COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 757

UNDERSTANDING SPURIOUS REGRESSIONS IN ECONOMETRICS

by

P. C. B. Phillips

July 1985

Revised: December 1985

UNDERSTANDING SPURIOUS REGRESSIONS IN ECONOMETRICS

by

P. C. B. Phillips*

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

This paper provides an analytical study of linear regressions involving the levels of economic time series. An asymptotic theory is developed for regressions that relate quite general integrated random processes. This includes the spurious regressions of Granger and Newbold (1974) and the recent cointegrating regressions of Granger and Engle (1985). An asymptotic theory is developed for the regression coefficients and for conventional significance tests. It is shown that the usual t and F ratio test statistics do not possess limiting distributions in this context but actually diverge as the sample size $T \uparrow \infty$. The limiting behavior of regression diagnostics such as the Durbin-Watson statistic, the coefficient of determination and the Box-Pierce statistic is also analyzed. The theoretical results that we present explain many of the earlier simulation findings of Granger and Newbold (1974, 1977).

May, 1985

Revised: December, 1985

---

*My thanks, as always, go to Glena Ames for her skill and effort in typing the manuscript of this paper. The research reported here was undertaken during the author's tenure of a John Simon Guggenheim Fellowship. My thanks go to the Guggenheim Foundation for their financial support during 1984/85. Thanks are also due to the NSF for support under Grant No. SES 8218792.

1. INTRODUCTION

In an important, influential and frequently cited article in this Journal, Granger and Newbold (1974) examined some of the likely empirical consequences of nonsense or spurious regressions in econometrics. Many of the points made by Granger-Newbold center on the classic textbook warning about the presence of serially correlated errors invalidating conventional procedures of inference in regression. The failure of conventional test procedures in this context was given dramatic demonstration by the Monte Carlo evidence reported in their article. More discussion and further evidence of the danger of spurious regressions can be found in their subsequent monograph, Granger and Newbold (1977, pp. 202-214) and in related work by Plosser and Schwert (1978).

A focal point of the Granger-Newbold study is the specification of regression equations in terms of the levels of economic time series. Granger and Newbold argue persuasively that: the levels of many economic time series are non stationary; their sample paths are well represented by integrated processes of the ARIMA type popularized by Box and Jenkins (1970); and often they appear to be near random walks. It is further argued that regression equations which relate such time series frequently have high R^2 yet also typically display highly autocorrelated residuals, indicated by very low Durbin-Watson statistics. In such situations they rightly contend that the usual significance tests about the regression coefficients are very misleading. The sampling experiments they conduct provide strong evidence that the conventional significance tests are seriously biased towards rejection of the null hypothesis of no relationship and hence acceptance of a spurious relationship, even when the series are generated as statistically independent random walks.

The failure of the conventional significance tests in the Granger-Newbold experiments has attracted a good deal of attention in econometric research and teaching programs. Yet, surprisingly, no analytical study has been made of what exactly goes wrong with the conventional tests in their Monte Carlo set up. Granger and Newbold themselves emphasize the inappropriateness of the usual tests, given the heavily autocorrelated residuals. They point to the difficulty of the distribution problem involved. But they provide no further analysis. Subsequent researchers appear to have ignored the problem.

The present paper develops an asymptotic theory for regressions that relate quite general integrated random processes. This includes spurious regressions of the Granger-Newbold type as a special case. It turns out that the correct asymptotic theory goes a long way towards explaining the experimental results that these authors obtained. In many cases their findings are quite predictable from the true asymptotic behavior of the relevant statistics. Thus, our theory demonstrates that in the Granger-Newbold regressions of independent random walks the usual t-ratio significance test does not possess a limiting distribution but actually diverges as the sample size $T \uparrow \infty$. Inevitably, therefore, the bias in this test towards the rejection of no relationship (based on a nominal critical value of 1.96) will increase with T. (In fact, $T = 50$ in the Granger-Newbold experiments.) We also show that the Durbin-Watson statistic actually converges in probability to zero, while the regression R^2 has a nondegenerate limiting distribution as $T \uparrow \infty$. These and other related results are given in Section 2 of the paper.

Section 3 extends the theory to multiple regressions in which the variables are generated by a very general vector integrated process. This framework allows for cointegrating regressions of the type recently advocated by Granger and Engle (1985). The latter authors work under a null hypothesis of no cointegration in the series. Under this null, the relevant asymptotics for such regressions are given by the general theory that we develop in this Section of the paper. Some discussion is also provided of the appropriate distribution theory under the alternative hypothesis that there is cointegration in the series in question (i.e. some linear combination of the series is stationary).

Some concluding remarks are made and some further extensions of the theory are discussed in Section 4 of the paper. A Mathematical Appendix is provided which contains proofs of results which appear in the body of the paper together with some related material on functional central limit theory and the multivariate Wiener process that is needed in our mathematical derivations.

2. LARGE SAMPLE $(T \uparrow \infty)$ ASYMPTOTICS FOR SPURIOUS REGRESSIONS

Granger and Newbold (1974) take the following stochastic environment as their prototype of a spurious regression. The variate y_t is regressed on a constant and another variate x_t giving the least squares regression:

$$(1) \qquad y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t \ ; \quad t = 1, \ldots, T .$$

In fact, y_t and x_t are generated by the independent random walks:

$$(2) \qquad y_t = y_{t-1} + v_t \ , \quad x_t = x_{t-1} + w_t \ ; \quad t = 1, 2, \ldots$$

in which v_t is $\mathrm{iid}(0, \sigma_v^2)$ and w_t is $\mathrm{iid}(0, \sigma_w^2)$. In their simulations Granger-Newbold set initial conditions as $y_0 = x_0 = 100$ and draw v_t, w_t from independent $N(0,1)$ populations.

For our own theoretical development we shall make much weaker assumptions about the innovations in (2). It will be convenient for our purpose here and for our analysis in the next Section of the paper to work at quite a general level. Thus, we introduce a sequence $\{\xi_t\}_1^\infty$ of random n-vectors defined on a probability space $(\Omega, \mathcal{B}, P)$. Let $S_t = \Sigma_{j=1}^t \xi_j$ be the partial sum process and set $S_0 = 0$. We require:

ASSUMPTION 2.1

(a) $E(\xi_t) = 0$, *all* t;

(b) $\sup_{i,t} E|\xi_{it}|^{\beta+\varepsilon} < \infty$ *for some* $\beta > 2$ *and* $\varepsilon > 0$;

(c) $\Sigma = \lim_{T\to\infty} T^{-1}E(S_T S_T')$ *exists and is positive definite;*

(d) $\{\xi_t\}_1^\infty$ *is strong mixing with mixing numbers* α_m *satisfying:*

$$\Sigma_1^\infty \alpha_m^{1-2/\beta} < \infty .$$

If we now set $n = 2$ and $\xi_t' = (v_t, w_t)$ then the conditions implied by Assumption 2.1 on the innovations of (2) are quite weak. In effect, they permit y_t and x_t to be rather general integrated processes (of order one) whose differences are weakly dependent and possibly heterogeneously distributed innovations. This includes a wide variety of possible data generating mechanisms, such as the ARIMA(p,1,q) model, under very general conditions on the underlying errors. Note that condition (b) of Assumption 2.1 controls the allowable heterogeneity of the process, whereas (d) controls the extent of permissable temporal dependence in the process in relation to the probability of outlier occurrences. Thus, the summability condition (d) is satisfied when the mixing decay rate is $\alpha_m = O(m^{-\lambda})$ for some $\lambda > \beta/(\beta-2)$. As β approaches 2 and the probability of outliers rises (under the weakening moment condition (b)) the mixing decay rate thereby increases and the effect of outliers is then required under (d) to wear off more quickly.

Note that if $\{\xi_t\}$ is weakly stationary then

$$\Sigma = E(\xi_1\xi_1') + \Sigma_{k=1}^{\infty}E(\xi_1\xi_k' + \xi_k\xi_1') \tag{3}$$

and the convergence of this series is implied by the mixing condition (d) (Ibragimov and Linnik (1971), theorem 18.5.3). Moreover, when $\xi_t' = (v_t, w_t)$ and v_t and w_t are independent, as in the spurious regressions context, we have

$$\Sigma = \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_w^2 \end{bmatrix}$$

where

$$\sigma_v^2 = \lim_{T\to\infty} T^{-1}E(P_T^2) \ , \quad \sigma_w^2 = \lim_{T\to\infty} T^{-1}E(Q_T^2)$$

and $\quad P_t = \Sigma_1^t v_j \ , \quad Q_t = \Sigma_1^t w_j \ .$

We denote the standard errors of $\hat{\alpha}$ and $\hat{\beta}$ in the regression (1) by $s_{\hat{\alpha}}$ and $s_{\hat{\beta}}$. The customary t-ratios are then $t_\alpha = \hat{\alpha}/s_{\hat{\alpha}}$ and $t_\beta = \hat{\beta}/s_{\hat{\beta}}$. Let DW be the usual Durbin-Watson d-statistic and R^2 be the coefficient of determination. The Box-Pierce statistic is $Q_k = T\Sigma_{s=1}^k r_s^2$ where $r_s = \Sigma_{s+1}^T \hat{u}_t\hat{u}_{t-s}/\Sigma_1^T\hat{u}_t^2$.

