter series may be obtained from (4) by means of a simple time dilation using $r=x/\pi$.

The representation (4) is one of many. For instance, we may replace the orthonormal system of trigonometric functions $\{\sqrt{2}\sin(k\pi r)\}_{k=1}^{\infty}$ by an orthonormal system of polynomials in r with its associated eigenvalues to produce alternative polynomial representations of $V(r)$ and $W(r)$. 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).

3. Reproduction of the Orthogonal Representation by 'Spurious' Regression

The existence of expansions like (3) and (4) 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 standardised 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 - see Durlauf and Phillips (1988).

In particular, 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

$$y_t=\sum_{k=1}^{K}\widehat{b}_k\varphi_k(\frac{t}{n})+\widehat{u}_t \tag{5}$$

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

$$\frac{y_t}{\sqrt{n}} = \sum_{k=1}^{K} \widehat{a}_k \varphi_k(\frac{t}{n}) + \frac{\widehat{u}_t}{\sqrt{n}} \tag{6}$$

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

$$\frac{y_{[n\cdot]}}{\sqrt{n}} \Rightarrow B(\cdot) \equiv BM(\sigma^2), \tag{7}$$

and the regressors φ_k form a complete orthonormal system in $L_2[0,1]$? All empirical applications of interest (including polynomial trends, trend breaks and sinusoidal trends) will be covered if we confine ourselves to orthonormal systems of piecewise continuous and differentiable functions φ_k.

In view of (2) and (7), we may very well expect that the regressors in (6) take on the role of the deterministic functions in the orthonormal representation of the limiting Brownian motion $B(\cdot)$. Perhaps, we can even go further than this. If $K \to \infty$ as $n \to \infty$, could (6) 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 $\widehat{a}_K = (\widehat{a}_k)$ be the coefficients and $\varphi_K = (\varphi_k)$ be the $K-$vector of regressors in (6). Let $c_K \in \mathbb{R}^K$ be any vector with $c_K'c_K = 1$, $t_{c_K'\widehat{a}_K}$ 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) $c_K'\widehat{a}_K \Rightarrow c_K'\left[\int_0^1 \varphi_K B\right] \stackrel{d}{=} N\left(0, c_K'\int_0^1\int_0^1 \varphi_K(r)\,(r\wedge s)\,\varphi_K(s)'dsdrc_K\right) = N(0, c_K'\Lambda_K c_K)$,

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

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

(d) $R^2 \Rightarrow 1 - \int_0^1 B_{\varphi_K}^2 / \int_0^1 B^2$, $\quad DW \xrightarrow{p} 0$,

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

3.2 REMARKS:

(a) Theorem 3.1 (a) shows that the fitted coefficients in the regression (6) 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, far from being a spurious regression, (6) 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 (6) as a partial but nonetheless correctly specified empirical version of an orthogonal representation of Brownian motion. We use the word 'partial' here because (6) has only K regressors, i.e. $\varphi_K = \left(\varphi_j\right)_{j=1}^{K}$. The model is correctly specified because the regressors that are omitted from (6), viz. $\varphi_\perp = \left(\varphi_{K+j}\right)_{j=1}^{\infty}$, are all orthogonal to the included variables. Hence, (6) 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 (5) is spurious and that the low DW statistic signals that inference is hazardous is inappropriate here.

(b) Part (c) of theorem 3.1 shows that the fitted coefficients are statistically significant with probability that goes to one as $n \to \infty$. Here, the t- ratios of the regression coefficients in (6) diverge at the rate $O_p(n^{1/2})$. The coefficients in (6) are not spuriously significant. The significant t-ratios correctly indicate the presence of the orthonormal representation

$$\begin{aligned} B(r) &= \sum_{k=1}^{\infty} \sqrt{\lambda_k}\varphi_k(r)\xi_k, \qquad \text{where } \xi_k \equiv iidN(0,1) \\ &= \sum_{k=1}^{\infty} \varphi_k(r)\eta_k, \qquad \text{where } \eta_k \equiv iidN(0,\lambda_k). \end{aligned} \tag{8}$$

In effect, the fitted regression (6) is an empirical model for (8). Setting $\eta_K = (\eta_k)_1^K$, we have

$$c_K'\widehat{a}_K \Rightarrow N(0, c_K'\Lambda_K c_K) \stackrel{d}{=} c_K'\eta_K. \tag{9}$$

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) An important feature of the true model (8) is that the coefficients η_k are random variables, whereas the variables $\varphi_k(r)$ are deterministic. The empirical regression (6) correctly reproduces this feature of the true model as $n \to \infty$, as is clear from (9).

(d) With some changes in notation, theorem 3.1 holds if the limiting behaviour of the dependent variable is a general continuous stochastic process $X(r)$ rather than Brownian motion. Suppose that for some $\alpha > 0$, $n^{-\alpha}y_{[n.]} \Rightarrow X(\cdot)$, a continuous stochastic process on $[0,1]$ with continuous covariance function $\gamma(r,s)$.

Instead of (6), we run the empirical regression

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

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

(i)

$$\begin{aligned} c_K'\widehat{a}_K &\Rightarrow c_K'\left[\int_0^1 \varphi_K X\right] \stackrel{d}{=} N\left(0, c_K'\left[\int_0^1\int_0^1 \varphi_K(r)\gamma(r,s)\varphi_K(s)'\right] c_K dsdr\right) \\ &= N(0, c_K'\Lambda_K c_K), \end{aligned}$$

(ii)

$$n^{-(1+2\alpha)}\sum_{t=1}^{n}\widehat{u}_t^2 \Rightarrow \int_0^1 X_{\varphi_K}^2,$$

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

(iii)

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

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 $X(r) = \int_0^r e^{(r-s)c}dW(s)$ for some constant c, and thereby (i)-(iii) 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 $n^{-1/2}y_{[n.]} \Rightarrow X(\cdot)$.

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* $c = (c_k)$, $\Lambda = diag(\lambda_1, \lambda_2, ...)$ *and* $c'c = 1$. *Moreover, if* $K \to \infty$ *and* $K/n \to 0$ *as* $n \to \infty$, *we have:*

(a) $c_K'\widehat{a}_K \Rightarrow N(0, \sigma_c^2)$;

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

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

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

3.4 REMARKS:

(a) Part (a) of theorem 3.3 gives the limiting distribution of the coefficients as both K and $n \to \infty$. In this case, c_K becomes infinite dimensional and $c_K' \widehat{a}_K$ becomes an l_2 inner product. As in the finite dimensional case, $c_K' \widehat{a}_K$ converges weakly to a random variable, but in place of (9) we now have

$$c_K' \widehat{a}_K \Rightarrow N(0, c'\Lambda c) \stackrel{d}{=} c'\eta,$$

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 (8).

(b) Part (c) of theorem 3.3 shows that the t-ratio $t_{c_K' \widehat{a}}$ diverges as both K and $n \to \infty$. As in the fixed K regressor case, all of the fitted coefficients are statistically significant as $n \to \infty$. 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 become 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 (6) 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 (6) 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) As in Remark 3.2 (d), theorem 3.3 can be extended to apply to more general stochastic processes than Brownian motion. But our proof of theorem 3.3 relies on the use of an extended probability space in which a strong invariance principle applies. Hence, the proof that we have given applies only in cases where such a result is valid.

4. A Brownian Motion Illustration

As an illustration, we will take an empirical regression that reproduces the explicit Brownian motion expansion given above in (4). This example is of additional interest because, as we discuss below, the deterministic functions that appear in (4) are not linearly independent. The complication of this example and the importance of the representation (4) make an extended discussion worthwhile.

We start with the following empirical regression

$$y_t = \widehat{b}_0 t + \sum_{k=1}^{K} \widehat{b}_k \frac{\sqrt{2} \sin\left[\frac{k\pi t}{n}\right]}{k\pi} + \widehat{u}_t \tag{10}$$

where $y_t = \Sigma_1^t u_s$ and $u_t \equiv iidN(0,1)$. Rewrite the fitted equation (10) in the following form with a normalization corresponding to (6), viz.

$$\frac{y_t}{\sqrt{n}} = \sqrt{n}\widehat{b}_0 \frac{t}{n} + \sum_{k=1}^{K} \frac{\widehat{b}_k}{\sqrt{n}} \frac{\sqrt{2} \sin\left[\frac{k\pi t}{n}\right]}{k\pi} + \frac{\widehat{u}_t}{\sqrt{n}} = \widehat{b}_{Kn}' g_K\left(\frac{t}{n}\right) + \frac{\widehat{u}_t}{\sqrt{n}}, \text{ say} \tag{11}$$

where $\widehat{b}'_{Kn} = \left[\sqrt{n}\widehat{b}_0, n^{-1/2}\widehat{b}_1, ..., n^{-1/2}\widehat{b}_K\right]$ and the regressor functions are defined by

$$g_K(r)' = \left[r, \frac{\sqrt{2}\sin \pi r}{\pi}, ..., \frac{\sqrt{2}\sin K\pi r}{K\pi}\right].$$

It is convenient to write the Wiener process representation (4) in the form $W(r) = g'_K\xi_K + g'_{K+}\xi_{K+}$, where

$$g_{K+}(r)' = \left[\frac{\sqrt{2}\sin(K+1)\pi r}{(K+1)\pi}, \frac{\sqrt{2}\sin(K+2)\pi r}{(K+2)\pi}, ...\right],$$

and $\xi'_{K+} = \left[\xi_{K+1,}\xi_{K+2}, ...\right]$. The limit distribution of the fitted coefficients and their t-ratios in (11) are given as follows.