Theorem 2.3 below provides the correct $(T \uparrow \infty)$ asymptotic theory for the least squares regression estimates in (1), the associated t-ratios and the commonly used regression diagnostics DW, R^2 and Q_k . The following Lemma is useful in the derivation of this theorem and our other results.

LEMMA 2.2. Suppose $\{y_t\}_1^\infty$ *and* $\{x_t\}_1^\infty$ *are generated by* (2). *If the innovation sequences* $\{v_t\}_1^\infty$ *and* $\{w_t\}_1^\infty$ *are independent and if* $\{(v_t, w_t)\}_1^\infty$ *satisfies Assumption* 2.1 *then, as* $T \uparrow \infty$ *:*

(a) $T^{-3/2}\Sigma_1^T x_t \Rightarrow \sigma_w \int_0^1 W(t)\,dt$; $T^{-3/2}\Sigma_1^T y_t \Rightarrow \sigma_v \int_0^1 V(t)\,dt$;

(b) $T^{-2}\Sigma_1^T x_t^2 \Rightarrow \sigma_w^2 \int_0^1 W(t)^2 dt$; $T^{-2}\Sigma_1^T y_t^2 \Rightarrow \sigma_v^2 \int_0^1 V(t)^2 dt$;

(c) $T^{-2}\Sigma_1^T (x_t - \bar{x})^2 \Rightarrow \sigma_w^2\left[\int_0^1 W(t)^2 dt - \{\int_0^1 W(t)\,dt\}^2\right]$;

$T^{-2}\Sigma_1^T (y_t - \bar{y})^2 \Rightarrow \sigma_v^2\left[\int_0^1 V(t)^2 dt - \{\int_0^1 V(t)\,dt\}^2\right]$;

(d) $T^{-2}\Sigma_1^T y_t x_t \Rightarrow \sigma_v \sigma_w \int_0^1 V(t)W(t)\,dt$;

(e) $T^{-1}\Sigma_r^T y_t(y_t - y_{t-r}) \Rightarrow (r/2)\{\sigma_v^2 V(1)^2 + \Omega_{v0}\} + \Sigma_{j=1}^r (r-j)\Omega_{vj}$;

$T^{-1}\Sigma_r^T x_t(x_t - x_{t-r}) \Rightarrow (r/2)\{\sigma_w^2 W(1)^2 + \Omega_{w0}\} + \Sigma_{j=1}^r (r-j)\Omega_{wj}$;

(f) $T^{-1}\Sigma_r^T y_t(x_t - x_{t-r}) + T^{-1}\Sigma_r^T x_t(y_t - y_{t-r}) \Rightarrow r\sigma_v\sigma_w V(1)W(1)$

where $W(t)$ *and* $V(t)$ *are independent Wiener processes on* $C[0,1]$ *and where*

$$\Omega_{vj} = \lim_{T\to\infty} T^{-1}\Sigma_{j+1}^T E(v_t v_{t-j}) \; ; \quad j = 0, 1, \ldots$$

$$\Omega_{wj} = \lim_{T\to\infty} T^{-1}\Sigma_{j+1}^T E(w_t w_{t-j}) \; ; \quad j = 0, 1, \ldots .$$

Moreover (a)-(f) *hold irrespective of the initial conditions assigned to* y_0 *and* x_0 .

In the statement of Lemma 2.2 $C[0,1]$ denotes the space of all real valued continuous functions on the interval $[0,1]$. The Wiener processes

$W(t)$ and $V(t)$ that occur in the Lemma are stochastically independent. Their sample paths lie in $C[0,1]$. Results (a)-(f) of the Lemma establish that suitably standardized sample moments of the sequences $\{y_t\}_1^\infty$ and $\{x_t\}_1^\infty$ converge weakly to appropriately defined functionals of the Wiener processes $W(t)$ and $V(t)$. Each of these functionals has a well defined non degenerate distribution. The notation " $\Rightarrow$ " in the Lemma is used to denote weak convergence of the relevant probability measures. Thus, in the case of (a) we deduce that $T^{-3/2}\Sigma_1^T x_t$ converges in distribution to the distribution of the functional $\sigma_w \int_0^1 W(t)dt$ of the Wiener process $W(t)$ on $C[0,1]$. Since $W(t)$ is Gaussian with mean zero and independent increments (see, for example, Billingsley (1968)) we further deduce that the limiting distribution of $T^{-3/2}\Sigma_1^T x_t$ is normal with zero mean and variance given by

$$\begin{aligned}\sigma_w^2 E\left\{\int_0^1\int_0^1 W(t)W(s)\,dtds\right\} &= 2\sigma_w^2\int_0^1\int_0^r E\{W(r)W(s)\}dsdr\\ &= 2\sigma_w^2\int_0^1\int_0^r s\,dsdr\\ &= \sigma_w^2/3\ .\end{aligned}$$

Lemma 2.2 and the subsequent results of this paper are proved in the Mathematical Appendix using functional central limit theory. An introductory discussion to this form of asymptotic theory and references to the recent literature on the subject are given there.

THEOREM 2.3. Suppose (1) *is estimated by least squares regression and the conditions of Lemma* 2.2 *are satisfied. Then, as* $T \uparrow \infty$:

(a) $$\hat{\beta} \Rightarrow \frac{\sigma_v\left\{\int_0^1 V(t)W(t)dt - \int_0^1 V(t)dt\int_0^1 W(t)dt\right\}}{\sigma_w\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)dt)^2\right\}} = (\sigma_v/\sigma_w)\zeta \ ;$$

(b) $$T^{-1/2}\hat{\alpha} \Rightarrow \sigma_v\left\{\int_0^1 V(t)dt - \zeta\int_0^1 W(t)dt\right\} \ ;$$

(c) $T^{-1/2}t_\beta \Rightarrow \mu/\nu^{1/2}$, *where*

$$\mu = \int_0^1 V(t)W(t)dt - \int_0^1 V(t)dt\int_0^1 W(t)dt \ , \quad \textit{and}$$

$$\nu = \left\{\int_0^1 V(t)^2 dt - (\int_0^1 V(t)dt)^2\right\}\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)dt)^2\right\} - \left\{\int_0^1 V(t)W(t)dt - \int_0^1 V(t)dt\int\right.$$

(d) $$T^{-1/2}t_\alpha \Rightarrow \left\{\int_0^1 V(t)dt - \zeta\int_0^1 W(t)dt\right\}\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)dt)^2\right\}/[\nu\int_0^1 W(t)^2 d$$

(e) $$R^2 \Rightarrow \frac{\zeta^2\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)dt)^2\right\}}{\int_0^1 V(t)^2 dt - (\int_0^1 V(t)dt)^2} \ ;$$

(f) $DW \underset{p}{\rightarrow} 0$; *and*

$$TDW \Rightarrow \{(\Omega_{v0}/\sigma_v^2) + \zeta^2(\Omega_{w0}/\sigma_w^2)\}\left[\int_0^1 V(t)^2 dt - (\int_0^1 V(t)dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\right.\right.$$

(g) *For all fixed* $s \geq 1$

$$T(r_s - 1) \Rightarrow -A_s/B$$

and

$$r_s = 1 + O_p(T^{-1})$$

where

$$A_s = (s/2)\left[\{V(1) - \zeta W(1)\} - \left\{\int_0^1 V(t)dt - \zeta\int_0^1 W(t)dt\right\}\right]^2 + (s/2)\left\{\int_0^1 V(t)dt - \zeta\int_0^1 W(t\right.$$

$$+\left\{\Omega_{v0}/\sigma_v^2 + \Sigma_{j=1}^s (s-j)\Omega_{vj}/\sigma_v^2\right\} + \zeta^2\left\{\Omega_{w0}/\sigma_w^2 + \Sigma_{j=1}^2 (s-j)\Omega_{wj}/\sigma_w^2\right\}$$

and

$$B = \int_0^1 V(t)^2 dt - (\int_0^1 V(t)dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)dt)^2\right\}$$

(h) $T^{-1}Q_k = \Sigma_{s=1}^k r_s^2 \underset{p}{\to} k$;

where $W(t)$ *and* $V(t)$ *are independent Wiener processes on* $C[0,1]$.