4.1 THEOREM: *For fixed K as $n \to \infty$*

$$\begin{aligned} \widehat{b}_{Kn} &\Rightarrow \left[\int_0^1 g_K g'_K\right]^{-1}\left[\int_0^1 g_K W\right] \\ &\equiv N(0, I_{K+1}) + \left[\int_0^1 g_K g'_K\right]^{-1}\left[\int_0^1 g_K g'_{K+}\right]\xi_{K+}. && (12) \\ &\equiv N\left(0, I_{K+1} + h_K e_K e'_K\right), && (13) \end{aligned}$$

where W is standard Brownian motion,

$$h_K = \left[\frac{1}{3} - \frac{2}{\pi^2}\sum_{k=1}^{K}\frac{1}{k^2}\right]^{-2}\left[\frac{2}{90} - \frac{2}{\pi^4}\sum_{k=1}^{K}\frac{1}{k^4}\right],$$

and $e_K = [1, -\sqrt{2}, \sqrt{2}, -\sqrt{2}, ..., (-1)^K\sqrt{2}]$ is a $(K+1)-$vector. Let $c_K \in \mathbb{R}^K$ be any vector with $c'_K c_K = 1$. Let $t_{c'_K\widehat{b}_{Kn}}$ be the t-ratio for the fitted linear combination $c'_K\widehat{b}_{Kn}$ in (11). For fixed K as $n \to \infty$

(a) $c'_K\widehat{b}_{Kn} \Rightarrow N\left(0, 1 + h_K\left(c'_K e_K\right)^2\right)$,

(b) $n^{-1/2}t_{c'_K\widehat{b}_{Kn}} \Rightarrow c'_K\left[\int_0^1 g_K g'_K\right]^{-1}\left[\int_0^1 g_K W\right] / \left[\left(\int_0^1 W^2_{g_K}\right) c'_K\left[\int_0^1 g_K g'_K\right]^{-1} c_K\right]^{1/2}$

where $W_{g_K}(\cdot) = W(\cdot) - \left(\int_0^1 W g'_K\right)\left(\int_0^1 g_K g'_K\right)^{-1} g_K(\cdot)$ is the L_2-projection residual of W on g_K,

4.2 LEMMA: $\lim_{k\to\infty} Kh_K = \frac{1}{6}$.

It follows that for large K, $h_K \sim 1/6K$, and (13) is approximately $N\left(0, I_{K+1} + (1/6K)\, e_K e'_K\right)$. Then, individual elements of the fitted coefficient vector $\widehat{b}_{Kn}$ are approximately $N(0,1)$ variates when K is large. When $K = 0$, there is only one regressor in

(10) , $h_K = 1/5$, and (13) is simply $N(0, 6/5)$, an outcome that is readily confirmed by calculating the limit

$$\widehat{b}_{0n} = \sqrt{n}\widehat{b}_0 \Rightarrow \left[\int_0^1 r^2\right]^{-1}\left[\int_0^1 rW\right] \equiv N(0, \frac{6}{5}),$$

directly - see Durlauf and Phillips (1988) for this special case. In general, for $K \geq 0$, the limiting variance of the first element of $\widehat{b}_{Kn}$ (i.e. $\sqrt{n}\widehat{b}_0$) is $\sigma^2_{0K} = 1 + h_K$, and the variances of subsequent elements of $\widehat{b}_{Kn}$ (viz. $n^{-1/2}\widehat{b}_k$, $k > 0$) are all given by $\sigma^2_{kK} = 1 + 2h_K$.

Figure 1: Limiting Variance σ^2_{0K}

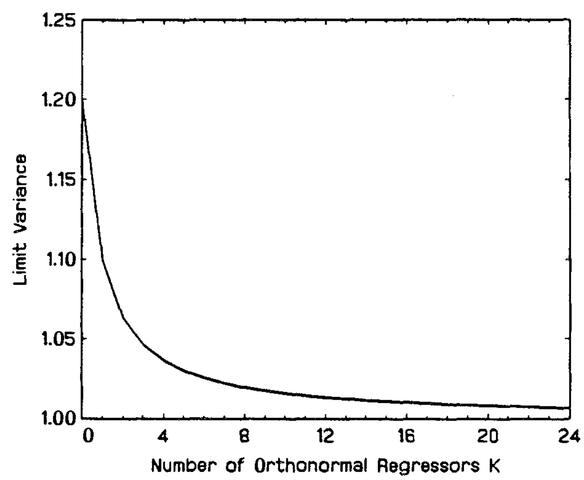


Fig. 1 graphs σ^2_{0K} as a function of K, showing the approach to unity as the number of regressors K in (10) becomes large. It follows that, for finite K, the empirical regression (10) only approximately reproduces the Wiener process representation given by (4). In particular, the random coefficients all have variances that are too large, i.e. $\sigma^2_{0K} = 1 + h_K$, and $\sigma^2_{kK} = 1 + 2h_K$ for $k > 0$. Moreover, in view of the form of the multivariate limit (13), there are linear combinations of the coefficients for which the variances are too large relative to the $iidN(0, 1)$ elements in the Wiener process representation even as $K \to \infty$. For example, let $\varepsilon_K = e_K / (e'_K e_K)^{1/2}$. Then

$$\varepsilon'_K \widehat{b}_{Kn} \Rightarrow N(0, 1 + h_K e'_K e_K) \equiv N(0, 1 + h_K(1 + 2K)) \Rightarrow N\left(0, \frac{4}{3}\right), \qquad (14)$$

as $K \to \infty$. Thus, in a certain sense, ordinary least squares regression on (10) is inadequate and does not fully reproduce the Wiener process representation even when $K \to \infty$. The remainder of this section studies this phenomenon.

One way of explaining the inadequacy of the least squares regression (10) is in terms of asymptotic collinearity in the regressors. The functions $\{\sqrt{2}\sin(k\pi r)\}_{k=1}^{\infty}$ constitute a complete orthonormal system for $L_2[0,1]$, and so the function in the first element of g_K, viz. $g_0(r)=r$, can itself be expanded in terms of $\{\sqrt{2}\sin(k\pi r)\}_{k=1}^{\infty}$. In particular, we have the following pointwise convergent Fourier sine series representation for $g_0(r)=r$ over the interval $[0,1)$ (e.g. Tolstov, 1976, pp. 27-28)

$$r=\frac{2}{\pi}\sum_{k=1}^{\infty}(-1)^{k-1}\frac{\sin k\pi r}{k}, \qquad \text{for } 0\le r<1. \tag{15}$$

Thus, the regressors in the fitted equation (11) become collinear in the limit as $K\to\infty$. This collinearity is formalized and shown to have some non-trivial consequences for the regression (11) in the following results.

4.3 LEMMA: *As $K\to\infty$, $e_K'\left[\int_0^1 g_K g_K'\right]e_K\to 0$, and the infinite matrix $\left[\int_0^1 g_K g_K'\right]_{K\to\infty}$ is positive semi-definite rather than positive definite.*

4.4 THEOREM: *Let $\varepsilon_K=e_K/\left(e_K'e_K\right)^{1/2}$, and $m_K=[m_L,0]$ be sequences of $(K+1)-$ vectors for which m_K has only a finite number of nonzero elements given in the L-vector m_L and $m_L'm_L=1$. Let $t_{c_K'\widehat{b}_{Kn}}$ be the usual least squares regression t-ratio for the estimated linear combination $c_K'\widehat{b}_{Kn}$. If $K\to\infty$ and $K/n\to 0$ as $n\to\infty$, then*

(a) $\varepsilon_K'\widehat{b}_{Kn}\Rightarrow N(0,4/3)$.

(b) $m_K'\widehat{b}_{Kn}\Rightarrow N(0,1)$.

(c) $n^{-1/2}t_{\varepsilon_K'\widehat{b}_{Kn}}, n^{-1/2}t_{m_K'\widehat{b}_{Kn}}$ *diverge.*

4.5 REMARKS:

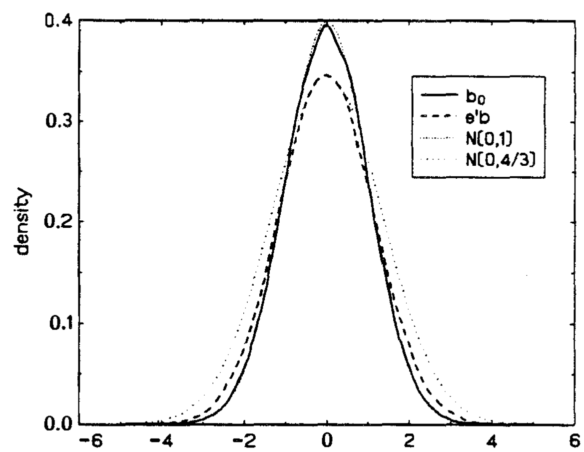
(a) Theorem 4.4 shows that there are linear combinations of the coefficients in the fitted regression (11)whose variances are larger than those of the Wiener process representation in the realistic situation where the number of regressors $K\to\infty$ as $n\to\infty$. This confirms the sequential ($n\to\infty$, then $K\to\infty$) asymptotics given in (14). However, these linear combinations must be infinite and involve all the regressors. As is apparent from Part (b) of the theorem, any finite linear combination of the coefficients has limiting variance that is the same as the coefficients in the Wiener representation.