Theorem 2.3 goes a long way towards explaining the Monte Carlo results reported by Granger and Newbold. In the first place, parts (c) and (d) of the Theorem show that the conventional t ratios, t_α and t_β , that are used to assess the significance of the coefficients in regression analysis do not have limiting distributions in this context. In fact, the distributions of t_α and t_β diverge as $T \uparrow \infty$, so that there are no asymptotically correct critical values for these conventional significance tests. We should expect the rejection rate when these tests are based on a critical value delivered from conventional asymptotics (such as 1.96) to continue to increase with the sample size. The high rejection rate that Granger and Newbold found in their experimental investigation (where $T = 50$) therefore comes as no surprise. Indeed, it is predicted by the correct asymptotic theory. From their experimental results, Granger and Newbold (1974, p. 115) suggest the use of a new critical value of 11.2 (rather than the usual value 1.96) when assessing the significance of the coefficient of x_t in the regression at the 5% level. Our results now show that this suggestion has no foundation in asymptotic theory. On the contrary, if asymptotics were to be used in this context then the correctly standardized statistic is $t_\beta' = t_\beta/\sqrt{T}$, whose limiting distribution is given by the functional defined in (c) rather than the standard $N(0,1)$ distribution that is used in conventional asymptotic tests. Note that after such standardization the critical value suggested by Granger and Newbold transforms as $11.2 \to 11.2/\sqrt{50} = 1.58$. This transformed value is, in fact, an approximation to the percentage point of the asymptotic distribution given in (c).

However, (a) and (b) of Theorem 2.3 show that, in contrast to the usual results of regression theory, the coefficients $\hat{\alpha}$ and $\hat{\beta}$ do not converge in probability to constants as $T \uparrow \infty$. In fact, $\hat{\beta}$ has a nondegenerate limiting distribution as $T \uparrow \infty$ and the distribution of $\hat{\alpha}$ actually diverge as $T \uparrow \infty$. Thus, the uncertainty about the regression (1) that stems from its spurious nature (y_t and x_t being generated by (2)) persists asymptotically in these limiting distributions. The contrast with usual regression theory extends further to the case where y_t and x_t are generated by independent stable autoregressive processes. In that case both $\hat{\alpha}$ and $\hat{\beta}$ converge in probability to zero.

The reason for the distinctive nature of the present results is that the processes y_t and x_t are nonergodic. In fact, the sample moments of y_t and x_t and their joint sample moments do not converge to constants, as they do for ergodic processes. As shown by the results of Lemma 2.2 quite different limiting behavior occurs. Upon appropriate standardization the sample moments of y_t and x_t actually converge weakly to random variables. Theorem 2.3 demonstrates that, when the regressors are nonergodic, we obtain limiting behavior for regression coefficients which is also quite different from that predicted by conventional theory.

Theorem 2.3 also shows that $DW \underset{p}{\to} 0$, whereas R^2 has a nondegenerate limiting distribution as $T \uparrow \infty$. Low values for the Durbin-Watson statistic DW and moderate values of the coefficient of determination R^2 are therefore to be expected in spurious regressions such as (1) with data generated by integrated processes such as (2). The asymptotic distribution of the standardized statistic TDW is also given in part (f) of the Theorem.

From part (g) of the Theorem we see that the serial correlation coefficients of the regression residuals converge in probability to unity. The limiting distribution of the standardized coefficient $T(r_s - 1)$ is also given in part (g) of the Theorem. From part (h) we deduce that the distribution of the commonly used Box-Pierce statistic Q_k diverges as $T \uparrow \infty$. All of these results differ from the conventional theory of regression with stationary processes.

3. EXTENSIONS TO MULTIPLE REGRESSIONS WITH INTEGRATED PROCESSES

The results of the previous section are readily extended to the multiple regressions of the form:

$$y_t = \hat{\alpha} + \hat{\beta}' x_t + \hat{u}_t \; ; \quad t = 1, \ldots, T \tag{5}$$

where y_t (a scalar) and x_t (an m-vector) are quite general integrated processes of order one. It is not necessary to require that y_t and x_t be independent. In fact, the main requirement is that the vector time series (y_t, x_t') is not cointegrated in the sense of Granger and Engle: that is, there does not exist a linear combination of (y_t, x_t') which is integrated of order zero (i.e. is a stationary process). When (y_t, x_t') is cointegrated, different results apply as we shall indicate below. The reader is referred to Phillips and Durlauf (1985) for the theory in this case.

For our development here we set $z_t' = (y_t, x_t')$ and suppose that z_t is a vector integrated process of dimension $n = m+1$ whose generating mechanism is:

(6) $$z_t = z_{t-1} + \xi_t \; ; \quad t = 1, 2, \ldots .$$

The process $\{\xi_t\}_1^\infty$ in (6) is required to satisfy Assumption 2.1 and we allow either of the commonly used initial conditions: (i) z_0 = const. with probability one; or (ii) z_0 is random with a certain specified distribution.

Our main result is the following:

THEOREM 3.1. If (5) *is estimated by least squares regression, if* z_t *is generated by* (6) *and if the innovation sequence* $\{\xi_t\}_1^\infty$ *satisfies Assumption* 2.1 *then as* $T \uparrow \infty$:

(a) $\hat{\beta} \Rightarrow A_{22}^{-1}a_{21}$;

(b) $T^{-1/2}\hat{\alpha} \Rightarrow b_1 - b_2'A_{22}^{-1}a_{21}$;

(c) $R^2 \Rightarrow a_{21}'A_{22}^{-1}a_{21}/a_{11}$;

(d) $T^{-1}F_\beta \Rightarrow (1/m)a_{21}'A_{22}^{-1}a_{21}/\{a_{11} - a_{21}'A_{22}^{-1}a_{21}\}$;

(e) $T^{-1/2}t_{\beta_i} \Rightarrow (A_{22}^{-1}a_{21})_i/\left\{(a_{11} - a_{21}'A_{22}^{-1}a_{21})^{1/2}([A_{22}^{-1}]_{ii})^{1/2}\right\}$;

(f) $TDW \Rightarrow \eta'\Sigma_\xi\eta/\{a_{11} - a_{21}'A_{22}^{-1}a_{21}\}$;

where

(7) $$A = \begin{array}{cc} & \begin{array}{cc} 1 & m \end{array} \\ & \begin{bmatrix} a_{11} & a_{21}' \\ a_{21} & A_{22} \end{bmatrix} \begin{array}{c} 1 \\ m \end{array} \end{array} = \Sigma^{1/2}\left\{\int_0^1 Z(t)Z(t)'dt - \int_0^1 Z(t)dt\int_0^1 Z(t)'dt\right\}\Sigma^{1/2} ,$$

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\begin{array}{c} 1 \\ m \end{array} = \Sigma^{1/2}\int_0^1 Z(t)\,dt ,$$

$$\eta' = (1, -a_{21}'A_{22}^{-1}) ,$$

$Z(t)$ *is a vector Wiener process on* C^n , Σ *is defined in Assumption* 2.1 *and*

$$\Sigma_\xi = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(\xi_t \xi_t') \ .$$

In part (d) of Theorem 3.1, F_β denotes the customary regression F statistic for testing the signficance of $\hat{\beta}$ in (5); and, in part (e), t_{β_i} represents the conventional t-statistic for assessing the significance of $\hat{\beta}_i$. We observe that the distributions of both F_β and t_{β_i} diverge as $T \uparrow \infty$ and so there are no asymptotically correct critical values for these statistics. As in the case of the t-ratio statistics considered in the previous section, the use of conventional asymptotics in setting the critical values of these tests leads to a rejection rate which increases with the sample size. We note that the divergence rate of the distribution of F_β is $O(T)$. This is greater than the divergence rate of $O(T^{1/2})$ for the individual t tests (and that of the t-ratio statistics t_α and t_β in the simple regression context). In a regression with many regressors, therefore, we might expect a noticeably greater rejection rate for the block F test than for the individual t tests or for a test with fewer regressors. This is, in fact, precisely what we do observe in the Granger and Newbold experiments. For regressions involving independent random walks their Table 2 (p. 116 of Granger and Newbold (1974)) reports a rejection rate of 76% when $m = 1$ (one regressor) and a rejection rate of 96% when $m = 5$ (with $T = 50$ in both cases). For regressions involving independent ARIMA(0,1,1) series the corresponding rejection rates are 64% and 90%. Thus, asymptotic theory is again helpful in explaining these simulation findings.

As in the case of simple regression, the regression coefficients $\hat{\alpha}$

and $\hat{\beta}$ do not converge in probability to constants as $T \uparrow \infty$. Once again $\hat{\beta}$ has a nondegenerate limiting distribution while the distribution of $\hat{\alpha}$ diverges as $T \uparrow \infty$. From part (f) of Theorem 3.1 we see that $DW \underset{p}{\to} 0$ and, from part (c), that R^2 has a nondegenerate limiting distribution, both as in the case of simple regression. Again, low values for the Durbin-Watson statistic and moderate values of R^2 are to be expected in regressions such as (5) that involve integrated processes.

It is worth emphasizing that in the present case y_t and x_t are, in general, correlated time series. In the previous section we retained the Granger-Newbold hypothesis of independent y_t and x_t to underscore the spurious nature of the regression. However, it is clear from the results of Theorem 3.1 that the major conclusions of the present theory continue to hold irrespective of whether y_t and x_t are independent or not. Of course, the correlation properties of these time series do have quantitative effects on the limiting distributions. These effects are introduced through the parameters of the limiting covariance matrices Σ and Σ_ξ. Under the conditions of the Theorem, however, these effects do not interfere with the main qualitative results of the theory: viz. that the regression coefficients $\hat{\alpha}$ and $\hat{\beta}$ do not converge in probability to constants; that the distributions of test statistics such as F_β and t_{β_i} diverge as $T \uparrow \infty$; and that the Durbin-Watson statistic converges in probability to zero whereas R^2 as a nondegenerate limiting distribution as $T \uparrow \infty$.