(b) Lemma 4.3 reveals that the vector e_K is in the null space of the matrix $\int_0^1 g_K g_K'$ as $K\to\infty$. This is the direction where we can expect the regressors to be least informative, as this is the only source of the collinearity in the regressors. (As shown above in (15), it arises from the dependence of $g_0(r)=r$ on the sine series). The effect of the collinearity is to increase the variance of the fitted coefficients. The linear combination $\varepsilon_K'\widehat{b}_{Kn}$ has limiting variance $\frac{4}{3}$. Other linear combinations like $c_K'\widehat{b}_{Kn}$, where $c_K'c_K = 1$, have a limiting variance as $n \to \infty$ that is given by $1 + \frac{1}{3}c_K'\varepsilon_K \leq \frac{4}{3}$, with equality occuring when $c_K = \lambda\varepsilon_K$ and $\lambda = \pm 1$. Thus, the greatest increment in the variance of the random coefficients over that of the true Wiener representation is $33\frac{1}{3}\%$.

(c) An interesting aspect of the inefficiency is that individual coefficients in the fitted regression are all independent $N(0,1)$, yet some linear combinations have variance greater than unity. To be precise, the matrix normal distribution $N(0, I_{K+1} + h_K e_K e_K')$ generates individual component $iidN(0,1)$ variates as $K \to \infty$, but all linear combinations of the components except those that are othogonal to $\varepsilon_K = e_K / (e_K' e_K)^{1/2}$ have variance that is greater than unity as $K \to \infty$. This example shows that, in infinite matrix normal distributions, it is possible to have $iidN(0,1)$ components and still have dependence in the elements that produces a variance higher than unity for some infinite linear combinations of the elements.

(d) As shown in the proof of theorem 4.4, the residual variance $n^{-2}\Sigma_1^n \widehat{u}_t^2 \xrightarrow{p} 0$, as $K \to \infty$ and $n \to \infty$. Hence, the regression t-ratio statistics diverge as $n \to \infty$, and do so at a faster rate than they do when K is fixed (i.e faster than $O(n^{1/2})$, c.f. Part (b) of theorem 4.1). In other words, when K is fixed and when $K \to \infty$, significance tests in the empirical regression (11) correctly signal the validity of the representation Wiener process in terms of the sine series (4).

Figure 2: Densities of $\sqrt{n}\widehat{b}_0$ and $\varepsilon_K' b_{Kn}$ for $K = 75$, $n = 200$



4.6 SIMULATIONS:

We ran 30,000 simulations of the fitted model (11) with $K = 75$ and $n = 200$. Fig. 2 shows the sampling distributions of the first coefficient, $\sqrt{n}\widehat{b}_0$, and the linear combination $\varepsilon'_K\widehat{b}_{Kn}$ against the $N(0,1)$ and $N\left(0, \frac{4}{3}\right)$ limit distributions. Clearly, the sampling distribution of $\varepsilon'_K\widehat{b}_{Kn}$ has greater dispersion than that of $\sqrt{n}\widehat{b}_0$, as the limit theory predicts. In both cases, the limit theory given in theorem 4.4 gives a reasonable approximation, but is better for the individual coefficient than it is for the linear combination $\varepsilon'_K\widehat{b}_{Kn}$.

Figure 3: Densities of t-ratio of $\sqrt{n}\widehat{b}_0$

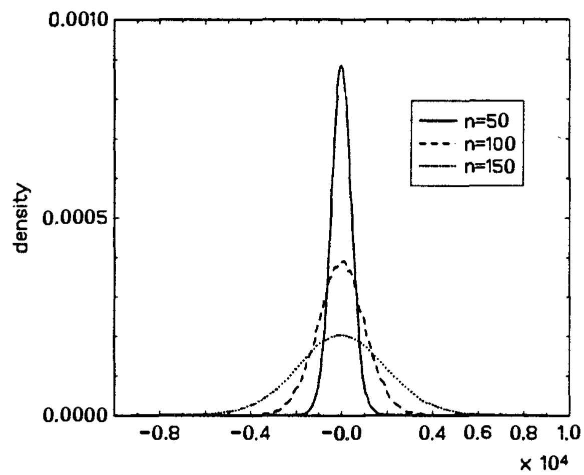


Fig. 3 shows the sampling densities of the t-ratio of $\sqrt{n}\widehat{b}_0$ for the parameter settings $(K = 25; n = 50, 100, 150)$. The substantial dispersion of these distributions and the divergence of the t-ratio statistic are evident in these simulation densities as n increases.

5. Efficient Regression and a Differential form of the Wiener Process Representation

In the previous section we used the collinearity of the regressors in (11) to explain the shortcomings of the fitted regression in fully reproducing the properties of the Wiener process representation. Note that the linear dependence of the functions in (4) applies only over the half open interval $[0, 1)$ because the Fourier series for r does not converge at $r = 1$. Thus, the linear term in (4) is not superfluous. Indeed, if we use (15) in (4) we get the alternative representation

$$W(r) = \begin{cases} \sqrt{2}\sum_{k=1}^{\infty} \frac{\sin(k\pi r)}{k\pi}\eta_k, & 0 \leq r < 1 \\ \xi_0 & r = 1 \end{cases}$$

where $\eta_k = \xi_k + (-1)^k \sqrt{2}\xi_0 \equiv N(0,3)$, and $E(\xi_0\eta_k) = (-1)^k \sqrt{2}$. Note that in this representation, the variance of the random coefficients increases to 3, and the coefficients are correlated. The random coefficient ξ_0 is needed to capture the value of the process $W(r)$ at its end point $r = 1$.

Another way of thinking about the regression (10), or (11) is that it is asymptotically statistically inefficient. To address the inefficiency, we could perform a generalized least squares regression, or equivalently transform the variables so that the regression is efficient. The dependent variable in (10) is the partial sum process $y_t = \Sigma_1^t u_j$, and u_j is a stationary time series. If the spectrum of u_j is continous we would expect, by the Grenander-Rosenblatt theorem (e.g. Grenander and Rosenblatt, 1957, p. 244), to get asymptotically efficient estimates from a regression of the time series u_j on appropriately transformed regressors. We pursue this intuition in what follows.

Transforming the variables in (10) by differencing, we get the new dependent variable $\Delta y_t = u_t$ and new regressors, comprised of an intercept and the cosine functions

$$\frac{\sin\frac{k\pi t}{n} - \sin\frac{k\pi(t-1)}{n}}{k\pi} \cong \frac{\left(\cos\frac{k\pi t}{n}\right)\frac{k\pi}{n}}{k\pi} = \frac{\cos\frac{k\pi t}{n}}{n} \quad k = 1, 2, ... \tag{16}$$

As in (10), we consider the fitted regression

$$\Delta y_t = \widetilde{b}_0 + \sum_{k=1}^{K} \widetilde{b}_k \frac{\sqrt{2}\cos\frac{k\pi t}{n}}{n} + \widetilde{u}_t \tag{17}$$

which we write in following standardized form that corresponds to (11) in differenced form

$$\frac{\Delta y_t}{\sqrt{n}} = \sqrt{n}\widetilde{b}_0\frac{1}{n} + \sum_{k=1}^{K}\frac{\widetilde{b}_k}{\sqrt{n}}\frac{\cos\frac{k\pi t}{n}}{n} + \frac{\widetilde{u}_t}{\sqrt{n}} = \widetilde{b}'_{Kn}\frac{\zeta_K\left(\frac{t}{n}\right)}{n} + \frac{\widetilde{u}_t}{\sqrt{n}} \tag{18}$$

where

$$\zeta_K\left(\frac{t}{n}\right)' = \left[1, \sqrt{2}\cos\frac{\pi t}{n}, \sqrt{2}\cos\frac{2\pi t}{n}, ..., \sqrt{2}\cos\frac{K\pi t}{n}\right],$$

and

$$\widetilde{b}'_{Kn} = \left[\sqrt{n}\widetilde{b}_0, \frac{\widetilde{b}_1}{\sqrt{n}}, ..., \frac{\widetilde{b}_K}{\sqrt{n}}\right].$$

Then we have

5.1 THEOREM:

(a) *For fixed K as $n \to \infty$,* $\widetilde{b}_{Kn} \Rightarrow \left[\int_0^1 \zeta_K(r)\zeta_K(r)'\right]^{-1}\left[\int_0^1 \zeta_K(r)\,dW(r)\right] \equiv N(0, I_{K+1})$.

(b) *The infinite matrix* $\left[\int_0^1 \zeta_K(r)\zeta_K(r)'\right]_{K\to\infty} = diag[1, 1, ...]$.

(c) *When* $K \to \infty$ *and* $K/n \to 0$ *as* $n \to \infty$, $c_K' \widehat{b}_{Kn} \Rightarrow N(0,1)$ *for any sequence* $c_K \in \mathbb{R}^K$ *for which* $c_K' c_K = 1$.