There is, in fact, one case of major importance where the correlation properties of y_t and x_t do interfere with these qualitative results. Assumption 2.1(c) requires that the limiting covariance matrix Σ be non singular. If we allow the matrix Σ to be singular then the asymptotic theory of Theorem 3.1 no longer holds as stated. We may suppose, for example, that Σ has rank $m = n-1$ and the submatrix Σ_{22} has full rank m in the partition:

$$\begin{array}{cc} & \begin{array}{cc} 1 & m \end{array} \\ \Sigma = & \left[\begin{array}{cc} \sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22} \end{array}\right] \begin{array}{c} 1 \\ m \end{array} \end{array} .$$

Then $\sigma_{11} = \sigma_{21}'\Sigma_{22}^{-1}\sigma_{21}$ and $\Sigma\gamma = 0$ where $\gamma' = (1, -a')$ and $a' = \sigma_{21}'\Sigma_{22}^{-1}$. The singularity of Σ is, it turns out, a necessary condition for (y_t, x_t) to be cointegrated in the sense of Granger and Engle (1985). In this case, the cointegrating vector is γ and under weak additional conditions (it is sufficient that $\gamma' z_t$ satisfy Assumption 2.1) we find that the regression coefficient $\hat{\beta}$ in (5) is a consistent estimator of $a = \Sigma_{22}^{-1}\sigma_{21}$. Thus, the asymptotic theory of regression for cointegrated series is quite different in certain respects from that given in Theorem 3.1. The reader is referred to Phillips and Durlauf (1985) for a detailed investigation of the regression theory in this case.

In developing tests of cointegration Granger and Engle (1985) prescribe as their null hypothesis that the series in question are not cointegrated. The asymptotic distribution theory under this null of the cointegrating regression (5) is then given directly by Theorem 3.1. To test this null hypothesis against the alternative that the series are cointegrated Granger and Engle (1985) suggest a number of different statistics. One of these is the Durbin-Watson statistic, DW, constructed from the residuals of the cointegrating regression. As we have seen above, under the null of no cointegration, $DW \xrightarrow{p} 0$ as $T \uparrow \infty$. Moreover, the limiting distribution of TDW given by part (f) of Theorem 3.1 may be used to construct an asymptotic critical value for the Granger-Engle test, thereby approximating the critical values reported in Tables II and III of Granger and Engle (1985) for this test.

This DW test was earlier suggested by Bhargava (1984). By virtue of its construction, it would seem to be an intuitively appealing test for discriminating between stationary and nonstationary alternatives or, in the present context of regression residuals, between cointegration and no cointegration. However, as Granger and Engle (1985) remark, the correct critical value for this test is parameter dependent upon the dynamics of the errors. Our result makes this dependency explicit and shows how the limiting distribution of the statistic TDW, and hence the implied asymptotic critical value for DW, depends on the serial correlation and heterogeneity characteristics of the innovation sequence $\{\xi_t\}_1^\infty$.

4. SOME CONCLUSIONS AND FURTHER EXTENSIONS

To the extent that the levels of economic time series are nonstationary and nonergodic, regressions that relate such variables will typically require the use of asymptotic methods and results that are quite different from those that are well established in current econometric theory and practice. The present paper has developed an asymptotic theory of regression which is applicable when the variables are quite general integrated processes. This includes the spurious regressions of the type considered in the simulation studies of Granger and Newbold (1974, 1977). When the correct asymptotic theory is brought into play in this context we have found that the simulation findings of Granger and Newbold offer no surprises. In many respects these findings are well predicted by the relevant theory.

The asymptotic theory of regression that we have developed here also applies to the cointegrating regressions that have recently been introduced by Granger and Engle (1985). Interestingly, the essential characteristics of the regression theory in this case are the same as they are for spurious regressions which relate independent time series. Some major differences in the theory do arise when the time series in question are cointegrated. However, methods similar to those that we have employed here may be used to analyze such regressions. For a detailed treatment of this case the reader is referred to recent work by Phillips and Durlauf (1985) and some related work by Stock (1985).

The theory derived in this paper is based on large sample $(T \uparrow \infty)$ asymptotics. An alternative asymptotic theory may be developed which works in terms of the time interval (h) between sampled observations. As $h \downarrow 0$ we obtain a continuous record of observations over a finite time span. Such continuous recording of data has been a feature of statistical data collection in certain physical and medical sciences for many years. Trends in this direction for economic and financial statistics are now well established. For example, with the electronic monitoring of activity in certain financial and foreign exchange markets it is now possible to work with data recorded at very high frequencies (daily, hourly or even minute by minute in some cases). It is, therefore, of natural interest to study the asymptotic behavior of statistical procedures when $h \downarrow 0$ as well as when $T \uparrow \infty$ for a given fixed h. Continuous record asymptotics of this type were first developed rigorously by the author in a recent paper (1985a). The regressions considered in the present paper may also be analyzed using continuous record asymptotics. When the innovation sequence $\{\xi_t\}$ is iid some especially interesting and intuitively appealing results are obtained by this approach. For example, we find that the regression coefficient $\hat{\beta}$ in (1) as the same limiting distribution in this case as $h \downarrow 0$ over a fixed span of data as it has when $T \uparrow \infty$ (with h fixed) over an infinite span. Thus, the same limiting distribution theory applies in two different directions. One might therefore expect the asymptotic distribution of this regression coefficient to yield an unusually good approximation in finite samples. The reader is referred to an earlier version of the present paper (Phillips (1985b)) for a detailed analysis along these lines.

It is also possible to gain insight into the adequacy of the asymptotic theory presented here by means of higher order asymptotics. Since the limiting distribution theory in the present paper is nonnormal and relies on functional central limit theory the mathematical development of such expansions is quite difficult and is of a very different character from the conventional theory of Edgeworth expansions. In another paper the author (1985c) has developed the theory of such higher order asymptotics in a general setting which extends to the present case. It may be shown from these results, for example, that the error on the asymptotic distribution of $\hat{\beta}$ given in Theorem 2.3 is of $O_p(T^{-1})$ under quite general conditions. This gives us another reason for expecting the asymptotic theory of the present paper to work well, at least in certain cases.

MATHEMATICAL APPENDIX

1. *Functional Limit Theory and the Wiener Process*

Define the partial sums $P_t = \Sigma_1^t v_j$, $Q_t = \Sigma_1^t w_j$ and set $P_0 = Q_0 = 0$. In view of (2) we may write $y_t = P_t + y_0$ and $x_t = Q_t + x_0$. In what follows we shall permit either of the commonly proposed initial conditions: (i) y_0 (respectively x_0) = c , a constant, with probability one; or (ii) y_0 (x_0) has a certain specified distribution. We construct the standardized sums:

$$Y_T(t) = \frac{1}{\sqrt{T}\,\sigma_v} P_{[Tt]} = \frac{1}{\sqrt{T}\,\sigma_v} P_{j-1} \;; \quad (j-1)/T \le t < j/T \quad (j = 1, \ldots, T)$$

$$X_T(t) = \frac{1}{\sqrt{T}\,\sigma_w} Q_{[Tt]} = \frac{1}{\sqrt{T}\,\sigma_w} Q_{j-1} \;; \quad (j-1)/T \le t < j/T \quad (j = 1, \ldots, T)$$

$$Y_T(1) = \frac{1}{\sqrt{T}\,\sigma_v} P_T \;, \quad X_T(1) = \frac{1}{\sqrt{T}\,\sigma_w} Q_T$$

where [a] denotes the integer part of a .

We shall also have occasion to work with the more general sequence $\{\xi_t\}_1^\infty$ of random n-vectors satisfying Assumption 2.1. Using the partial sum process $S_t = \Sigma_1^t \xi_j$ $(S_0 = 0)$ we construct:

$$Z_T(t) = T^{-1/2}\Sigma^{-1/2}S_{[Tt]} = T^{-1/2}\Sigma^{-1/2}S_{j-1} \;; \quad (j-1)/T \le t < j/T \quad (j = 1, \ldots, T)$$

$$Z_T(1) = T^{-1/2}\Sigma^{-1/2}S_T \;.$$

When $\xi_t' = (v_t, w_t)$ and the sequences $\{v_t\}$ and $\{w_t\}$ are independent we have:

$$\Sigma = \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_w^2 \end{bmatrix}, \quad Z_T(t) = \begin{bmatrix} Y_T(t) \\ X_T(t) \end{bmatrix}.$$

$X_T(t)$ and $Y_T(t)$ are random elements of the function space $D[0,1]$, the space of all real valued functions on $[0,1]$ that are right continuous at each point of $[0,1]$ and have finite left limits. Similarly, $Z_T(t)$ is an element of the product space $D^n = D[0,1] \times \ldots \times D[0,1]$ (n copies). We endow D^n with the metric

$$d_n(f,g) = \max_i\{d_0(f_i, g_i) : i = 1, \ldots, n;\ f_i, g_i \in D[0,1]\}$$

where d_0 is the modified Skorohod metric [Billingsley (1968), p. 112]. With this metric D^n is a separable and complete metric space. Under quite general conditions on the underlying process $\{\xi_t\}$ we may establish a central limit theory for $Z_T(t)$ on the function space D^n. We shall, in particular, make use of the following result which is proved in Phillips (1985c, theorem 2.2):

LEMMA A.1. Let $\{\xi_t\}_1^\infty$ *be a sequence of random* n*-vectors satisfying Assumption* 2.1. *Then, as* $T \uparrow \infty$, $Z_T(t) \Rightarrow Z(t)$ *a multivariate Wiener process on* $C^n = C[0,1] \times \ldots \times C[0,1]$.

The notation " $\Rightarrow$ " in the statement of Lemma A.1 is used to signify the weak convergence of the probability measure of $Z_T(t)$ to the probability measure (here, multivariate Wiener measure) of the random function $Z(t)$. The result is a multivariate functional central limit theorem (CLT) i.e. a CLT on the function space D^n. It may also be described as a multivariate invariance principle following early (univariate) work by Donsker (1951) and Erdos and Kac (1946). Univariate results similar to Lemma A.1 were obtained by McLeish (1975a) and Herrndorf (1984). The reader is referred to Billingsley (1968), Hall and Heyde (1980) and Pollard (1984) for an introduction to the subject and excellent reviews of the literature.

The limit process $Z(t)$ in Lemma A.1 is popularly known as the vector Wiener process or as vector Brownian motion. The sample paths of $Z(t)$ lie almost surely (Wiener measure) in $C^n = C[0,1] \times \ldots \times C[0,1]$ (n copies) where $C[0,1]$ is the space of all real valued continuous functions on $[0,1]$. Moreover, the vector random function $Z(t)$ is Gaussian, with independent increments (so that $Z(s)$ is independent of $Z(t) - Z(s)$ for $0 < s < t \leq 1$) and with independent elements (so that $Z_i(t)$ is independent of $Z_j(t)$, $i \neq j$).

In the case where $n = 2$, $\xi_t' = (v_t, w_t)$ and the sequences $\{v_t\}$ and $\{w_t\}$ are independent we obtain

$$X_T(t) \Rightarrow W(t) \ , \quad Y_T(t) \Rightarrow V(t) \ ; \quad \text{as} \quad T \uparrow \infty$$

where $W(t)$ and $V(t)$ are independent Wiener processes on $C[0,1]$.

2. *Proof of Lemma 2.2*

To prove (a)-(d) we write each statistic as a functional of $X_T(t)$ or $Y_T(t)$ or both as is appropriate. Thus, in the case of (a) we have:

$$\begin{aligned}
T^{-3/2}\Sigma_1^T x_t &= T^{-3/2}\Sigma_{i=1}^T(Q_{i-1} + w_i + x_0) \\
&= \sigma_w T^{-1}\Sigma_1^T(1/\sqrt{T}\,\sigma_w)Q_{i-1} + T^{-3/2}\Sigma_1^T(w_i + x_0) \\
&= \sigma_w \Sigma_1^T \int_{(i-1)/T}^{i/T} X_T(t)\,dt + o_p(1) \\
&= \sigma_w \int_0^1 X_T(t)\,dt + o_p(1) \\
&\Rightarrow \sigma_w \int_0^1 W(t)\,dt \ ; \quad \text{as} \quad T \uparrow \infty
\end{aligned}$$

The third line of the argument follows since $T^{-1/2}x_0 = o_p(1)$ (under either of the initial conditions (i) or (ii)) and $T^{-1}\Sigma_1^T w_i \to 0$ a.s. as $T \uparrow \infty$ by the strong law of McLeish for weakly dependent (α-mixing) sequences [McLeish (1975b), theorem 2.10 or White (1984), Corollary 3.48]. The final line of the derivation follows from Lemma A.1 and the continuous mapping theorem (e.g. Billingsley (1968), pp. 30-31). Note that the result holds irrespective of whether the initial condition x_0 is prescribed to be a fixed constant or a random variate with a specified distribution. Since the same point applies in all the derivations that follow (at least for large T asymptotics) we will set $x_0 = y_0 = 0$ without loss of generality to simplif derivations.

Arguments entirely analogous to those of the proof of (a) yield results (b)-(d). For example, in the case of (d) we have:

$$
\begin{aligned}
T^{-2}\Sigma_1^T y_t x_t &= T^{-1}\sigma_v\sigma_w\Sigma_1^T(1/\sqrt{T}\,\sigma_v)P_{i-1}(1/\sqrt{T}\,\sigma_w)Q_{i-1} \\
&\quad + T^{-2}\Sigma_1^T(v_t x_{t-1} + y_{t-1}w_t + v_t w_t) \\
\text{(A1)} \qquad &= \sigma_v\sigma_w\Sigma_1^T \int_{(i-1)/T}^{i/T} Y_T(t)X_T(t)\,dt + o_p(1) \\
&= \sigma_v\sigma_w\int_0^1 Y_T(t)X_T(t)\,dt + o_p(1) \\
&\Rightarrow \sigma_v\sigma_w\int_0^1 V(t)W(t)\,dt \;; \quad \text{as} \quad T \uparrow \infty .
\end{aligned}
$$

The final line of the argument follows once again from Lemma A.1 and the continuous mapping theorem. The second line of the derivation is a consequence of the fact that

$$T^{-1}\Sigma_1^T v_t x_{t-1} = \sigma_v\sigma_w \Sigma_1^T \int_{(i-1)/T}^{i/T} X_T(t)\,dY_T(t)$$

$$= \sigma_v\sigma_w \int_0^1 X_T(t)\,dY_T(t)$$

$$\text{(A2)} \qquad \Rightarrow \sigma_v\sigma_w \int_0^1 W(t)\,dV(t) \quad ; \quad \text{as} \quad T \uparrow \infty$$

as in Lemma 3.1(e) of Phillips and Durlauf (1985). The integral in (A2) is, of course, a stochastic integral. Similarly, we have

$$T^{-1}\Sigma_1^T y_{t-1} w_t \Rightarrow \sigma_v\sigma_w \int_0^1 V(t)\,dW(t) \quad ; \quad \text{as} \quad T \uparrow \infty$$

and by the strong law of McLeish for weakly dependent (α-mixing) sequences

$$T^{-1}\Sigma_1^T v_t w_t \to 0 \quad \text{a.s.}$$

This verifies the $o_p(1)$ error on the right side of (A1). Thus, (d) is established.

To prove (e) we first consider the case $r = 1$. We shall demonstrate the result for y_t, writing

$$T^{-1}\Sigma_1^T y_t(y_t - y_{t-1}) = T^{-1}\Sigma_1^T y_t v_t$$

$$\text{(A3)} \qquad = T^{-1}\Sigma_1^T v_t^2 + T^{-1}\Sigma_1^T y_{t-1} v_t \ .$$

We note that the process $\{v_t^2 - E(v_t^2)\}_1^\infty$ is a measurable function of (a finite stretch of) $\{\xi_t\}_1^\infty$ and is, therefore, strong mixing with mixing numbers that satisfy the summability condition (d) of Assumption 2.1 (e.g. White (1984), theorem 3.49). Set $\beta = 2r$ with $r > 1$ in this summability condition and note that in view of Assumption 2.1(a) $\sup_t E|V_t|^{2r+\varepsilon} < \infty$ for some $\varepsilon > 0$. Thus, by the strong law of McLeish for α-mixing sequences we have

$$\text{(A4)} \qquad T^{-1}\Sigma_1^T v_t^2 \xrightarrow[\text{a.s.}]{} \lim_{T\to\infty} T^{-1}\Sigma_1^T E(v_t^2) = \Omega_{v0}$$

as $T \uparrow \infty$. Taking the second term of (A3) we write:

$$\begin{aligned} T^{-1}\Sigma_1^T y_{t-1}v_t &= \tfrac{1}{2}\left\{T^{-1}\Sigma_1^T(y_{t-1} + v_t)^2 - T^{-1}\Sigma_1^T y_{t-1}^2 - T^{-1}\Sigma_1^T v_t^2\right\} \\ &= (\sigma_v^2/2)\Sigma_{i=1}^T [Y_T(t)^2]_{(i-1)/T}^{i/T} - (1/2T)\Sigma_1^T v_t^2 \\ &= (\sigma_v^2/2)Y_T(1)^2 - (1/2T)\Sigma_1^T v_t^2 \\ \text{(A5)} \qquad &\Rightarrow (\sigma_v^2/2)V(1)^2 - (1/2)\Omega_{v0} \end{aligned}$$

as $T \uparrow \infty$. (A5) follows by Lemma A.1 and the continuous mapping theorem (applied to the first term) and the strong law of large numbers (A4). Combining (A3), (A4) and (A5) we deduce that

$$\text{(A6)} \qquad T^{-1}\Sigma_1^T y_t(y_t - y_{t-1}) \Rightarrow (\sigma_v^2/2)V(1)^2 + (1/2)\Omega_{v0} \ ; \quad \text{as} \quad T \uparrow \infty$$

which is the stated result with $r = 1$.