5.2 REMARKS:

(a) Theorem 5.1 shows that the fitted coefficients in the regression model (18) are asymptotically $N(0, I_{K+1})$ when the number of regressors is fixed, and that linear combinations of unit length of the coefficients are all $N(0,1)$ even when $K \to \infty$ as $n \to \infty$. Thus, the regression (18) reproduces accurately the Wiener process representation (4) in the following differential format

$$dW(r) = dr\xi_0 + \sqrt{2}\sum_{k=1}^{\infty}\cos(k\pi r)dr\xi_k. \tag{19}$$

Of course, the series (19) is purely formal and it should be interpreted as a way of writing

$$W(r+h) - W(r) = \left[h\xi_0 + \sqrt{2}\sum_{k=1}^{\infty}\frac{\sin k\pi(r+h) - \sin k\pi r}{k\pi}\xi_k\right]$$

for $h > 0$. Nevertheless, it is of some independent interest outside of the present context. In particular, let (19) hold over the two sided interval $[-1,1]$, and let n be an arbitrary positive integer. Evaluating the following integral formally term by term we get

$$\frac{1}{\sqrt{2}}\int_{-1}^{1} e^{-in\pi r}dW(r) = \frac{1}{\sqrt{2}}\int_{-1}^{1} e^{-in\pi r}\left[\xi_0 + \sqrt{2}\sum_{k=1}^{\infty}\cos(k\pi r)\xi_k\right]dr = \xi_n,$$

that is

$$\xi_n = \int_{-\pi}^{\pi} e^{-in\lambda}dZ(\lambda),$$

where $Z(\lambda) = W(\lambda/2\pi)$ is a process with independent increments and variance $E\left(dZ(\lambda)^2\right) = d\lambda/2\pi$. Hence, (19) is the formal inverse of the Cramér representation of the $iidN(0,1)$ process ξ_n.

(b) The reason why (18) is successful in reproducing the correct Wiener process representation is that the regressor functions $\{1, \sqrt{2}\cos\pi r, \sqrt{2}\cos 2\pi r, ...\}$ in (18) form a complete orthonormal system for $L_2[0,1]$, and so there are no redundant regressors in the fitted regression. In effect, the differential form (19) of the Wiener process is better suited to least squares regression. Similar comments will apply to empirical reproductions in terms of deterministic functions of other stochastic processes. However, in every case the most appropriate transformation of the representation will be supplied by the mapping that corresponds to the use of generalised least squares. More will be said of this in later work.

(c) Since the regressors form a complete orthonormal set, the infinite matrix

$$\left[\int_0^1 \zeta_K(r)\,\zeta_K(r)'\right]_{K\to\infty} = diag\,[1,1,...]$$

is positive definite. Thus, the regression (18) encounters none of the difficulties of (11).

(d) Interestingly, (18) and theorem 5.1 show that the best way to reproduce the form of a limiting stochastic trend in terms of its deterministic function representation is to do the regression using the stationary differenced components of the time series rather than run the regression in levels. This is because regressions of stationary time series on deterministic functions is asymptotically efficient (by the Grenander-Rosenblatt theory), whereas the same is not true of regressions involving stochastic trends - see Phillips and Lee (1997) for further discussion and illustrations of this point.

6. 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.

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

6.2 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

$$\alpha_0 + \sum_{k=1}^{K} (\alpha_k \sin(kr) + \beta_k \cos(kr))\,. \tag{20}$$

In this series approximation, the coefficients $\{\alpha_k, \beta_k\}$ are non random and the functions are deterministic continuous functions. In an analogous way, Part (a) of theorem 6.1 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 $\sum_{i=1}^{N} d_i W_i(r)$ of Wiener processes.

(b) Part (b) of theorem 6.1 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

$$\lim_{N\to\infty} \int_0^1 \left[f(r) - \sum_{i=1}^{N} d_i W_i(r) \right]^2 dr = 0, \tag{21}$$

and thus

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

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.

$$\left.\begin{array}{l} V_1 = W_1 \\ V_2 = W_2 - \left(\int_0^1 W_2 V_1\right)\left(\int_0^1 V_1^2\right)^{-1} V_1 \\ V_3 = W_3 - \left(\int_0^1 W_3 V_a\right)\left(\int_0^1 V_a^2\right)^{-1} V_a, \quad V_a' = [V_1, V_2] \\ \quad \text{etc.} \end{array}\right\} \tag{22}$$

In place of (21), 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

$$f(r) \sim \sum_{i=1}^{\infty} e_i V_i(r), \quad \text{with } e_i = \left(\int_0^1 f V_i\right)\left(\int_0^1 V_i^2\right)^{-1}, \tag{23}$$

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 6.1 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 6.1 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 formalise this as follows.

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

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

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

6.4 REMARKS:

(a) Part (c) of theorem 6.3 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 $B(\cdot)$.

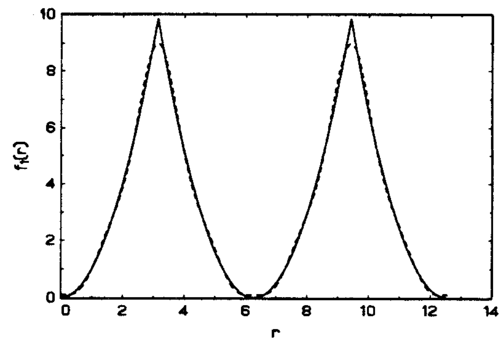
(b) Part (c) of theorem 6.3 also gives us a model for the classic 'spurious' regression of independent random walks. In this model, the role of the regressors and the coefficients becomes reversed. The coefficients d_i are random and they are co-dependent with the dependent variable $B(r)$. The variables $W_i(r)$ 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 (22). Then, in place of Part (c) we have, as in (23),

$$B(r) \sim \sum_{i=1}^{\infty} e_i V_i(r), \quad \text{with } e_i = \left(\int_0^1 BV_i \right) \left(\int_0^1 V_i^2 \right)^{-1}. \tag{24}$$

(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 othogonal, then we reproduce a finite sample version of the representation (24).

Figure 4: Fourier series $f_1(r)$: 3 terms



6.5 EXAMPLE:

As an illustration, consider the quadratic function $f_1(r) = r^2$, for $-\pi \leq r \leq \pi$, combined with its periodic extension outside this interval. The Fourier series for this function is (c.f. Tolstov, 1973, pp. 24-25)

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

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 is shown in Fig. 4. Fig. 5 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 quite well.

6.6 EXAMPLE:

Next, consider the function

$$f_2(r) = \begin{cases} \cos\frac{\pi r}{4} & \text{for} \quad 0 \leq r \leq 2 \\ 0 & \text{for} \quad 2 < r \leq 6 \\ \cos\frac{\pi r}{4} & \text{for} \quad 6 < r \leq 8 \end{cases}$$

combined with its periodic extension outside this interval. The Fourier series for this function is (c.f. Tolstov, 1973, pp. 37)

$$f_2(r) \sim \frac{1}{\pi} + \frac{1}{2}\cos\frac{\pi r}{4} - \frac{2}{\pi}\sum_{k=1}^{\infty}\frac{(-1)^k}{4k^2 - 1}\cos\frac{\pi k r}{2},$$

and this series converges to $f_2(r)$ in the interval $[0, 8]$ and to its periodic extension outside of this interval.

Figure 5: $f_1(r)$: 125 Wiener terms

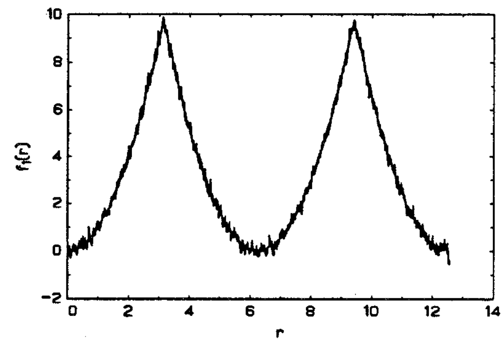
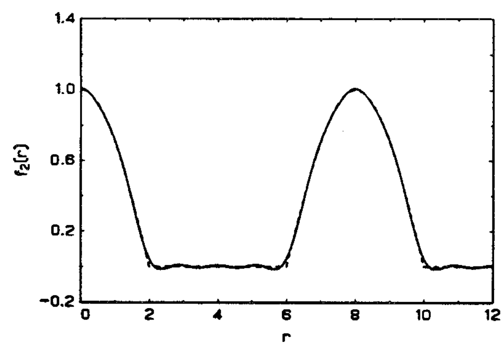


Figure 6: Fourier series $f_2(r)$: 3 terms

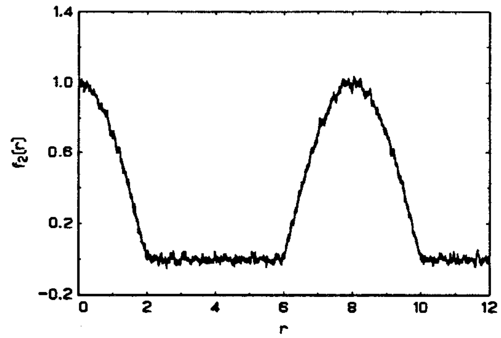


The function together with three terms of its Fourier series is shown in Fig. (6). Note that the fourier series is well suited to the function f_2 as the sinusoidal part is captured precisely by the second term of the series. Fig. 7shows the same function with its approximation in terms of N independent Wiener processes with $N = 125$. The coefficients in the approximation are calculated using least squares regression of $f_2(r)$ on $1,000$ observations generated from 125 independent random walks. With this number of terms, the Wiener process series adequately captures the sinusoidal component and the flat component of the function f_2.