Suppose (e) holds for some $r \geq 1$. Write

$$\text{(A7)} \qquad T^{-1}\Sigma_{r+1}^T y_t(y_t - y_{t-r-1}) = T^{-1}\Sigma_{r+1}^T y_t(y_t - y_{t-r}) + T^{-1}\Sigma_{r+1}^T y_t v_{t-r}$$

and

$$T^{-1}\Sigma_{r+1}^{T}y_t v_{t-r} = T^{-1}\Sigma_{r+1}^{T}y_{t-(r+1)}v_{t-r} + T^{-1}\Sigma_{r+1}^{T}\Sigma_{j=t-r}^{t}v_j v_{t-r}$$

$$= T^{-1}\Sigma_{r+1}^{T}y_{t-(r+1)}v_{t-r} + \Sigma_{i=0}^{r}T^{-1}\Sigma_{r+1}^{T}v_{t-i}v_{t-r}$$

$$\text{(A8)} \qquad \Rightarrow (\sigma_v^2/2)V(1)^2 - (1/2)\Omega_{v0} + \Sigma_{j=0}^{r}\Omega_{vj} \; ; \quad \text{as} \quad T \uparrow \infty \; .$$

The first term of (A8) follows in the same way as (A5) above. The second term follows from the strong law applied to $T^{-1}\Sigma_{r+1}^{T}\{v_{t-i}v_{t-r} - E(v_{t-i}v_{t-r})\}$ for $i = 0, 1, \ldots, r$. In particular, it is easily verified that the sequence $\{(v_{t-i}v_{t-r} - E(v_{t-i}v_{t-r}))\}_{r+1}^{\infty}$ satisfies the moment condition

$$\sup_t E|v_{t-i}v_{t-r} - E(v_{t-i}v_{t-r})|^{r+\delta} < \infty$$

in view of Assumption 2.1(b) with $\beta = 2r$, $r > 1$ and $\varepsilon = 2\delta > 0$; and $(v_{t-i}v_{t-r} - E(v_{t-i}v_{t-r}))$ is a measurable function of a finite stretch of the process ξ_t, so that it is also strong mixing with mixing numbers that satisfy the summability condition (d) of Assumption 2.1. It follows by the strong law of McLeish that

$$T^{-1}\Sigma_{r+1}^{T}(v_{t-i}v_{t-r} - E(v_{t-i}v_{t-r})) \xrightarrow[\text{a.s.}]{} 0 \; ; \quad \text{as} \quad T \uparrow \infty$$

or

$$T^{-1}\Sigma_{r+1}^{T}v_{t-i}v_{t-r} \xrightarrow[\text{a.s.}]{} \lim_{T\to\infty} T^{-1}\Sigma_{j+1}^{T}E(v_s v_{s-j}) = \Omega_{vj} \; , \quad j = r-i \; .$$

This proves (A8). Combining (A6), (A7) and (A8) we obtain:

$$T^{-1}\Sigma_{r+1}^{T}y_t(y_t - y_{t-r-1}) \Rightarrow ((r+1)/2)(\sigma_v^2V(1)^2 + \Omega_{v0}) + \Sigma_{j=1}^{r+1}(r+1-j)\Omega_{vj}$$

as $T \uparrow \infty$, proving (e) by induction on r.

To prove (f) we first consider the case $r = 1$. Write

$$\Sigma_{j=1}^{T} v_j \Sigma_{k=1}^{T} w_k = \Sigma_1^T v_j w_j + \Sigma_{j=1}^{T} (\Sigma_{k=1}^{j-1} w_k) v_j + \Sigma_{k=1}^{T} (\Sigma_{j=1}^{k-1} v_j) w_k$$

from which we deduce that

$$T^{-1}\Sigma_1^T y_{t-1}(x_t - x_{t-1}) + T^{-1}\Sigma_1^T x_{t-1}(y_t - y_{t-1})$$

$$= \sigma_v \sigma_w Y_T(1) X_T(1) - T^{-1}\Sigma_1^T v_t w_t$$

$$\Rightarrow \sigma_v \sigma_w V(1) W(1) \quad ; \quad \text{as} \quad T \uparrow \infty$$

by Lemma A.1 and the continuous mapping theorem (applied to the first term) and by the strong law applied to $T^{-1}\Sigma_1^T v_t w_t$ since $E(v_t w_t) = 0$ all t.

Suppose (f) holds for some $r \geq 1$. We write

$$T^{-1}\Sigma_{r+1}^T y_t(x_t - x_{t-r-1}) + T^{-1}\Sigma_{r+1}^T x_t(y_t - y_{t-r-1})$$

$$= T^{-1}\Sigma_{r+1}^T y_t(x_t - x_{t-r} + w_{t-r}) + T^{-1}\Sigma_{r+1}^T x_t(y_t - y_{t-r} + v_{t-r}) \ .$$

Now consider

$$T^{-1}\Sigma_{r+1}^T y_t w_{t-r} + T^{-1}\Sigma_{r+1}^T x_t v_{t-r}$$

$$= T^{-1}\Sigma_{r+1}^T y_{t-r-1} w_{t-r} + T^{-1}\Sigma_{r+1}^T x_{t-r-1} v_{t-r} + \Sigma_{i=0}^{r} \{T^{-1}\Sigma_{r+1}^T v_{t-i} w_{t-r} + T^{-1}\Sigma_{r+1}^T w_{t-i} v_{t-r}\}$$

$$= T^{-1}\Sigma_{r+1}^T (y_{t-r-1} w_{t-r} + x_{t-r-1} v_{t-r}) + o_p(1) \ ,$$

since v_t and w_s are independent. This summation converges weakly to $\sigma_v \sigma_w V(1) W(1)$ as in the case $r = 1$ considered above. (f) now follows by induction on r.

3. *Proof of Theorem 2.3*

$$\hat{\beta} = \frac{\Sigma y_t(x_t - \bar{x})}{\Sigma(x_t - \bar{x})^2} = \frac{T^{-2}\Sigma y_t x_t - T^{-1}\bar{y}\,\bar{x}}{T^{-2}\Sigma(x - \bar{x})^2}$$

$$\Rightarrow \frac{\sigma_V\sigma_W\left\{\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt\int_0^1 W(t)\,dt\right\}}{\sigma_W^2\left[\int_0^1 W(t)^2 dt - \{\int_0^1 W(t)\,dt\}^2\right]} \quad ; \quad \text{as} \quad T \uparrow \infty$$

by Lemma 2.2 and the continuous mapping theorem. We define

$$\zeta = \frac{\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt\int_0^1 W(t)\,dt}{\int_0^1 W(t)^2 dt - \{\int_0^1 W(t)\,dt\}^2}$$

as in the statement of the theorem, so that $\hat{\beta} \Rightarrow (\sigma_V/\sigma_W)\zeta$, proving (a).

Again by direct application of Lemma 2.2 we have:

$$T^{-1/2}\hat{\alpha} = T^{-3/2}\Sigma y_t - \hat{\beta}T^{-3/2}\Sigma x_t$$

$$\Rightarrow \sigma_V\left\{\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\right\} \quad ; \quad \text{as} \quad T \uparrow \infty$$

proving (b). Define $s^2 = T^{-1}\Sigma_1^T(y_t - \hat{\alpha} - \hat{\beta}x_t)^2$ and then:

$$T^{-1}s^2 = T^{-2}\Sigma\{(y_t - \bar{y}) - \hat{\beta}(x_t - \bar{x})\}^2$$

$$= T^{-2}\Sigma(y_t - \bar{y})^2 - \hat{\beta}^2 T^{-2}\Sigma(x_t - \bar{x})^2$$

$$\text{(A9)} \qquad \Rightarrow \sigma_V^2\left[\int_0^1 V(t)^2 dt - (\int_0^1 V(t)\,dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}\right] \quad ; \quad \text{as} \quad T \uparrow \infty$$

once again by Lemma 2.2. Now

$$T^{-\frac{1}{2}}t_{\beta} = \frac{\hat{\beta}}{T^{\frac{1}{2}}s_{\hat{\beta}}} = \frac{\hat{\beta}}{T^{\frac{1}{2}}s(\Sigma(x_t - \overline{x})^2)^{-\frac{1}{2}}}$$

$$= \frac{\hat{\beta}T(T^{-2}\Sigma(x_t - \overline{x})^2)^{\frac{1}{2}}}{T(T^{-\frac{1}{2}}s)}$$

$$\Rightarrow \zeta\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}^{\frac{1}{2}}$$