6.7 EXAMPLE:

Finally, we consider the standard Gaussian random walk $y_t = \sum_{j=1}^{t} u_{0j}$, where $u_{0j} \equiv iidN(0,1)$. Let $x_t = (x_{kt}) = \left(\sum_{j=1}^{t} u_{kj}\right)_{k=1}^{K}$ be K independent Gaussian random walks all of which are independent of y_t. Consider the linear regression $y_t = \widehat{b}_x' x_t + \widehat{u}_t$, based on n $(> K)$ observations of these series. The large n asymptotic

Figure 7: $f_2(r)$: 125 Wiener terms



behaviour of $\widehat{b}_x$ is given by (Phillips, 1986)

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

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

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

$$\begin{aligned} z_{1t} &= x_{1t} \\ z_{2t} &= x_{2t} - (x_{2\cdot}'x_{1\cdot})\left(x_{1\cdot}'x_{1\cdot}\right)^{-1}x_{1t} \\ z_{3t} &= x_{3t} - (x_{3\cdot}X_a)\left(X_a'X_a\right)^{-1}x_{at}, \quad X_a := [x_{1\cdot}, x_{2\cdot}] := [x_{a\cdot}'] \\ &\text{etc.} \end{aligned}$$

By standard weak convergence arguments, we find

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

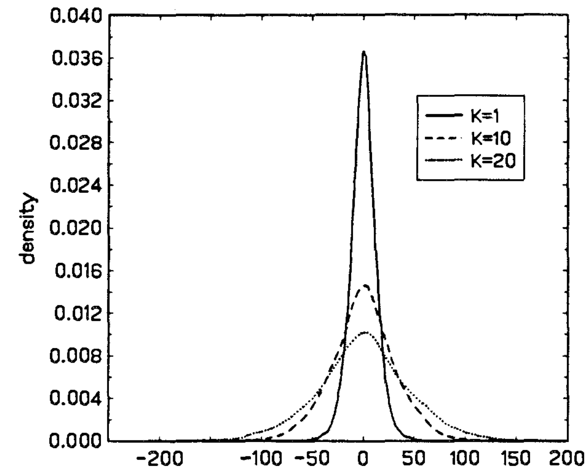
Now let $z_t = (z_{kt})_1^K$, and consider the regression $y_t = \widehat{b}_{zK}' z_t + \widehat{u}_t$. In this case, writing $\widehat{b}_{zK} = \left(\widehat{b}_{zk}\right)_1^K$, we have the limit

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

as in (24). 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 othogonalised coordinate system formed from K independent standard Brownian motions. The regression t-ratios are $t_{b_k} = \widehat{b}_{zk}/s_{\widehat{b}_{zk}}$ and these have the limiting behaviour

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

Figure 8: Densities of t-ratio $t_{b_1} : n = 100$



where $W_{VK}(\cdot) = W_y(\cdot) - \left(\int_0^1 W_y V_K\right)\left(\int_0^1 V_K V_K'\right)^{-1} V_K(\cdot)$, and $V_K(\cdot) = (V_k(\cdot))_{k=1}^K$. As in the case of deterministic regressors (c.f 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 (24), $\int_0^1 W_{VK}^2 \to 0$ $a.s$ as $K \to \infty$, and we can expect the divergence rate of these t-ratios to increase when both $K, n \to \infty$. Fig. 8 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 behaviour of the R^2 in the regression $y_t = \widehat{b}_{zK}' z_t + \widehat{u}_t$ is:

$$R^2 \Rightarrow 1 - \int_0^1 W_{VK}^2 / \int_0^1 W_y^2, \quad \text{for fixed } K,$$
$$R^2 \xrightarrow{p} 1 \quad \text{when } K \to \infty \text{ as } n \to \infty.$$

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.

7. Conclusion

The results presented in this paper put the old idea of a spurious regression in a very different perspective. In contrast to established thinking, we have shown that there is a clear mathematical model underlying such classic 'spurious' regressions as the regression of a random walk on deterministic trends. This mathematical model is based on the orthonormal representation of a continuous stochastic process in terms of deterministic functions. The idea is analogous to the Fourier series representation of a continuous function, but in the stochastic process case the Fourier coefficients are random variables. We have shown that regressions that have in the past been regarded as purely spurious are in fact nothing other than an approximation to these series representations. In effect, 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 standardised version of the dependent variable in the regression. Moreover, we have shown that, if the number of regressors in such regressions is allowed to grow with the sample size (n), these regressions succeed in accurately reproducing the full series representation in the limit as $n \to \infty$. Our theory 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, this model is consistently estimable in the limit as $n \to \infty$. The existence of this valid underlying model also explains why the coefficients in such regressions are found to be statistically significant. The fact that regression statistics like t-tests diverge as $n \to \infty$ has been interpreted in the past (including the author's own work, 1986) as a primary symptom of 'spurious' regression. Yet, this behaviour is simply a correct manifestation of the existence of the series representation. Indeed, as we have shown, the fact that the series representation exists and can be accurately reproduced in the limit as $n \to \infty$ increases the divergence rate of the regression t-tests.

We hope that these new results will help both theorists and practitioners move toward a more sympathetic understanding of empirical regressions with trending economic time series. The radical conclusion that emerges from this study is that there are no spurious regressions for trending time series, just alternative valid representations of the limiting dependent variable process in terms of other processes and deterministic functions of time. It is by no means accidental that such empirical relationships that have in the past been deemed 'spurious' are found, because they are the inevitable empirical manifestation of an underlying mathematical representation. However, just as some series representations of continuous functions converge faster than others, it turns out that some regression representations of trends will inevitably be better approximations than others.

The results presented here also have implications for unit root testing. In recent years much of the 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. Our results show that such specifications are not, in fact, really 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 favour of a trend 'alternative' when in fact that alternative model is nothing other than an alternate representation of the unit root process itself. A fuller study of the impact of such considerations on empirical work are left for a future paper.

8. Proofs

8.1 PROOF OF THEOREM 3.1: Since $\varphi_K([n\cdot]/n) \to \varphi_K(\cdot)$, we have $n^{-1}\sum_{t=1}^{n}\varphi_K(\frac{t}{n})\varphi_K(\frac{t}{n})' \to \int_0^1 \varphi_K\varphi_K' = I_K$. Then, using (7), we obtain $n^{-1}\Sigma_1^n\varphi_K(\frac{t}{n})y_t/\sqrt{n} \Rightarrow \int_0^1 \varphi_K B$. Let Φ_K be the observation matrix of the regressors and let $y = (y_t)_1^n$ in (5). Then, we have

$$\begin{aligned} c_K'\widehat{a}_K &= c_K'\left(\frac{\Phi_K'\Phi_K}{n}\right)^{-1}\left(\frac{1}{n}\Phi_K'\frac{y_K}{\sqrt{n}}\right) \Rightarrow c_K'\int_0^1 \varphi_K B \\ &\overset{d}{=} N\left(0, c_K'\int_0^1\int_0^1 \varphi_K(r)\,(r\wedge s)\,\varphi_K(s)'dsdrc_K\right), \end{aligned} \tag{25}$$

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 $r\wedge s$ corresponding to φ_k. Write this representation in the form

$$B(\cdot) = \varphi_K(\cdot)'\Lambda_K^{1/2}\xi_K + \varphi_\perp(\cdot)'\Lambda_\perp^{1/2}\xi_\perp, \tag{26}$$

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 $\Lambda_K = diag(\lambda_1,...,\lambda_K)$, $\Lambda_\perp = diag(\lambda_K, \lambda_K, ...)$. Using this representation of the Brownian motion $B(\cdot)$, we get

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

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

$$\int_0^1\int_0^1 \varphi_K(r)\,(r\wedge s)\,\varphi_K(s)'dsdr = \int_0^1 \varphi_K(r)\varphi_K(r)'dr\Lambda_K = \Lambda_K.$$

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

$$s^2_{c_K'\widehat{a}_K} = \left(n^{-1}\sum_{t=1}^{n}\left(n^{-1/2}\widehat{u}_t\right)^2\right)c_K'\left(\Phi_K'\Phi_K\right)^{-1}c_K. \tag{27}$$

A simple calculation reveals that $n^{-2}\sum_{t=1}^n \widehat{u}_t^2 \Rightarrow \int_0^1 B^2_{\varphi_K}$, where $B_{\varphi_K}(\cdot) = B(\cdot) - \left(\int_0^1 B\varphi_K'\right)\left(\int_0^1 \varphi_K\varphi_K'\right)^{-1}\varphi_K(\cdot)$ is the L_2 -projection residual of B on φ_K, giving Part (b). Further, $ns^2_{c_K'\widehat{a}_K} = \left(n^{-2}\sum_{t=1}^n\widehat{u}_t^2\right)c_K'\left(n^{-1}\Phi_K'\Phi_K\right)^{-1}c_K \Rightarrow \int_0^1 B^2_{\varphi_K}$,and we deduce that

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

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

$$
\begin{aligned}
DW &= \frac{\sum \left(n^{-1/2}\Delta\widehat{u}_t\right)^2}{\sum \left(n^{-1/2}\widehat{u}_t\right)^2} = \frac{\frac{1}{n^2}\sum \left(\Delta\widehat{u}_t\right)^2}{\frac{1}{n}\sum \left(n^{-1/2}\widehat{u}_t\right)^2} \\
&= \frac{\frac{1}{n^2}\sum \left[u_t - \widehat{b}_K'\Delta\varphi_{Kt}\right]^2}{\frac{1}{n}\sum \left(n^{-1/2}\widehat{u}_t\right)^2} = O_p\left(n^{-1}\right).
\end{aligned}
$$

8.2 PROOF OF THEOREM 3.2: First, note that $\Sigma_1^\infty c_k^2 = 1$, and $\Sigma_1^\infty \lambda_k = \int_0^1 \gamma(r,r)dr = \int_0^1 rdr$. Hence, $\Sigma_1^\infty c_k^4 < \infty$, and $\Sigma_1^\infty \lambda_k^2 < \infty$. It follows that

$$
c_K'\Lambda_K c_K = \Sigma_1^K c_k^2\lambda_k \leq \left(\Sigma_1^K c_k^4\right)^{1/2}\left(\Sigma_1^K \lambda_k^2\right)^{1/2} \leq \left(\Sigma_1^\infty c_k^4\right)^{1/2}\left(\Sigma_1^\infty \lambda_k^2\right)^{1/2} < \infty.
$$

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 $c = (c_k)$, $\Lambda = diag(\lambda_1, \lambda_2, ...)$ and $c'c = 1$.