$$\div \left[\int_0^1 V(t)^2 dt - (\int_0^1 V(t)\,dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}\right]^{\frac{1}{2}} \quad ; \quad \text{as} \quad T \uparrow \infty$$

$$= \int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt \int_0^1 W(t)\,dt$$

$$\div \left[\left\{\int_0^1 V(t)^2 dt - (\int_0^1 V(t)\,dt)^2\right\}\left\{\int_0^1 W(t)^2 - (\int_0^1 W(t)\,dt)^2\right\} - \left\{\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt \int_0^1 W(t)\,dt\right\}^2\right]^{\frac{1}{2}}$$

$$= \mu/\nu^{\frac{1}{2}}$$

proving (c). Next

$$T^{-1/2}t_{\alpha} = \frac{\hat{\alpha}}{T^{1/2}s_{\hat{\alpha}}} = \frac{\hat{\alpha}(T\Sigma(x_t - \bar{x})^2)^{1/2}}{T^{1/2}s(\Sigma x_t^2)^{1/2}}$$

$$= \frac{(\hat{\alpha}/\sqrt{T})(T^{-2}\Sigma(x_t - \bar{x})^2)^{1/2}}{(s/\sqrt{T})(T^{-2}\Sigma x_t^2)^{1/2}}$$

$$\Rightarrow \left\{\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\right\}\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}^{1/2}$$

$$\div \left[\int_0^1 V(t)^2 dt - (\int_0^1 V(t)\,dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}\right]^{1/2} (\int_0^1 W(t)^2 dt)^{1/2},$$

as $T \uparrow \infty$,

$$= \left\{\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\right\}\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}$$

$$\div\left[\left[\left\{\int_0^1 V(t)^2 dt - (\int_0^1 V(t)\,dt)^2\right\}\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}\right.\right.$$

$$\left.\left. - \left\{\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt\int_0^1 W(t)\,dt\right\}^2\right]^{1/2} (\int_0^1 W(t)^2 dt)^{1/2}\right]$$

as required for (d).

The coefficient of determination is

$$R^2 = \frac{\Sigma(\hat{y}_t - \bar{y})^2}{\Sigma(y_t - \bar{y})^2} = \frac{\hat{\beta}^2 T^{-2}\Sigma(x_t - \bar{x})^2}{T^{-2}\Sigma(y_t - \bar{y})^2}$$

$$\Rightarrow \frac{\zeta^2\left\{\int_0^1 W(t)^2 dt - (\int_0^1 W(t)\,dt)^2\right\}}{\int_0^1 V(t)^2 dt - (\int_0^1 V(t)\,dt)^2}\,; \quad \text{as } T \uparrow \infty,$$

proving (e).

Next we consider the Durbin-Watson statistic:

$$DW = \frac{\Sigma_2^T(\hat{u}_t - \hat{u}_{t-1})^2}{\Sigma_1^T\hat{u}_t^2} = \frac{1}{T}\,\frac{T^{-1}\Sigma_2^T(v_t - \hat{\beta}w_t)^2}{T^{-2}\Sigma_1^T(y_t - \bar{y} - \hat{\beta}(x_t - \bar{x}))^2} .$$

Now

$$T^{-1}\Sigma_2^T(v_t - \hat{\beta}w_t)^2 \Rightarrow \Omega_{v0} + \sigma_v^2\zeta^2\Omega_{w0}/\sigma_w^2 \ ; \quad \text{as} \quad T \uparrow \infty ,$$

whereas

$$T^{-2}\Sigma_1^T(y_t - \bar{y} - \hat{\beta}(x_t - \bar{x}))^2 \Rightarrow \sigma_v^2\left[\int_0^1 V(t)^2dt - (\int_0^1 V(t)dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2dt - (\int_0^1 W(t)dt\right\}\right]$$

as $T \uparrow \infty$. Thus, $DW \xrightarrow[p]{} 0$ proving the first part of (f). The asymptotic distribution of the standardized statistic, TDW, is given by:

$$TDW \Rightarrow \{(\Omega_{v0}/\sigma_v^2) + \zeta^2(\Omega_{w0}/\sigma_w^2)\}/\left[\int_0^1 V(t)^2dt - (\int_0^1 V(t)dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2dt - (\int_0^1 W(t)dt)^2\right\}\right]$$

as required for the second part of (f).

To prove (g) we first write:

$$T(r_s - 1) = -\Sigma_{s+1}^T\hat{u}_t(\hat{u}_t - \hat{u}_{t-s})/\Sigma_1^T\hat{u}_t^2 - \{T^{-1}\Sigma_1^s\hat{u}_t^2\}/\{T^{-2}\Sigma_1^T\hat{u}_t^2\} .$$

Then, in view of (A9) we have:

$$\text{(A10)} \quad T^{-2}\Sigma_1^T\hat{u}_t^2 \Rightarrow \sigma_v^2\left[\int_0^1 V(t)^2dt - (\int_0^1 V(t)dt)^2 - \zeta^2\left\{\int_0^1 W(t)^2dt - (\int_0^1 W(t)dt)^2\right\}\right]$$

as $T \uparrow \infty$; and since s is fixed

(A11) $T^{-1}\Sigma_1^s \hat{u}_t^2 = T^{-1}\Sigma_1^s\{y_t - \bar{y} - \hat{\beta}(x_t - \bar{x})\}^2$

$$\Rightarrow s\sigma_v^2\left\{\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\right\}^2$$

as $T \uparrow \infty$. Finally,

$$T^{-1}\Sigma_{s+1}^T \hat{u}_t(\hat{u}_t - \hat{u}_{t-s}) = T^{-1}\Sigma_{s+1}^T\{y_t - \bar{y} - \hat{\beta}(x_t - \bar{x})\}\{y_t - y_{t-s} - \hat{\beta}(x_t - x_{t-s})\}$$

and by Lemma 2.2 we have as $T \uparrow \infty$:

(A12) $$T^{-1}\Sigma_s^T y_t(y_t - y_{t-s}) \Rightarrow (s/2)\{\sigma_v^2 V(1)^2 + \Omega_{v0}\} + \Sigma_{j=1}^s (s-j)\Omega_{vj}$$

(A13) $$T^{-1}\Sigma_s^T x_t(x_t - x_{t-s}) \Rightarrow (s/2)\{\sigma_w^2 W(1)^2 + \Omega_{w0}\} + \Sigma_{j=1}^s (s-j)\Omega_{wj}$$

and

(A14) $$T^{-1}\Sigma_s^T x_t(y_t - y_{t-s}) + T^{-1}\Sigma_s^T y_t(x_t - x_{t-s}) \Rightarrow s\sigma_v\sigma_w V(1)W(1)$$

as $T \uparrow \infty$. Moreover, $y_t - y_{t-s} = \Sigma_{t-s+1}^t v_j$ and $x_t - x_{t-s} = \Sigma_{t-s+1}^t w_j$ depend only on a finite stretch of the process $\{\xi_t' = (v_t, w_t)\}$ and thereby satisfy the moment and mixing conditions of Assumption 2.1. It follows that as $T \uparrow \infty$:

(A15) $$T^{-1/2}\Sigma_s^T(y_t - y_{t-s}) \Rightarrow s\sigma_v V(1) \equiv N(0,\ s^2\sigma_v^2) \ ,$$

(A16) $$T^{-1/2}\Sigma_s^T(x_t - x_{t-s}) \Rightarrow s\sigma_w W(1) \equiv N(0,\ s^2\sigma_w^2) \ .$$

Combining (A10)-(A16) we deduce, after a little simplification, that:

$$T(r_s - 1) \Rightarrow -A_s'/B' = -A_s/B$$

where

$$A'_s = \sigma_v^2\Big[(s/2)\{V(1)^2 + \Omega_{v0}/\sigma_v^2\} + \Sigma_{j=1}^s(s-j)\Omega_{vj}/\sigma_v^2\Big]$$

$$+\sigma_v^2\zeta^2\Big[(s/2)\{W(1)^2 + \Omega_{w0}/\sigma_w^2\} + \Sigma_{j=1}^s(s-j)\Omega_{wj}/\sigma_w^2\Big]$$

$$- s\sigma_v^2\zeta V(1)W(1) - s\sigma_v^2[V(1) - \zeta W(1)]\Big[\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\Big] + s\sigma_v^2\Big\{\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\Big\}^2$$

$$= (s\sigma_v^2/2)\Big[\{V(1) - \zeta W(1)\} - \Big\{\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\Big\}\Big]^2 + (s\sigma_v^2/2)\Big\{\int_0^1 V(t)\,dt - \zeta\int_0^1 W(t)\,dt\Big\}^2$$

$$+\sigma_v^2\Big[\{\Omega_{v0}/\sigma_v^2 + \Sigma_{j=1}^s(s-j)\Omega_{vj}/\sigma_v^2\} + \zeta^2\{\Omega_{w0}/\sigma_w^2 + \Sigma_{j=1}^s(s-j)\Omega_{wj}/\sigma_w^2\}\Big]$$

$$B' = \sigma_v^2\Big[\int_0^1 V(t)^2dt - (\int_0^1 V(t)\,dt)^2 - \zeta^2\Big\{\int_0^1 W(t)^2dt - (\int_0^1 W(t)\,dt)^2\Big\}\Big]$$

and

$$A = A'_s/\sigma_v^2\ ,\quad B = B'/\sigma_v^2\ .$$

This proves part (g) of the theorem. Part (h) follows immediately.