To prove Part (a) we write, as in (25), $c_K'\widehat{a}_K = c_K'\left(n^{-1}\Phi_K'\Phi_K\right)^{-1}\left(n^{-1}\Phi_K' y_K/n^{1/2}\right)$. Using the Hungarian strong approximation (e.g. Csörgõ and Horváth,1993) to the partial sum process $y_k = \sum_{i=1}^k u_j$, we can construct a expanded probability space with a Brownian motion $B(\cdot)$ for which

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

or

$$
\sup_{0\leq k\leq n} \left|\frac{y_k}{\sqrt{n}} - B\left(\frac{k}{n}\right)\right| = o_{a.s.}(1). \tag{28}
$$

This gives the representation

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

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

$$
n^{-1}\Sigma_1^n \varphi_K\left(\frac{t}{n}\right)\left(\frac{y_t}{\sqrt{n}}\right) = \int_0^1 \varphi_K(r)B(r) + o_{a.s.}(1).
$$

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

$$
n^{-1}\Sigma_1^n \varphi_K\left(\frac{t}{n}\right)\varphi_K\left(\frac{t}{n}\right)' = \int_0^1 \varphi_K(r)\varphi_K(r)'dr + o(1) = I_K + o(1),
$$

leading to

$$
\begin{aligned}
c_K'\widehat{a}_K &= c_K'\left[I_K + o(1)\right]^{-1}\left[\int_0^1 \varphi_K(r)B(r)dr + o_{a.s.}(1)\right] \\
&= c_K'\int_0^1 \varphi_K(r)B(r)dr + o_{a.s.}(1).
\end{aligned} \tag{29}
$$

Now use the orthonormal representation (26) of the Brownian motion $B(\cdot)$ in (29), and since the series converges uniformly we may integrate term by term, leading to

$$
\begin{aligned}
c_K'\widehat{a}_K &\overset{d}{=} c_K' \int_0^1 \varphi_K(r)\left[\varphi_K(r)'\Lambda_K^{1/2}\xi_K + \varphi_\perp(r)'\Lambda_\perp^{1/2}\xi_\perp\right]dr + o_{a.s.}(1) \\
&= c_K'\Lambda_K^{1/2}\xi_K + c_K'\int_0^1 \varphi_K(r)\varphi_\perp(r)'dr\Lambda_\perp^{1/2}\xi_\perp dr + o_{a.s.}(1) = c_K'\Lambda_K^{1/2}\xi_K + o_{a.s.}(1),
\end{aligned}
$$

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

$$
c_K'\widehat{a}_K \overset{d}{=} c_K'\Lambda_K^{1/2}\xi_K + o_{a.s.}(1) \overset{d}{=} N(0, c_K'\Lambda_K c_K) + o_{a.s.}(1) \Rightarrow N(0, c'\Lambda c), \tag{30}
$$

as required for Part (a).

For Parts (b) and (c), we have $n^{-1/2}t_{c_K'\widehat{a}_K} = c_K'\widehat{a}_K / \left(n^{1/2}s_{c_K'\widehat{a}_K}\right)$. The behaviour of the numerator is given in (30). The square of the denominator is

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

Now

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

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

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

where

$$\begin{aligned} B_{\varphi_K}(r) &= B(r) - \left(\int_0^1 B\varphi_K'\right)\left(\int_0^1 \varphi_K\varphi_K'\right)^{-1}\varphi_K(r) \\ &= B(r) - \left(\int_0^1 B\varphi_K'\right)\varphi_K(r) \\ &= B(r) - \Sigma_{k=1}^K \left(\int_0^1 B(s)\varphi_k(s)ds\right)\varphi_k(r). \end{aligned} \tag{31}$$

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

$$B(r) = \Sigma_1^\infty \varphi_k(r)\left[\int_0^1 \varphi_k(s)B(s)ds\right] \tag{32}$$

in quadratic mean. It follows from (31) and (32) 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

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

giving Part (b). In consequence,

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

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

8.3 PROOF OF THEOREM 4.1: Working directly from (11), a standard weak convergence argument gives the limit $\widehat{b}_{Kn} \Rightarrow \left[\int_0^1 g_K g_K'\right]^{-1}\left[\int_0^1 g_K W\right]$. Since $W(r)$ has the decomposition $W(r) = g_K'\xi_K + g_{K+}'\xi_{K+}$, (12) follows immediately.

To prove (13), we proceed by direct calculation. First,

$$\begin{aligned} \int_0^1 g_K g_K' &= \begin{bmatrix} \int_0^1 r^2 & \frac{\sqrt{2}}{\pi}\int_0^1 r\sin\pi r & \cdots & \frac{\sqrt{2}}{K\pi}\int_0^1 r\sin K\pi r \\ \frac{\sqrt{2}}{\pi}\int_0^1 r\sin\pi r & \frac{2}{\pi^2}\int_0^1(\sin\pi r)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sqrt{2}}{K\pi}\int_0^1 r\sin K\pi r & 0 & \cdots & \frac{2}{K^2\pi^2}\int_0^1(\sin K\pi r)^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3} & \frac{\sqrt{2}}{\pi^2} & \cdots & \frac{\sqrt{2}(-1)^{K+1}}{K^2\pi^2} \\ \frac{\sqrt{2}}{\pi^2} & \frac{1}{\pi^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sqrt{2}(-1)^{K+1}}{K^2\pi^2} & 0 & \cdots & \frac{1}{K^2\pi^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & d_K' \\ d_K & D_K \end{bmatrix}, \text{ say} \end{aligned}$$

and

$$\int_0^1 g_K g'_{K+} = \begin{bmatrix} \frac{\sqrt{2}(-1)^{K+2}}{(K+1)^2\pi^2} & \frac{\sqrt{2}(-1)^{K+2}}{(K+2)^2\pi^2} & \cdots \\ 0 & 0 & \cdots \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \end{bmatrix} = \begin{bmatrix} f'_{K+} \\ 0 \end{bmatrix}, \text{ say.}$$

Then

$$\begin{aligned} \left[\int_0^1 g_K g'_K\right]^{-1}\left[\int_0^1 g_K g'_{K+}\right]\xi_{K+} &= \begin{bmatrix} \frac{1}{3} & d'_K \\ d_K & D_K \end{bmatrix}^{-1}\begin{bmatrix} f'_{K+} \\ 0 \end{bmatrix}\xi_{K+} \\ &= \begin{bmatrix} \left(\frac{1}{3} - d'_K D_K^{-1} d_K\right)^{-1} f'_{K+}\xi_{K+} \\ -D_K^{-1} d_K \left(\frac{1}{3} - d'_K D_K^{-1} d_K\right)^{-1} f'_{K+}\xi_{K+} \end{bmatrix} \\ &= e_K\left(\frac{1}{3} - d'_K D_K^{-1} d_K\right)^{-1} f'_{K+}\xi_{K+}, \qquad (33) \end{aligned}$$

where

$$e_K = \left[1, -d'_K D_K^{-1}\right]' = \left[1, -\sqrt{2}, \sqrt{2}, ..., (-1)^K\sqrt{2}\right]'. \qquad (34)$$

Next observe that

$$d'_K D_K^{-1} d_K = \frac{2}{\pi^2}\sum_{k=1}^{K}\frac{1}{k^2}, \text{ and } f'_{K+}f_{K+} = \frac{2}{\pi^4}\sum_{k=K+1}^{\infty}\frac{1}{k^4} = \frac{2}{90} - \frac{2}{\pi^4}\sum_{k=1}^{K}\frac{1}{k^4}.$$

It follows that

$$\left(\frac{1}{3} - d'_K D_K^{-1} d_K\right)^{-1} f'_{K+}\xi_{K+} \equiv N\left(0, \left[\frac{1}{3} - \frac{2}{\pi^2}\sum_{k=1}^{K}\frac{1}{k^2}\right]^{-2}\left[\frac{2}{90} - \frac{2}{\pi^4}\sum_{k=1}^{K}\frac{1}{k^4}\right]\right), \qquad (35)$$

and hence

$$\left[\int_0^1 g_K g'_K\right]^{-1}\left[\int_0^1 g_K g'_{K+}\right]\xi_{K+} \equiv N\left(0, \left[\frac{1}{3} - \frac{2}{\pi^2}\sum_{k=1}^{K}\frac{1}{k^2}\right]^{-2}\left[\frac{2}{90} - \frac{2}{\pi^4}\sum_{k=1}^{K}\frac{1}{k^4}\right] e_K e'_K\right).$$

Since ξ_K and ξ_{K+} are orthogonal, combining this last result with (12) gives (13). Part (a) follows immediately.

To prove Part (b), note that $n^{-1/2}t_{c'_K\widehat{b}_{Kn}} = c'_K\widehat{b}_{Kn}/\left(n^{1/2}s_{c'_K\widehat{b}_{Kn}}\right)$. As in (28) we have

$$ns^2_{c'_K\widehat{b}_{Kn}} = \left(n^{-2}\sum_{t=1}^{n}\widehat{u}_t^2\right) c'_K\left(\frac{1}{n}\sum_{t=1}^{n} g_K\left(\frac{t}{n}\right) g_K\left(\frac{t}{n}\right)'\right)^{-1} c_K.$$

Next,

$$c'_K\left(\frac{1}{n}\sum_{t=1}^{n} g_K\left(\frac{t}{n}\right) g_K\left(\frac{t}{n}\right)'\right)^{-1} c_K \to c'_K\left[\int_0^1 g_K g'_K\right]^{-1} c_K,$$

and $n^{-2}\sum_{t=1}^{n}\widehat{u}_t^2 \Rightarrow \int_0^1 W_{g_K}^2$, where $W_{g_K}(\cdot) = W(\cdot) - \left(\int_0^1 W g_K'\right)\left(\int_0^1 g_K g_K'\right)^{-1} g_K(\cdot)$ is the L_2 -projection residual of W on g_K. We deduce that

$$n^{-1/2} t_{c_K' \widehat{b}_{Kn}} \Rightarrow c_K' \left[\int_0^1 g_K g_K'\right]^{-1} \left[\int_0^1 g_K W\right] / \left[\left(\int_0^1 W_{g_K}^2\right) c_K' \left[\int_0^1 g_K g_K'\right]^{-1} c_K\right]^{1/2},$$

as required.

8.4 PROOF OF LEMMA 4.2: As $K \to \infty$, we have

$$\frac{1}{3} - \frac{2}{\pi^2}\sum_{k=1}^{K}\frac{1}{k^2} = \frac{2}{\pi^2}\sum_{k=K+1}^{\infty}\frac{1}{k^2} \to 0,$$

and

$$\frac{2}{90} - \frac{2}{\pi^4}\sum_{k=1}^{K}\frac{1}{k^4} = \frac{2}{\pi^4}\sum_{k=K+1}^{\infty}\frac{1}{k^4} \to 0.$$

However, the convergence of the latter sequence to zero is faster than the former. To see this, note that

$$\frac{2}{\pi^2(K+1)} = \frac{2}{\pi^2}\int_{K+1}^{\infty}\frac{1}{x^2} < \frac{2}{\pi^2}\sum_{k=K+1}^{\infty}\frac{1}{k^2} < \frac{2}{\pi^2}\int_{K}^{\infty}\frac{1}{x^2} = \frac{2}{\pi^2 K},$$

and similarly,

$$\frac{2}{3\pi^4(K+1)^3} = \frac{2}{\pi^4}\int_{K+1}^{\infty}\frac{1}{x^4} < \frac{2}{\pi^4}\sum_{k=K+1}^{\infty}\frac{1}{k^4} < \frac{2}{\pi^4}\int_{K}^{\infty}\frac{1}{x^4} = \frac{2}{3\pi^4 K^3}.$$

It follows that

$$\frac{\frac{2}{3\pi^4(K+1)^3}}{\left[\frac{2}{\pi^2 K}\right]^2} < \frac{\frac{2}{90} - \frac{2}{\pi^4}\sum_{k=1}^{K}\frac{1}{k^4}}{\left[\frac{1}{3} - \frac{2}{\pi^2}\sum_{k=1}^{K}\frac{1}{k^2}\right]^2} = \frac{\frac{2}{\pi^4}\sum_{k=K+1}^{\infty}\frac{1}{k^4}}{\left[\frac{2}{\pi^2}\sum_{k=K+1}^{\infty}\frac{1}{k^2}\right]^2} < \frac{\frac{2}{3\pi^4 K^3}}{\left[\frac{2}{\pi^2(K+1)}\right]^2}$$

Hence,

$$\lim_{K\to\infty} K\left[\frac{1}{3} - \frac{2}{\pi^2}\sum_{k=1}^{K}\frac{1}{k^2}\right]^{-2}\left[\frac{2}{90} - \frac{2}{\pi^4}\sum_{k=1}^{K}\frac{1}{k^4}\right] = \frac{1}{6},$$

as required.

8.5 PROOF OF LEMMA 4.3: First observe that since the elements of g_K are linearly independent for any finite value of K, we have

$$e_K'\left[\int_0^1 g_K g_K'\right] e_K = \int_0^1 \left(e_K' g_K\right)^2 > 0,$$

for any $e_K \in \mathbb{R}^K$ with $e_K \neq 0$. Hence $\int_0^1 g_K g_K'$ is positive definite for all finite K. Now, let $e_K' = [1, -d_K' D_K^{-1}] = \left[1, -\sqrt{2}, \sqrt{2}, ..., (-1)^K \sqrt{2}\right]$, as in (34). We find

$$\begin{aligned} e_K' \left[\int_0^1 g_K g_K'\right] e_K &= [1, -d_K' D_K^{-1}] \begin{bmatrix} \frac{1}{3} & d_K' \\ d_K & D_K \end{bmatrix} \begin{bmatrix} 1 \\ -D_K^{-1} d_K \end{bmatrix} \\ &= \frac{1}{3} - d_K' D_K^{-1} d_K = \frac{1}{3} - \frac{2}{\pi^2} \sum_{k=1}^{K} \frac{1}{k^2} \to 0, \end{aligned}$$

as $K \to \infty$, because $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$. Hence, $\int_0^1 g_K g_K'$ is positive semi-definite, not positive definite as $K \to \infty$.

8.6 PROOF OF THEOREM 4.4: As in the proof of theorem 3.2, we use the strong approximation (28), i.e. $\sup_{0 \le k \le n} \left| \frac{y_k}{\sqrt{n}} - B(\frac{k}{n}) \right| = o_{a.s.}(1)$, by expanding the probability space as necessary. Then, using the same argument as that leading to (29), we obtain

$$\begin{aligned} \widehat{b}_{Kn} &= \left[\int_0^1 g_K g_K'\right]^{-1} \left[\int_0^1 g_K W\right] + o_{a.s.}(1) \\ &\equiv N\left(0, I_{K+1} + h_K e_K e_K'\right) + o_{a.s.}(1) \end{aligned} \tag{36}$$

as in (13). It follows that $\varepsilon_K' \widehat{b}_{Kn} \equiv N\left(0, 1 + h_K \left(\varepsilon_K' e_K\right)^2\right) + o_{a.s.}(1)$. But,

$$h_K \left(\varepsilon_K' e_K\right)^2 = \frac{1}{6K} e_K' e_K = \frac{1 + 2K}{6K} \to \frac{1}{3},$$

as $K \to \infty$. Then, on the original probability space, letting $K \to \infty$ and $K/n \to 0$ as $n \to \infty$, we deduce that $\varepsilon_K' \widehat{b}_{Kn} \Rightarrow N(0, 1 + 1/3)$, giving Part (a) of the result.

To prove Part (b), use the representation (36) giving

$$m_K' \widehat{b}_{Kn} \equiv N\left(0, 1 + h_K \left(m_K' e_K\right)^2\right) + o_{a.s.}(1) \Rightarrow N(0, 1)$$

since

$$h_K \left(m_K' e_K\right)^2 = \frac{1}{6K} m_K' e_K = \frac{1}{6K} \left[m_1 + \sqrt{2} \sum_{k=2}^{L} (-1)^{k-1} m_k\right] \to 0,$$

as $K \to \infty$.

To prove Part (c), we first find the weak limit of the residual variance

$$\begin{aligned} \frac{1}{n^2} \sum_{t=1}^{n} \widehat{u}_t^2 &= \frac{1}{n} \sum_{t=1}^{n} \left(\frac{y_t}{\sqrt{n}}\right)^2 - \left(\frac{1}{n} \sum_{t=1}^{n} \frac{y_t}{\sqrt{n}} g_K\left(\frac{t}{n}\right)'\right) \\ &\quad \times \left(\frac{1}{n} \sum_{t=1}^{n} g_K\left(\frac{t}{n}\right) g_K\left(\frac{t}{n}\right)'\right)^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} g_K\left(\frac{t}{n}\right) \frac{y_t}{\sqrt{n}}\right) \end{aligned}$$

$$
\begin{aligned}
&= \left(\int_0^1 W(r)^2 dr + o_{a.s.}(1)\right) - \left(\int_0^1 W(r) g_K(r)' dr + o_{a.s.}(1)\right) \\
&\quad \times \left(\int_0^1 g_K(r) g_K(r)' dr + o(1)\right)^{-1} \left(\int_0^1 g_K(r) W(r) dr + o_{a.s.}(1)\right) \\
&= \int_0^1 W(r)^2 dr - \left(\int_0^1 W g_K'\right)\left(\int_0^1 g_K g_K'\right)^{-1}\left(\int_0^1 g_K W\right) + o_{a.s.}(1) \\
&= \int_0^1 W(r)^2 dr - \left(\int_0^1 W g_K'\right)\left(\int_0^1 g_K g_K'\right)^{-1}\left(\int_0^1 g_K g_K' \xi_K + \int_0^1 g_K g_{K+}' \xi_{K+}\right) + o_{a.s.}(1) \\
&= \int_0^1 W(r)^2 dr - \left(\int_0^1 W g_K'\right)\xi_K - \left(\int_0^1 W g_K'\right)\left(\int_0^1 g_K g_K'\right)^{-1}\left(\int_0^1 g_K g_{K+}' \xi_{K+}\right) + o_{a.s.}(1) \\
&= \int_0^1 W(r)^2 dr - \left(\int_0^1 W g_K'\right)\xi_K - \left(\int_0^1 W g_K'\right) e_K \left(\frac{1}{3} - d_K' D_K^{-1} d_K\right)^{-1} f_{K+}' \xi_{K+} + o_{a.s.}(1) \\
&= \int_0^1 W(r)^2 dr - \left(\int_0^1 W g_K' \xi_K\right) - \left(\int_0^1 W g_K' e_K\right) N(0, h_K) + o_{a.s.}(1)
\end{aligned}
$$