4. *Proof of Theorem 3.1*

We shall make use of the following result, which is proved in Phillips and Durlauf (1985, Lemma 3.1):

LEMMA A.2. If $\{z_t\}_1^\infty$ *is generated by* (6) *and if the innovation sequence* $\{\xi_t\}_1^\infty$ *satisfies Assumption* 2.1 *then as* $T \to \infty$ *:*

(A17) $$T^{-3/2}\Sigma_1^T z_t \Rightarrow \Sigma^{1/2}\int_0^1 Z(t)\,dt\ ;$$

(A18) $$T^{-2}\Sigma_1^T z_t z_t' \Rightarrow \Sigma^{1/2}\int_0^1 Z(t)Z(t)'dt\,\Sigma^{1/2}\ ;$$

(A19) $$T^{-2}\Sigma_1^T(z_t - \bar{z})(z_t - \bar{z})' \Rightarrow \Sigma^{1/2}\Big\{\int_0^1 Z(t)Z(t)'dt - \int_0^1 Z(t)\,dt\int_0^1 Z(t)'dt\Big\}\Sigma^{1/2}$$

where $Z(t)$ *is a vector Wiener process on* C^n .

To prove part (a) of Theorem 3.1 we simply note that

$$\hat{\beta} = \left\{T^{-2}\Sigma_1^T(x_t - \bar{x})(x_t - \bar{x})'\right\}^{-1}\left\{T^{-2}\Sigma_1^T(x_t - \bar{x})(y_t - \bar{y})\right\}$$

and by appealing to (A19) and the continuous mapping theorem we deduce that as $T \uparrow \infty$:

$$\text{(A20)} \quad \hat{\beta} \Rightarrow \left\{\left[\Sigma^{1/2}\left\{\int_0^1 Z(t)Z(t)'dt - \int_0^1 Z(t)dt\int_0^1 Z(t)'dt\right\}\Sigma^{1/2}\right]_{22}\right\}^{-1}$$

$$\cdot\left[\Sigma^{1/2}\left\{\int_0^1 Z(t)Z(t)'dt - \int_0^1 Z(t)dt\int_0^1 Z(t)'dt\right\}\Sigma^{1/2}\right]_{21}$$

$$= A_{22}^{-1}a_{21} ,$$

as required. To prove (b) we have

$$\hat{\alpha} = \bar{y} - \bar{x}'\hat{\beta}$$

so that using (A17) and (A20) we obtain:

$$T^{-1/2}\hat{\alpha} \Rightarrow b_1 - b_2'A_{22}^{-1}a_{21} \; ; \quad \text{as} \quad T \uparrow \infty ,$$

as required. Now

$$R^2 = \hat{\beta}'T^{-2}\Sigma_1^T(x_t - \bar{x})(x_t - \bar{x})'\hat{\beta}/T^{-2}\Sigma_1^T(y_t - \bar{y})^2$$

$$\Rightarrow a_{21}'A_{22}^{-1}a_{21}/a_{11} \; ; \quad \text{as} \quad T \uparrow \infty ,$$

proving (c). Next

$$T^{-1}F = \frac{T-m-1}{Tm}\,\frac{R^2}{1-R^2}$$

$$\Rightarrow \frac{1}{m}\,\frac{a_{21}'A_{22}^{-1}a_{21}}{a_{11} - a_{21}'A_{22}^{-1}a_{21}} \quad ; \quad \text{as} \quad T \uparrow \infty$$

proving (d). To prove (e)

$$t_{\beta_i} = \hat{\beta}_i/s_{\beta_i} \ , \quad s_{\beta_i}^2 = s^2\left[\left\{\Sigma_1^T(x_t - \overline{x})(x_t - \overline{x})'\right\}^{-1}\right]_{ii} \ .$$

Now as $T \uparrow \infty$

$$\hat{\beta}_i \Rightarrow (A_{22}^{-1}a_{21})_i$$

$$T^{-1}s^2 \Rightarrow a_{11} - a_{21}'A_{22}a_{21}$$

and

$$t_{\beta_i} = \beta_i/\left\{\sqrt{T}(s/\sqrt{T})T^{-1}\left(\left[\left\{T^{-2}\Sigma_1^T(x_t - \overline{x})(x_t - \overline{x})'\right\}^{-1}\right]_{ii}\right)^{1/2}\right\}$$

so that

$$T^{-1/2}t_{\beta_i} \Rightarrow (A_{22}^{-1}a_{21})_i/\left\{(a_{11} - a_{21}'A_{22}a_{21})^{1/2}([A_{22}^{-1}]_{ii}{}^{1/2})\right\}$$

as required for (e). Finally,

$$TDW = T^{-1}\Sigma_2^T\left\{y_t - y_{t-1} - \hat{\beta}'(x_t - x_{t-1})\right\}^2/T^{-2}\Sigma_1^T\left\{y_t - \overline{y} - \hat{\beta}'(x_t - \overline{x})\right\}^2$$

$$= \hat{b}'(T^{-1}\Sigma_2^T\dot{\xi}_t\dot{\xi}_t')\hat{b}/\hat{b}'\left(T^{-2}\Sigma_1^T(z_t - \overline{z})(z_t - \overline{z})'\right)\hat{b} \ ; \quad \hat{b}' = (1, -\hat{\beta}') \ .$$

Now, by the strong law of McLeish

$$T^{-1}\Sigma_2^T \xi_t \xi_t' \xrightarrow[\text{a.s.}]{} \Sigma_\xi = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(\xi_t \xi_t')$$

and

$$T^{-2}\Sigma_1^T (z_t - \bar{z})(z_t - \bar{z})' \Rightarrow \Sigma^{\frac{1}{2}}\left\{\int_0^1 Z(t)Z(t)'dt - \int_0^1 Z(t)\,dt \int_0^1 Z(t)'dt\right\}\Sigma^{\frac{1}{2}}$$

as $T \uparrow \infty$ by (A19). Moreover, $\hat{b} \Rightarrow \eta$ as $T \uparrow \infty$ in view of (a) above. Thus,

$$\text{TDW} \Rightarrow \eta'\Sigma_\xi\eta/\eta'A\eta = \eta'\Sigma_\xi\eta/\{a_{11} - a_{21}'A_{22}a_{21}\}$$

as required for (f).

REFERENCES

Bhargava, A. (1984). "On the theory of testing for unit roots in observed time series," mimeographed, London School of Economics.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley: New York.

Box, G. E. P. and G. M. Jenkins (1970). *Time Series Analysis: Forecasting and Control.* Holden Day: San Francisco.

Donsker, M. D. (1951). "An invariance principle for certain probability limit theorems," *Memoirs of the American Mathematical Society*, 6, 1-12.

Erdös, P. and M. Kac (1946). "On certain limit theorems in the theory of probability," *Bulletin of the American Mathematical Society*, 52, 292-302.

Granger, C. W. J. and P. Newbold (1974). "Spurious regressions in econometrics," *Journal of Econometrics*, 2, 111-120.

Granger, C. W. J. and P. Newbold (1977). *Forecasting Economic Time Series.* Academic Press: New York.

Granger, C. W. J. and R. F. Engle (1985). "Dynamic model specification with equilibrium constraints: Cointegration and error correction," UCSD Discussion Paper No. 85-18.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Applications.* New York: Academic Press.

Herrndorf, N. (1984). "A functional central limit theorem for weakly dependent sequences of random variables," *Annals of Probability*, 12, 141-153.

Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and Stationary Sequences of Random Variables.* Groningen: Wolters-Noordhoff.

McLeish, D. L. (1975a). "Invariance principles for dependent variables," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32, 165-178.

McLeish, D. L. (1975b). "A maximal inequality and dependent strong laws," *Annals of Probability*, 3, 829-839.

Plosser, C. I. and G. W. Schwert (1978). "Money income and sunspots: Measuring Economic Relationships and the Effects of Differencing," *Journal of Monetary Economics*, 4, 637-660.

Phillips, P. C. B. (1985a). "Time series regression with unit roots," Cowles Foundation Discussion Paper No. 740, Yale University.

Phillips, P. C. B. (1985b). "Understanding spurious regressions in econometrics," Cowles Foundation Discussion Paper No. 757, July 1985.

Phillips, P. C. B. (1985c). "Asymptotic expansions in nonstationary vector autoregressions," Cowles Foundation Discussion Paper No. 765.

Phillips, P. C. B. and S. N. Durlauf (1985). "Multiple time series regression with integrated processes," Cowles Foundation Discussion Paper No. 768.

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer-Verlag.

White, H. (1984). *Asymptotic Theory for Econometricians.* New York: Academic Press.