using (33) and (35) in the final two lines of the above argument. From lemma 4.2, $K h_K \to 1/6$, so that $\left(\int_0^1 W g_K' e_K\right) N(0, h_K) \xrightarrow{p} 0$, as $K \to \infty$. In view of the uniform convergence of the series representation $W(r) = g_K' \xi_K + g_{K+}' \xi_{K+}$ (c.f. (4) above), we deduce that $g_K' \xi_K \to W(r)$ $a.s$ and uniformly in r as $K \to \infty$. It follows that, if $K \to \infty$ and $K/n \to 0$ as $n \to \infty$, we have $n^{-2} \sum_{t=1}^n \widehat{u}_t^2 \xrightarrow{p} 0$. Then

$$
n s^2_{\varepsilon_K' \widehat{b}_{Kn}} = \left(\frac{1}{n^2} \sum_{t=1}^n \widehat{u}_t^2\right) \varepsilon_K' \left(\frac{1}{n} \sum_{t=1}^n g_K\left(\frac{t}{n}\right) g_K\left(\frac{t}{n}\right)'\right)^{-1} \varepsilon_K \xrightarrow{p} 0.
$$

In consequence,

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

diverges as $n \to \infty$ when $K \to \infty$ and $K/n \to 0$. A similar argument applies to the scaled t-ratio $n^{-1/2} t_{m_K' \widehat{a}_K}$, thereby establishing Part (c)

8.7 Proof of theorem 5.1: Direct least squares regression on (18) yields

$$
\begin{aligned}
\widetilde{b}_{Kn} &= \left[\frac{1}{n^2} \sum_{t=1}^n \zeta_K\left(\frac{t}{n}\right) \zeta_K\left(\frac{t}{n}\right)'\right]^{-1} \left[\frac{1}{n} \sum_{t=1}^n \zeta_K\left(\frac{t}{n}\right) \frac{\Delta y_t}{\sqrt{n}}\right] \\
&= \left[\frac{1}{n} \sum_{t=1}^n \zeta_K\left(\frac{t}{n}\right) \zeta_K\left(\frac{t}{n}\right)'\right]^{-1} \left[\sum_{t=1}^n \zeta_K\left(\frac{t}{n}\right) \frac{\Delta y_t}{\sqrt{n}}\right],
\end{aligned}
$$

and by standard weak convergence arguments we obtain

$$\widetilde{b}_{Kn} \Rightarrow \left[\int_0^1 \zeta_K(r)\,\zeta_K(r)'\right]^{-1}\left[\int_0^1 \zeta_K(r)\,dW(r)\right] \equiv N(0, I_{K+1}),$$

giving Part (a). Since $\{1, \sqrt{2}\cos\pi r, \sqrt{2}\cos 2\pi r, \ldots\}$ is a complete orthonormal system for $L_2[0,1]$ we have $\int_0^1 \zeta_K(r)\,\zeta_K(r)' = I_{K+1}$, and then the infinite matrix $\left[\int_0^1 \zeta_K(r)\,\zeta_K(r)'\right]_{K\to\infty} = diag[1,1,\ldots]$, which is positive definite, giving Part (b). To prove Part (c), we use the strong approximation (28) and obtain in the expanded probability space

$$\begin{aligned} c_K'\widetilde{b}_{Kn} &= c_K'\left[\int_0^1 \zeta_K(r)\,\zeta_K(r)' + o_{a.s.}(1)\right]^{-1}\left[\int_0^1 \zeta_K(r)\,dW(r) + o_{a.s.}(1)\right] \\ &= c_K'\left[I_{K+1} + o_{a.s.}(1)\right]^{-1}\left[\int_0^1 \zeta_K(r)\,dW(r) + o_{a.s.}(1)\right] \\ &= c_K'\left[\int_0^1 \zeta_K(r)\,dW(r)\right] + o_{a.s.}(1) \\ &\equiv N(0,1) + o_{a.s.}(1). \end{aligned}$$

It follows that in the original probability space $c_K'\widetilde{b}_{Kn} \Rightarrow N(0,1)$, when $K\to\infty$ and $K/n \to 0$ as $n\to\infty$, giving Part (c).

8.8 Proof of Theorem 6.1: Let $\{W_i(r)\}$ be any sequence of independent Wiener processes on the $[0,1]$ interval. Using the series representation (4) for each process $W_i(r)$ in the sequence we may write

$$W_i(r) = r\xi_{i0} + \sqrt{2}\sum_{k=1}^{\infty}\frac{\sin(k\pi r)}{k\pi}\xi_{ik}, \tag{37}$$

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

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

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

$$a_0 r + \sum_{k=1}^{K} a_k\left(\frac{\sqrt{2}\sin(k\pi r)}{k\pi}\right) = a_K'\psi_K(r), \quad \text{say,}$$

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

$$\sup_{r\in[0,1]}\left|f(r) - a_0 r + \sum_{k=1}^{K} a_k \left(\frac{\sqrt{2}\sin(k\pi r)}{k\pi}\right)\right| < \frac{\varepsilon}{2}. \tag{38}$$

We now seek to combine (37) and (38) to produce an arbitrarily close approximation to $f(r)$ by Wiener processes. Given a fixed K for which (38) 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

$$W_i(r) = r\xi_{i0} + \sqrt{2}\sum_{k=1}^{\infty}\frac{\sin(k\pi r)}{k\pi}\xi_{ik} = \psi_K(r)'\,\xi_{iK} + \psi_\perp(r)'\,\xi_{i\perp}, \tag{39}$$

where $\xi_{iK}' = [\xi_{i0}, \xi_{i1}, ..., \xi_{iL}]$, $\xi_{i\perp}' = [\xi_{iL+1}, \xi_{iL+2}, ...]$, and

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

The series (39) are known to converge almost surely and uniformly in $r \in [0,1]$ - e.g. see Hida (1980, p.73).

Taking the linear least squares approximation to the first term of (39) based on N observations $(i = 1, ..., N)$, we obtain $\widehat{\psi}_K = (\Xi_{KN}'\Xi_{KN})^{-1}\Xi_{KN}'W_N$, and

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

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

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

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

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

as $N\rightarrow\infty$. 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

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

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

$$\left|f(r)-a_K'\widehat{\psi}_K(r)\right|\leq\left|f(r)-a_K'\psi_K(r)\right|+\left|a_K'\psi_K(r)-a_K'\widehat{\psi}_K(r)\right|$$

and

$$\sup_{r\in[0,1]}\left|f(r)-a_K'\widehat{\psi}_K(r)\right| \leq \sup_{r\in[0,1]}\left|f(r)-a_K'\psi_K(r)\right|+\sup_{r\in[0,1]}\left|a_K'\psi_K(r)-a_K'\widehat{\psi}_K(r)\right|$$
$$\leq \frac{\varepsilon}{2}+\sum_{k=1}^{K}|a_k|\sup_{r\in[0,1]}\left|\psi_K(r)-\widehat{\psi}_K(r)\right|\leq\varepsilon, \tag{40}$$

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

Now note that we can write

$$a_K'\widehat{\psi}_K(r)=a_K'\left(\Xi_{KN}'\Xi_{KN}\right)^{-1}\Xi_{KN}'W_N=\sum_{i=1}^{N}d_iW_i(r) \tag{41}$$

with $d_i=a_K'\left(\Xi_{KN}'\Xi_{KN}\right)^{-1}\xi_{iK}$. If follows from (40) and (41) that

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

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

8.9 PROOF OF THEOREM 6.3: Let $(\Omega_b=C[0,1],\mathcal{F}_b,P_b)$ 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 set C with $P_b(C)=1$ such that, for all $\omega_b\in C$, the sample path $B(r,\omega_b)$ is continuous. Take any such $\omega_b\in C$. We can apply theorem 6.1 to $B(r,\omega_b)$. We expand the probability space to the product space

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

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

$$\sup_{r\in[0,1]}\left|B(r,\omega_b) - \sum_{i=1}^N d_i W_i(r)\right| < \varepsilon, \quad \int_0^1 \left[B(r,\omega_b) - \sum_{i=1}^N d_i W_i(r)\right]^2 dr < \varepsilon \quad a.s.\,(P_W) \tag{42}$$

as $N \to \infty$. This is possible for all $\omega_b \in C$ and, as is clear from the construction of the coefficients d_i in the proof of theorem 6.1, 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. Since (42) holds for all $\omega_b \in C$ as $N \to \infty$, 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

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

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

9. Notation

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

References

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

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

Durlauf, S. N. and P. C. P. Phillips (1988). "Trends versus random walks in time series anaylysis," *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.* Wiley: New York.

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

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

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.* 2nd Edition. Springer-Verlag: New York.

Loève, M. (1963). *Probability Theory.* 3'rd Edition. Van Nostrand: New York.

Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311-340.

Phillips, P. C. B. and C. C. Lee (1997) "Efficiency Gains from Quasi-Differencing under Nonstationarity" in P. M. Robinson and M. Rosenblatt (Eds.) *Essays in memory of E. J. Hannan* (forthcoming).

Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics.* Wiley: New York.

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

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.